

Published in Journals: Atmosphere, Climate, Geomatics, Hydrology,  
Remote Sensing and Water

Topic Reprint

---

# Advances in Hydrological Remote Sensing

---

Edited by  
Hailong Liu and Liangliang Jiang

[mdpi.com/topics](https://mdpi.com/topics)





# **Advances in Hydrological Remote Sensing**



# Advances in Hydrological Remote Sensing

Topic Editors

**Hailong Liu**

**Liangliang Jiang**



Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

*Topic Editors*

Hailong Liu  
School of Resources  
and Environment  
University of Electronic  
Science and Technology  
of China  
Chengdu  
China

Liangliang Jiang  
School of Geography  
and Tourism  
Chongqing Normal  
University  
Chongqing  
China

*Editorial Office*

MDPI AG  
Grosspeteranlage 5  
4052 Basel, Switzerland

This is a reprint of the Topic, published open access by the journals *Atmosphere* (ISSN 2073-4433), *Climate* (ISSN 2225-1154), *Geomatics* (ISSN 2673-7418), *Hydrology* (ISSN 2306-5338), *Remote Sensing* (ISSN 2072-4292) and *Water* (ISSN 2073-4441), freely accessible at: <https://www.mdpi.com/topics/7K510482ZO>.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

|                                                                                                            |
|------------------------------------------------------------------------------------------------------------|
| Lastname, A.A.; Lastname, B.B. Article Title. <i>Journal Name</i> <b>Year</b> , Volume Number, Page Range. |
|------------------------------------------------------------------------------------------------------------|

**ISBN 978-3-7258-8843-6 (Hbk)**

**ISBN 978-3-7258-8844-3 (PDF)**

**<https://doi.org/10.3390/books978-3-7258-8844-3>**

© 2026 by the authors. Articles in this reprint are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The reprint as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

# Contents

About the Editors . . . . . vii

**Esthela Campos Lara, Julián González-Trinidad, David Armando Contreras Solorio, Ada Rebeca Rodríguez Contreras, Hugo Enrique Júnez-Ferreira, Sandra Dávila-Hernández, et al.**

Comparative Performance of SCS-CN and Green-Ampt Methods in HEC-HMS Under Spatio-Temporal Rainfall Variability in a Semi-Arid Mexican Basin  
Reprinted from: *Hydrology* **2026**, *13*, 216, <https://doi.org/10.3390/hydrology13080216> . . . . . 1

**Mostafa Mashal, Doaa Amin, Mona A. Hagraas and Ashraf M. Elmoustafa**

Assessing Flood Susceptibility Using Machine Learning in Arid Regions  
Reprinted from: *Geomatics* **2026**, *6*, 78, <https://doi.org/10.3390/geomatics6040078> . . . . . 31

**Thiago Lappicy, Daniel Beltrão, Luana Oliveira Sales, Tati Almeida, Guilherme Gomes Pessoa, Saulo Souza, et al.**

A Large-Scale Evaluation of SWOT-Derived Water Surface Elevations: Precision Drivers and Strategies to Enhance Data Availability  
Reprinted from: *Remote Sens.* **2026**, *18*, 1609, <https://doi.org/10.3390/rs18101609> . . . . . 62

**Tong Zhao, Chuanxun Hou, Zhili Zhang and Zhaofa Zhou**

An Intelligent Gated Fusion Network for Waterbody Recognition in Multispectral Remote Sensing Imagery  
Reprinted from: *Remote Sens.* **2026**, *18*, 1088, <https://doi.org/10.3390/rs18071088> . . . . . 86

**Youssef Laalaoui, Naïma El Assaoui, Oumaima Ouahine, Thanh Thi Nguyen and Ahmed M. Saqr**

Hybrid Ensemble Learning for TWSA Prediction in Water-Stressed Regions: A Case Study from Casablanca–Settat Region, Morocco  
Reprinted from: *Hydrology* **2026**, *13*, 53, <https://doi.org/10.3390/hydrology13020053> . . . . . 123

**Qingyuan Luo, Dongzhi Wang, Lina Liu, Caihong Hu and Chengshuai Liu**

Fusion and Evaluation of Multi-Source Satellite Remote Sensing Precipitation Products Based on Transformer Machine Learning  
Reprinted from: *Water* **2026**, *18*, 358, <https://doi.org/10.3390/w18030358> . . . . . 148

**Kehang Fang, Feng Wu, Xing Gao and Zhihui Li**

Multi-Source Data Fusion and Heuristic-Optimized Machine Learning for Large-Scale River Water Quality Parameters Monitoring  
Reprinted from: *Remote Sens.* **2026**, *18*, 320, <https://doi.org/10.3390/rs18020320> . . . . . 169

**Jin Zhao, Kexin Yu, Shuaiqi Xu, Maorong Wang, Yiguang Yang, Degang Xu, et al.**

Advances in Laser-Induced Acoustic Technology for Underwater Detection  
Reprinted from: *Water* **2025**, *17*, 3285, <https://doi.org/10.3390/w17223285> . . . . . 189

**Hanmin Yin, Xiaohan Liao, Huping Ye, Jie Bai, Wentao Yu, Yue Li, et al.**

Spatiotemporal Patterns of Climate-Vegetation Regulation of Soil Moisture with Phenological Feedback Effects Using Satellite Data  
Reprinted from: *Remote Sens.* **2025**, *17*, 3714, <https://doi.org/10.3390/rs17223714> . . . . . 210

**Hailong Liu, Aminjon Gulakhmadov and Firdavs Shaimuradov**

Impact Analysis of Climate Change on Hydropower Resource Development in the Vakhsh River Basin of Tajikistan  
Reprinted from: *Hydrology* **2025**, *12*, 294, <https://doi.org/10.3390/hydrology12110294> . . . . . 243

|                                                                                                                                                                                                                                                                                                                                                                                                                                 |            |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>Yong Chang, Nan Mu, Yaoyong Qi and Ling Liu</b><br>Quantifying the Contribution of Global Precipitation Product Uncertainty to Ensemble Discharge Simulations and Projections: A Case Study in the Liujiang Catchment, Southwest China<br>Reprinted from: <i>Atmosphere</i> <b>2025</b> , 16, 1260, <a href="https://doi.org/10.3390/atmos16111260">https://doi.org/10.3390/atmos16111260</a> . . . . .                      | <b>263</b> |
| <b>Aikaterini Pappa, John Kalogiros, Maria Tombrou, Christos Spyrou, Marios N. Anagnostou, George Varlas, et al.</b><br>Assessing the Impact of Assimilated Remote Sensing Retrievals of Precipitation on Nowcasting a Rainfall Event in Attica, Greece<br>Reprinted from: <i>Hydrology</i> <b>2025</b> , 12, 198, <a href="https://doi.org/10.3390/hydrology12080198">https://doi.org/10.3390/hydrology12080198</a> . . . . .  | <b>282</b> |
| <b>Andrés Gabriel Arguedas Chaverri, Rogério Hartung Toppa and Kelly Cristina Tonello</b><br>From Climate to Cloud: Advancing Fog Detection Through Satellite Imagery<br>Reprinted from: <i>Climate</i> <b>2025</b> , 13, 110, <a href="https://doi.org/10.3390/cli13060110">https://doi.org/10.3390/cli13060110</a> . . . . .                                                                                                  | <b>310</b> |
| <b>Liangliang Jiang, Guangming Wu, Qijin Li and Xiaoran Liu</b><br>Vegetation Response to Flash Drought Events Considering Resilience in Southwestern China<br>Reprinted from: <i>Water</i> <b>2025</b> , 17, 653, <a href="https://doi.org/10.3390/w17050653">https://doi.org/10.3390/w17050653</a> . . . . .                                                                                                                  | <b>325</b> |
| <b>Fátima Goretti García-Miranda, Claudia Muro, Yolanda Alvarado, José Luis Expósito-Castillo and Héctor Víctor Cabadas-Báez</b><br>Eutrophication Conditions in Two High Mountain Lakes: The Influence of Climate Conditions and Environmental Pollution<br>Reprinted from: <i>Hydrology</i> <b>2025</b> , 12, 32, <a href="https://doi.org/10.3390/hydrology12020032">https://doi.org/10.3390/hydrology12020032</a> . . . . . | <b>354</b> |
| <b>Long Jiang, Zhiqiang Lin, Zhenbo Zhou, Hongxin Luo, Jiafeng Zheng, Dongsheng Su and Minhong Song</b><br>Monitoring and Disaster Assessment of Glacier Lake Outburst in High Mountains Asian Using Multi-Satellites and HEC-RAS: A Case of Kyagar in 2018<br>Reprinted from: <i>Remote Sens.</i> <b>2024</b> , 16, 4447, <a href="https://doi.org/10.3390/rs16234447">https://doi.org/10.3390/rs16234447</a> . . . . .        | <b>370</b> |
| <b>Zhimin Hu, Min Feng, Yijie Sui, Dezhao Yan, Kuo Zhang, Jinhao Xu, Rui Liu and Earina Sthapit</b><br>Identifying Alpine Lakes with Shoreline Features<br>Reprinted from: <i>Water</i> <b>2024</b> , 16, 3287, <a href="https://doi.org/10.3390/w16223287">https://doi.org/10.3390/w16223287</a> . . . . .                                                                                                                     | <b>390</b> |

# About the Editors

## Hailong Liu

Hailong Liu is a Professor and PhD supervisor at the School of Resources and Environment, University of Electronic Science and Technology of China, and Director of Sichuan Engineering Research Center for Monitoring and Early Warning of Natural Disasters in Extreme Environments. He earned his PhD in Cartography and Geographic Information System from the Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, and was a visiting scholar at KU Leuven, Belgium in 2014 before joining UESTC in 2018. His research concentrates on quantitative remote sensing, hydrology and water resources, watershed ecohydrology, intelligent eco-environment monitoring, multi-source environmental big data and deep learning-based prediction. He has led and participated in more than 30 research projects including NSFC programs, the National Key R&D Program, the 973 Program and the Third Xinjiang Scientific Expedition. He has published over 100 papers (more than 60 SCI articles) and three monographs, and received several provincial and ministerial scientific and technological awards. He acts as director of the Arid Region Committee of Geographical Society of China and is an Editorial Board Member of the *Journal of Image and Graphics*, and regularly serves as a reviewer for many journals on remote sensing, hydrology and environment.

## Liangliang Jiang

Liangliang Jiang is an Associate Professor and Master's Supervisor, Assistant to the Dean, Deputy Director of the Department of Geographic Information Science at the School of Geography and Tourism, Chongqing Normal University. He is Young Scholar of Bayu Program and a council member of the Geographical Society of Chongqing. He obtained a joint doctoral degree in Geography from Ghent University (Belgium) and the University of Chinese Academy of Sciences in 2020, and completed his postdoctoral research at Chongqing Meteorological Bureau. His research focuses on remote sensing monitoring of resources and environment. He has presided over 12 provincial and national research projects including the National Natural Science Foundation of China, the sub-project of the National Basic Science and Technology Resources Investigation Program, the China Postdoctoral Science Foundation and the Key Program of Chongqing Natural Science Foundation. He has published more than 60 academic papers (over 40 SCI/EI articles), 35 papers as first or corresponding author, authored four monographs and obtained eight patents. He has undertaken multiple teaching reform projects and guided students to win first prizes in the National GIS Skills Competition. He received the President's Scholarship of the Chinese Academy of Sciences.





## Article

# Comparative Performance of SCS-CN and Green-Ampt Methods in HEC-HMS Under Spatio-Temporal Rainfall Variability in a Semi-Arid Mexican Basin

Esthela Campos Lara <sup>1</sup>, Julián González-Trinidad <sup>2</sup>, David Armando Contreras Solorio <sup>2,\*</sup>,  
Ada Rebeca Rodríguez Contreras <sup>2</sup>, Hugo Enrique Júnez-Ferreira <sup>2</sup>, Sandra Dávila-Hernández <sup>3,\*</sup>,  
Manuel Ibarra Reyes <sup>2</sup>, Ana Isabel Veyna Gómez <sup>1</sup>, Raúl Ulices Silva Avalos <sup>2</sup> and Cruz Octavio Robles Rovelto <sup>2</sup>

<sup>1</sup> Doctoral Program in Science and Technology of Light and Matter, Universidad Autónoma de Zacatecas, Circuito Marie Curie S/N, Parque de Ciencia y Tecnología QUANTUM Ciudad del Conocimiento, Zacatecas 98160, Mexico; esthela.campos.lara@uaz.edu.mx (E.C.L.); ana.veyna85@uaz.edu.mx (A.I.V.G.)

<sup>2</sup> Faculty of the Doctoral Program in Science and Technology of Light and Matter, Universidad Autónoma de Zacatecas, Circuito Marie Curie S/N, Parque de Ciencia y Tecnología QUANTUM Ciudad del Conocimiento, Zacatecas 98160, Mexico; jgonza@uaz.edu.mx (J.G.-T.); rebeca.contreras@uaz.edu.mx (A.R.R.C.); hejunez@uaz.edu.mx (H.E.J.-F.); mibarralumet@uaz.edu.mx (M.I.R.); raul.silva@uaz.edu.mx (R.U.S.A.); octavio.robles@uaz.edu.mx (C.O.R.R.)

<sup>3</sup> Program in Topography and Hydrography, Universidad Autónoma de Zacatecas, Ramón López Velarde 801, Centro, Zacatecas 98600, Mexico

\* Correspondence: dacs@uaz.edu.mx (D.A.C.S.); sandra.davila@uaz.edu.mx (S.D.-H.)

## Abstract

Pronounced spatio-temporal variability of rainfall in semi-arid basins remains a central challenge for rainfall–infiltration–runoff modeling, particularly in ungauged or newly instrumented catchments where continuous soil-moisture data are unavailable. This study evaluates the comparative predictive performance of the SCS-CN and Green-Ampt (GA) infiltration methods, implemented within HEC-HMS at the sub-basin scale, using eight rainfall–runoff analysis time windows recorded during the 2020–2025 rainy seasons within a monitoring network operational since October 2019 in an instrumented semi-arid basin in Mexico. A blind-validation framework was adopted: parameters for both methods were derived a priori from tabulated sources indexed by land use, hydrologic soil group, soil textural class, and locally supported by textural analysis at three depths per sub-basin, in situ testing of saturated hydraulic conductivity, and gravimetric determination of field capacity; the initial moisture content required by GA was set equal to the measured field capacity ( $\theta_i = \theta_{fc}$ ) to equate initial conditions between the two methods. Spatially distributed rainfall was captured by four monitoring stations under a one-to-one gauge–sub-basin assignment scheme, with monthly rainfall depth varying from 22.8 to 204.9 mm across the four sub-basins. Both methods reproduced observed discharge with varying levels of agreement: SCS-CN yielded very good performance (Pearson  $R = 0.95$ ; Nash–Sutcliffe efficiency  $NSE = 0.76$ ), whereas Green-Ampt yielded moderate correlation but unsatisfactory  $NSE$  ( $R = 0.70$ ;  $NSE = 0.45$ ) against the Levellogger records. Contrary to the initial expectation that the physically based GA would outperform SCS-CN, SCS-CN yielded substantially higher performance across windows, with the two simulated discharge series differing by a mean absolute deviation of  $36.9 \text{ m}^3/\text{s}$ . A systematic sensitivity analysis ( $\pm 6\%$ ,  $\pm 10\%$ ,  $\pm 20\%$  perturbations) revealed an asymmetric response: SCS-CN was highly sensitive to Curve Number perturbations (mean-deviation amplitude  $65.9 \text{ m}^3/\text{s}$ ), whereas Green-Ampt was nearly insensitive to its compound soil-hydraulic parameterization (amplitude  $3.2 \text{ m}^3/\text{s}$ ), indicating a structural limitation of the physically based method under blind validation.

**Keywords:** hydrological simulation; curve number; HEC-HMS; blind validation; rainfall; infiltration; runoff; semi-arid hydrology; sub-basin scale

## 1. Introduction

Rainfall in semi-arid basins is characterized by pronounced spatial and temporal variability, determined primarily by short, localized storms whose intensity, duration, and spatial extent depart substantially from those of systems in humid regimes. Under a warming climate, this variability is being amplified. Recent decades have witnessed both more intense torrential rainfall and longer dry spells, and semi-arid regions, which currently cover more than 40% of the Earth's terrestrial surface and continue to expand [1], have been identified as highly vulnerable environments [2]. Human-induced climate change is intensifying the global water cycle, altering precipitation regimes and further amplifying the spatio-temporal heterogeneity of rainfall across most regions of the world [2]. In this context, deepening the understanding of rainfall behavior between the surface-subsurface boundaries and quantifying the influence of climate change on its spatio-temporal distribution remain central scientific priorities [3], with direct implications for aquifer recharge, surface storage, and water-availability assessments under mounting anthropogenic pressure on natural resources [4].

Precipitation interacts with the soil to partition into infiltration and direct runoff, the latter integrated at a control point as surface storage [5]. These processes have historically been analyzed as isolated entities because of instrumental limitations, but the progressive instrumentation of hydrological basins is enabling a more integrated monitoring of the components of the hydrological cycle, supporting more reliable estimates of water availability across topography, land use, and soil type [6–8]. This shift is critical because infiltration (the process by which surface water enters the soil matrix) depends on multiple soil-hydraulic and topographic factors, including antecedent moisture content, saturation, grain-size distribution, hydraulic conductivity, slope, and rainfall intensity [9–15]; and requires a proper hydrodynamic characterization of the porous medium [16,17]. Mathematical representations of the rainfall–infiltration–runoff relationship have evolved since the 1970s, and their coupling with geographic information systems has substantially expanded their applicability across regions and data settings [10,18–22].

Hydrological models provide a cost-effective framework to estimate infiltration and runoff from rainfall inputs and are particularly valuable in ungauged or newly instrumented basins characteristic of semi-arid regions. Among widely applied tools, the Hydrologic Modeling System (HEC-HMS), developed by the U.S. Army Corps of Engineers, has become a standard semi-distributed platform for representing hydrological processes at sub-basin scale [23–26]. HEC-HMS relies on three basic data sets—meteorological, hydrometric, and geospatial (digital elevation, land use, soil type)—and supports multiple infiltration formulations [5,9,27]. Two of these formulations are extensively used: the Soil Conservation Service Curve Number method (SCS-CN), an empirical–lumped approach that estimates runoff from land use, hydrologic soil group, and antecedent moisture condition without explicitly resolving rainfall intensity or duration [24]; and the Green-Ampt method (GA), a physically based formulation that represents wetting-front dynamics as a function of soil hydraulic properties and rainfall intensity and that is better suited to stratified soils and to storms of variable duration [24]. Further comparative exploration of both methods in semi-arid basins remains an open research need [28].

In Mexico, SCS-CN remains the predominant method for design-discharge estimation, typically applied with 24 h maximum rainfall values and return-period criteria, whereas

comparative evaluations of SCS-CN and GA under measured spatio-temporal rainfall variability remain scarce.

Despite decades of hydrological progress, conventional practice still frequently relies on the assumption of spatially uniform rainfall across the basin during a given storm, an assumption repeatedly questioned in semi-arid settings, where highly localized rainfall introduces biases in infiltration and runoff estimates [29]. Field evidence indicates that direct runoff tends to dominate in the lower portions of the basin, while infiltration prevails in the upper sections and along the channels [30], and that rainfall recharge can be reasonably estimated through infiltration-factor approaches complemented by empirical equations in controlled and uncontrolled areas during the non-monsoon season [31]. Closing this gap requires comparative evaluations that: (i) use measured (rather than interpolated) spatial rainfall from a purposefully instrumented network, (ii) operate at the sub-basin scale, (iii) cover multi-annual periods to include a diverse population of events, and (iv) adopt a blind-validation framework in which parameters are derived a priori from independent sources, thereby avoiding the overfitting artifacts that may inflate reported performance under conventional calibration schemes [32].

The present study addresses this gap by evaluating the comparative predictive performance of the SCS-CN and Green-Ampt methods, implemented within HEC-HMS at sub-basin scale, using eight rainfall–runoff analysis time windows recorded during the 2020–2025 rainy seasons in an instrumented semi-arid basin in central Mexico, monitored continuously since October 2019. Two working hypotheses are tested: (1) under a blind-validation framework, both methods will exhibit comparable predictive skill for short events with pronounced spatial rainfall variability; and (2) given the limited availability of continuous antecedent-moisture information typical of semi-arid regions, the operational performance of SCS-CN will be at least equivalent to that of GA, despite the stronger physical basis of the latter. To test these hypotheses, four specific objectives are pursued: (a) To document the spatio-temporal variability of rainfall over the study basin during 2019–2025 through a distributed network of four monitoring stations and to characterize the eight analysis time windows retained for examination. (b) To derive a priori parameter sets for the SCS-CN and Green-Ampt methods from standard tabulated sources indexed by land use, hydrologic soil group, and antecedent moisture (SCS-CN) and by soil textural class (Green-Ampt), locally supported by Bouyoucos textural analysis at three depths per sub-basin. Pronounced spatio-temporal variability of rainfall in semi-arid basins remains a central challenge for rainfall–infiltration–runoff modeling, particularly in ungauged or newly instrumented catchments where continuous soil-moisture data are unavailable. This study evaluates the comparative predictive performance of the SCS-CN and Green-Ampt (GA) infiltration methods implemented within HEC-HMS at sub-basin scale using eight rainfall–runoff analysis time windows recorded during the 2020–2025 rainy seasons within a monitoring network operational since October 2019, in an instrumented semi-arid basin in Mexico. A blind-validation framework was adopted: parameters for both methods were derived a priori from tabulated sources indexed by land use, hydrologic soil group, and soil textural class and locally supported by textural analysis at three depths per sub-basin, in situ testing of saturated hydraulic conductivity, and gravimetric determination of field capacity; the initial moisture content required by GA was set equal to the measured field capacity ( $\theta_i = \theta_{fc}$ ) to equate initial conditions between the two methods. Spatially distributed rainfall was captured by four monitoring stations under a one-to-one gauge–sub-basin assignment scheme, with observed rainfall depth ranging from 9.1 to 90.6 mm per analysis window across the four sub-basins. Both methods reproduced observed discharge with varying levels of agreement: SCS-CN yielded very good performance (Pearson  $R = 0.95$ ; Nash–Sutcliffe efficiency  $NSE = 0.76$ ), whereas Green-Ampt yielded moderate correlation but unsatisfactory  $NSE$  ( $R = 0.70$ ;  $NSE = 0.45$ ) against the Levelogger records. Contrary to the initial expectation that the physically

based GA would outperform SCS-CN, SCS-CN yielded substantially higher performance across windows, with the two simulated discharge series differing by a mean absolute deviation of  $36.9 \text{ m}^3/\text{s}$ . A systematic sensitivity analysis ( $\pm 6\%$ ,  $\pm 10\%$ ,  $\pm 20\%$  perturbations) revealed an asymmetric response: SCS-CN was highly sensitive to Curve Number perturbations (mean-deviation amplitude  $65.9 \text{ m}^3/\text{s}$ ), whereas Green-Ampt was nearly insensitive to its compound soil-hydraulic parameterization (amplitude  $3.2 \text{ m}^3/\text{s}$ ), indicating a structural limitation of the physically based method under blind validation, in situ double-ring infiltrometer testing of saturated hydraulic conductivity, and gravimetric determination of field capacity. (c) To evaluate, under a blind-validation framework, the predictive performance of the SCS-CN and Green-Ampt methods against observed discharge, using standard goodness-of-fit metrics at window scale. (d) To assess the robustness of each method to uncertainty in its a priori parameter estimates through a systematic sensitivity analysis of the main model parameters. To the best of our knowledge, this study contributes the first multi-annual (2019–2025), instrumented, sub-basin-scale, blind-validation comparison of SCS-CN and Green-Ampt under measured spatio-temporal rainfall variability in a semi-arid Mexican basin.

## 2. Materials and Methods

### 2.1. Instrumented Basin

The study area is called the “Chilitas” basin, located in central Mexico (Figure 1) in the state of Zacatecas. It has an approximate area of  $100 \text{ km}^2$  and is a rural basin with fewer than 500 inhabitants distributed across two small towns. There is no wastewater discharge into the watercourses, and there are practically no reservoirs or appreciable natural depressions. For hydrological modeling, the basin was delineated into four sub-basins, hereafter referred to as Lower, Middle, Upper, and Outlet, matching the location of the four rainfall monitoring stations. The land use and cover are distributed as follows: 62.4% natural pasture, 18.6% rainfed annual agriculture, 9.3% succulent scrubland, 4.2% irrigated annual agriculture, 3.6% secondary shrub vegetation, 0.43% urban settlements, and less than 0.03% oak forest; the remaining  $\sim 1.4\%$  corresponds to bare soil. The predominant soil type is Phaeozem (61.8%), followed by Kastanozem (32.5%) and Leptosol (3.7%), with the remaining  $\sim 2\%$  corresponding to unclassified types [33]. The climate is semi-arid with summer rainfall, mainly from June to September; average minimum and maximum temperatures are  $-2.8^\circ\text{C}$  and  $29^\circ\text{C}$ , respectively, and average annual rainfall is approximately 428 mm [34].

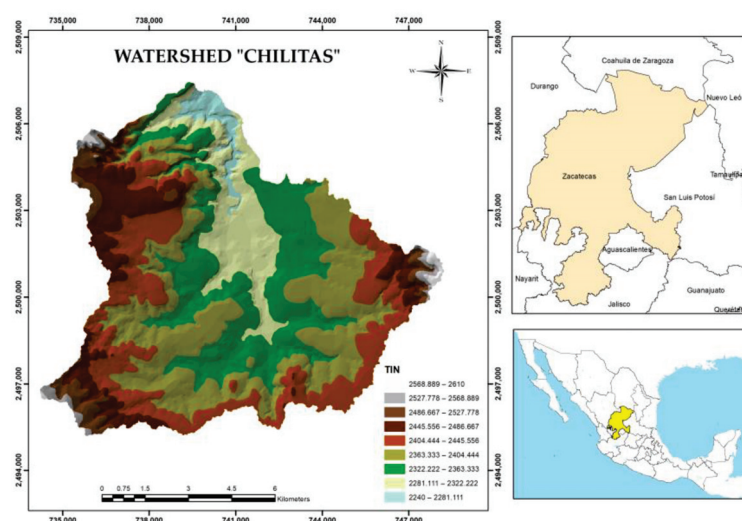


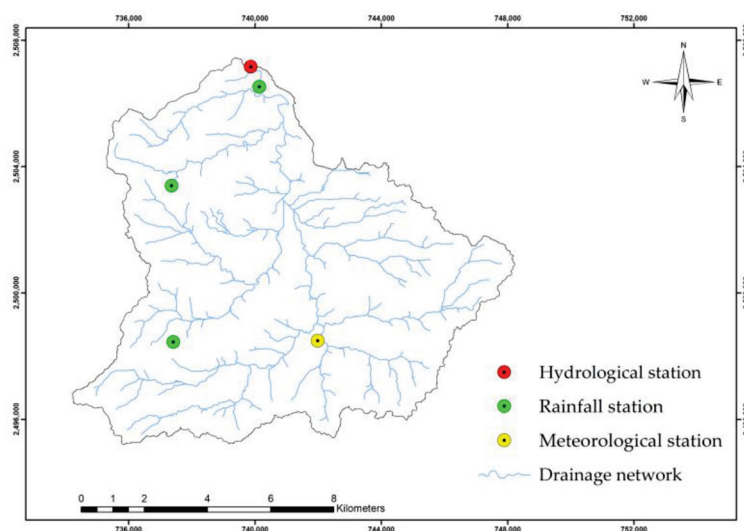
Figure 1. Chilitas basin, Zacatecas, Mexico.

## 2.2. Data Collection and Instrumentation

The comparative analysis undertaken in this research relies on hydrometeorological records collected within the basin by a permanent monitoring network operational since October 2019 and covering the 2020–2025 rainy seasons within the broader 2019–2025 study period. The network comprises four rainfall stations distributed across the drainage area and one pressure transducer at the basin outlet, complemented by a site-specific stage–discharge rating curve. The instrumentation and network configuration are described in Sections 2.2.1 and 2.2.2; the criteria used to select the rainfall–runoff analysis time windows, for the modeling comparison, in Section 2.2.3; and the rating curve used to convert stage to discharge at the basin outlet, in Section 2.2.4.

### 2.2.1. Rainfall Monitoring Network

Rainfall within the basin is monitored by four automatic stations distributed strategically across the drainage area to capture the pronounced spatio-temporal variability that constitutes a central object of this study (Figure 2): three tipping-bucket pluviographs and one automatic weather station equipped with a rainfall sensor. Each of the four monitoring stations is located within one of the four delineated sub-basins, enabling a direct one-to-one assignment between rain gauges and sub-basins. All rainfall sensors are equipped with self-emptying tipping buckets connected to internal dataloggers configured to record rainfall at fixed 5 min intervals during the rainy season, with a tipping resolution of 0.2 mm for all four instruments. The network was commissioned in October 2019 and has been in continuous operation since then; because commissioning occurred after the end of the regional monsoon season, no rainfall–runoff analysis time windows satisfying the selection criteria (detailed in Section 2.2.3) were recorded during the remaining months of 2019, and the analyzed windows therefore span the 2020–2025 rainy seasons.



**Figure 2.** Hydrographic network of the “Chilitas” basin and measuring equipment.

### 2.2.2. Streamflow Monitoring

Streamflow at the basin outlet (Hydrological station in Figure 2) was monitored using a pressure transducer (Solinst Levellogger model CRS451; Solinst Canada Ltd., Georgetown, ON, Canada) installed at a stable control cross-section whose geometry closely reproduces the natural channel morphology, operational since October 2019 in parallel with the rainfall network. The instrument integrates pressure and temperature sensors and reports water depth with an accuracy of  $\pm 0.1\%$  of full scale, expressed as Total Error Band (TEB). The



transducer was configured in event-triggered logging mode: a stage record is stored whenever a positive or negative pressure change exceeding 5 mm (equivalent to approximately 5 mm of water depth) is detected between the current reading and the previous stored value. This sampling strategy is well suited to the intermittent flow regime of the study basin, characteristic of semi-arid environments, and offers two operational advantages: (i) it captures with high temporal density the dynamic phases of runoff events (rising limb, peak, and recession limb) where stage changes are large and rapid; and (ii) it minimizes memory and battery consumption during prolonged dry-channel periods. Given the 5 mm event-triggering threshold, the maximum uncertainty in any individual stage value is bounded and remains well below the natural variability of the observed hydrographs; consequently, it does not represent a limiting factor for the comparative evaluation of the SCS-CN and Green-Ampt methods.

### 2.2.3. Selection of Analysis Time Windows

From the continuous 2020–2025 hydrometeorological record (within the broader monitoring period initiated in October 2019), eight rainfall–runoff analysis time windows were selected for the comparative analysis. Rather than isolated storm events, each window corresponds to a bounded time interval, defined by an explicit start date/time and end date/time, during which the rainfall–runoff dynamics could be fully documented across the four monitoring stations and the pressure transducer at the outlet; each window may include one or several rainfall pulses together with their associated recession phases. Window selection was regulated by three explicit and independent criteria, applied sequentially:

- (i) Simultaneous availability of complete rainfall and discharge records; that is, all four rainfall stations and the outlet pressure transducer were required to be operational continuously during the window, as well as in the preceding and following recession phases, ensuring that both the model inputs and the reference variable are fully documented in space and time within the window.
- (ii) Absence of gaps or missing values in either the rainfall or the discharge series during the event window. Because the window boundaries themselves were delineated to satisfy the first criterion, no gap-filling or interpolation procedures were applied to the input series within the analyzed windows.
- (iii) Runoff-generating capacity: only windows producing measurable direct runoff at the outlet were retained, ensuring that the hydrological response was quantifiable above the detection threshold of the instrumentation and that the comparative evaluation of the SCS-CN and Green-Ampt methods was applied to hydrologically meaningful periods.

The resulting eight windows are summarized in Table 1. Because the monitoring network was commissioned in October 2019 (after the end of the regional monsoon season), no rainfall–runoff windows satisfying these criteria were recorded during the remaining months of that year; consequently, the eight windows span the 2020–2025 rainy seasons. Their durations range from 29.00 h (Window 8, July 2025) to 216.00 h (Window 7, June 2025), with a median of approximately 71 h, reflecting the fact that individual windows may include one or several rainfall pulses together with their recession phases. All windows fall within the rainy season of the study region (June–September), consistent with the North American monsoon regime that dominates convective rainfall generation in semi-arid central Mexico.

This selection strategy prioritizes data quality and completeness over statistical representativeness of the full climatological distribution of storms in the basin; the implications

of the resulting sample size (8 windows) for the generality of the reported performance metrics are addressed in Section 4.6.

**Table 1.** Rainfall–runoff analysis time windows selected for the comparative modeling exercise.

| Analysis Time Windows | Date/Time                |                          | Duration (h) |
|-----------------------|--------------------------|--------------------------|--------------|
|                       | Start                    | End                      |              |
| 1                     | 29 August 2020, 16:00    | 30 August 2020, 22:30    | 30.50        |
| 2                     | 19 July 2021, 04:55      | 22 July 2021, 08:00      | 75.08        |
| 3                     | 14 September 2021, 19:30 | 17 September 2021, 23:55 | 76.42        |
| 4                     | 26 August 2022, 19:30    | 30 August 2022, 21:40    | 98.17        |
| 5                     | 30 August 2023, 13:10    | 31 August 2023, 23:30    | 34.33        |
| 6                     | 23 July 2024, 15:15      | 26 July 2024, 09:30      | 66.25        |
| 7                     | 21 June 2025, 23:50      | 30 June 2025, 23:50      | 216.00       |
| 8                     | 16 July 2025, 16:00      | 17 July 2025, 21:00      | 29.00        |

#### 2.2.4. Rating Curve

The stage series recorded by the pressure transducer was converted to discharge using a site-specific rating curve developed for the control cross-section at the basin outlet, following the USGS standard practice for discharge measurements at gauging stations [35]. The rating curve was constructed through a combined hydraulic-modeling and field-calibration approach: (i) two-dimensional hydraulic simulation of the control cross-section was performed with HEC-RAS (version 5.0) to obtain the stage–discharge relationship based on the geometry, roughness, and flow regime of the reach; (ii) a complementary one-dimensional hydraulic analysis provided an independent check on the 2D solution; and (iii) the resulting relationship was calibrated against direct discharge measurements collected in the field during runoff windows spanning the 2020–2025 period. The final rating curve is expressed as the cubic polynomial:

$$Q = -0.9002h^3 + 18.1847h^2 - 2.5329h + 0.0131 \quad (1)$$

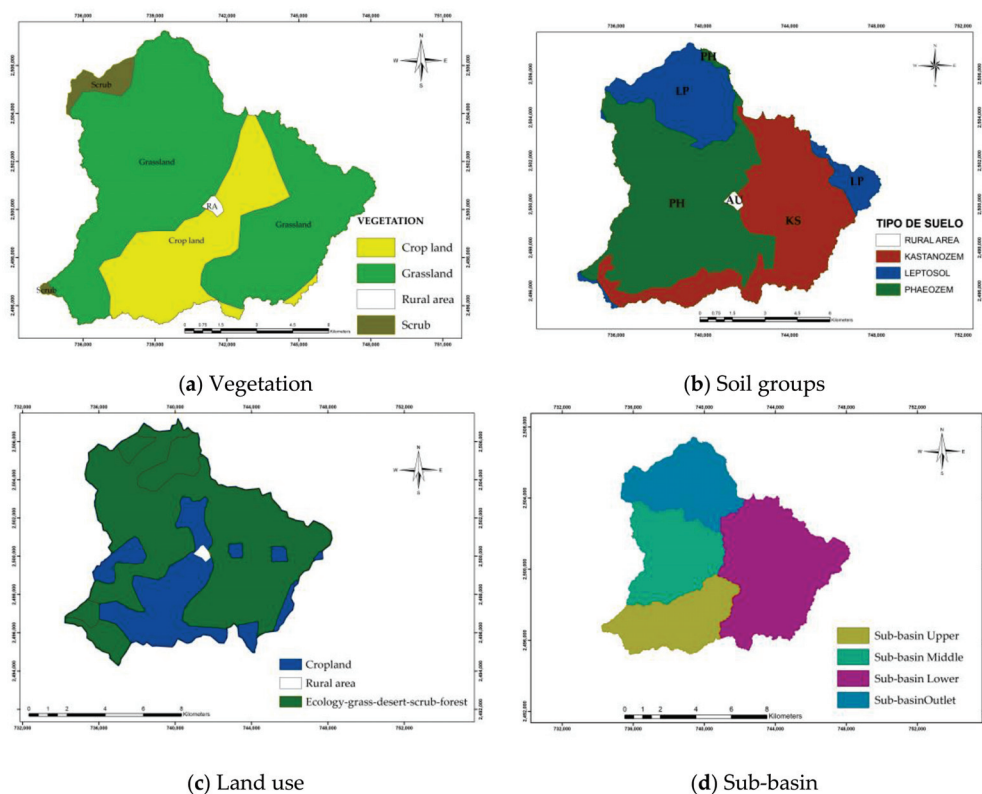
where  $Q$  is discharge ( $\text{m}^3/\text{s}$ ), and  $h$  is water depth (m), with a coefficient of determination  $R^2 = 0.95$  against the direct discharge measurements. The rating curve was applied only within its calibrated stage range, and all analysis windows examined in this study produced peak stages within this range; therefore, no extrapolation of the rating curve was required.

The observed discharge series recorded by the pressure transducer at the basin outlet was used exclusively as a reference for performance evaluation, not for parameter estimation of either infiltration method.

#### 2.3. Methodology

Starting with a 5 m resolution digital elevation model (DEM) obtained from INEGI [36] and supporting cartography, the vegetation cover (Figure 3a), edaphic diversity (soil-type distribution; Figure 3b), land use (Figure 3c), and sub-basin delineation (Figure 3d) of the basin were derived using geoprocessing tools in ArcGIS 10.8. For the hydrological monitoring and analysis of rainfall, the basin was accordingly divided into four sub-basins (referred to hereafter as Lower, Middle, Upper, and Outlet), allowing the spatio-temporal behavior of rainfall to be analyzed. All derived layers were verified against field observations.

Table 2 summarizes the geomorphological parameters of the four sub-basins, which are subsequently used to simulate rainfall–infiltration–runoff dynamics under the eight analysis time windows described in Section 2.2.3.



**Figure 3.** Thematic maps of the “Chilitas” basin: (a) vegetation cover; (b) soil-type distribution; (c) land use; and (d) delineation of the four sub-basins (Lower, Middle, Upper, and Outlet).

**Table 2.** Geomorphological parameters.

| Sub-Basin | Latitude      | Length         | Altitude | A (km <sup>2</sup> ) | L (km) | S <sub>c</sub> (m/m) | T <sub>c</sub> (min) |
|-----------|---------------|----------------|----------|----------------------|--------|----------------------|----------------------|
| Lower     | 22°34′33.5″ N | 102°38′47.4″ W | 2326     | 39.58                | 12.85  | 0.018                | 133.23               |
| Middle    | 22°37′15.3″ N | 102°41′26.7″ W | 2467     | 21.78                | 9.56   | 0.020                | 101.63               |
| Upper     | 22°34′34.4″ N | 102°41′27.5″ W | 2409     | 19.00                | 9.72   | 0.019                | 104.65               |
| Outlet    | 22°38′55.9″ N | 102°39′47.7″ W | 2244     | 20.71                | 11.2   | 0.023                | 108.48               |

### 2.3.1. Blind-Validation Framework and a Priori Parameter Estimation

This study adopts a blind-validation framework in which model parameters for both the SCS-CN and Green-Ampt methods are derived a priori, entirely independently of the observed discharge series recorded at the basin outlet. No parameter was iteratively adjusted or tuned against the reference streamflow data. This design choice serves three complementary purposes: (i) to ensure a methodologically fair comparison between the two infiltration methods under equal conditions of prior information; (ii) to eliminate the risk of overfitting inherent in event-based calibration; and (iii) to produce estimates that are directly transferable to ungauged semi-arid basins, where operational calibration is not feasible.

Parameter estimation followed three parallel workflows:

#### (a) SCS-CN parameters

Curve Number (CN) values for each sub-basin were obtained from the standard tabulations of NRCS-USA [37], based on the combination of land use, hydrologic soil group (HSG), and antecedent moisture condition (AMC II reference condition). The spatial distribution of land use was derived from INEGI Land Use and Vegetation Series [36], and HSG classification was based on the soil textural characterization described in paragraph (b). Sub-basin CN values were computed as area-weighted averages of the CN values assigned to each land



use  $\times$  HSG combination within the sub-basin. The initial abstraction ratio was retained at its standard value ( $\lambda = 0.20$ ), consistent with the original SCS-CN formulation [9], and the mathematical expression  $I_a = \lambda S$  is presented in Section 2.3.2 (Equation (6)).

(b) Green-Ampt parameters

The soil-hydraulic parameters required by the Green-Ampt method—saturated hydraulic conductivity ( $K_s$ ), wetting-front suction head ( $\psi_s$ ), saturated moisture content ( $\theta_s$ ), initial moisture content ( $\theta_i$ ), field capacity ( $\theta_{fc}$ ), and effective porosity ( $\theta_e$ )—were derived a priori from published tabulations [38,39] indexed by soil textural class, as determined below.

Soil texture was determined locally through field sampling: one soil sampling site was established within each of the four sub-basins, and samples were collected at three depths (0–20, 20–40, and 40–60 cm). Textural composition (% sand, % silt, % clay) was measured in the laboratory using the Bouyoucos hydrometer method [40]. For each sub-basin, the parameter values corresponding to the textural class of each depth were determined, and a single representative value per parameter was computed as the arithmetic mean across the three depths.

Two parameters were measured independently to support the a priori parameterization: Saturated hydraulic conductivity ( $K_s$ ) was cross-validated through in situ double-ring infiltrometer tests performed at each sub-basin sampling site, following standard procedures [41]. At the Outlet sub-basin, the field value (16.7 mm/h) differed from the tabulated (29.9 mm/h) and pedotransfer (Equation (13), 24 mm/h) estimates; the field-derived value was therefore adopted in the parameterization.

Field capacity ( $\theta_{fc}$ ) was determined by direct gravimetric analysis on soil samples collected at the same three depths per sub-basin, following standard laboratory procedures [42].

Because continuous soil-moisture monitoring was not available in the study basin during the analyzed period, and to enable a methodologically fair comparison with the SCS-CN framework (which incorporates antecedent-moisture status through the AMC II reference condition), the initial moisture content ( $\theta_i$ ) required by Green-Ampt was set equal to the measured field capacity for each sub-basin ( $\theta_i = \theta_{fc}$ ). This choice represents a deliberate design decision that anchors the initial condition to a locally measured well-defined soil-hydraulic state; removes any systematic advantage or disadvantage that arbitrary  $\theta_i$  choices might confer on Green-Ampt relative to SCS-CN; and ensures that observed performance differences between the two methods reflect their intrinsic formulations rather than an imbalance in initial-condition specification.

The resulting sub-basin-scale Green-Ampt parameter set was thus fully specified prior to the modeling exercise.

(c) Unit hydrograph and routing parameters

The time of concentration ( $T_c$ ), defined as the time required for excess rainfall to travel from the most distant point of the watershed to its outlet, is a fundamental input for semi-distributed hydrological simulation. For each sub-basin,  $T_c$  was estimated using the Kirpich empirical equation [28,43], based on the sub-basin morphometric properties of Table 2: main channel length ( $L$ ), average channel slope ( $S_c$ ), and drainage area ( $A$ ).

$$T_c = 0.06628 \left[ \frac{L^{0.77}}{S_c^{0.385}} \right] \quad (2)$$

where  $T_c$  is the time of concentration (min),  $L$  is the main channel length (km), and  $S_c$  is the average channel slope (m/m).

For the implementation in HEC-HMS,  $T_c$  was converted to lag time through the standard 0.6 conversion factor [37]:

$$\text{Lag Time} = 0.6 \times T_c \quad (3)$$

where *Lag Time* (min) is the delay between the centroid of excess rainfall and the peak of the direct runoff hydrograph.

### 2.3.2. Simulation Platform and Infiltration Formulations

HEC-HMS was selected as the simulation platform, since it incorporates a wide range of methods to represent the rainfall–infiltration–runoff relationship, including the SCS-CN and Green-Ampt loss methods, as well as Clark’s unit hydrograph for direct runoff transformation.

The SCS-CN method is grounded in a water-balance hypothesis, mathematically expressed as

$$Q = \frac{(P - I_a)^2}{(P - I_a) + S} \quad (4)$$

where  $Q$  is the direct runoff depth (mm),  $P$  is the total precipitation (mm) ( $P \geq Q$ ),  $I_a$  is the initial abstraction (mm), and  $S$  is the potential maximum retention of the basin (mm) ( $S \geq F$ ). The retention  $F$  (mm) after the onset of runoff is expressed as

$$F = (P - I_a) - Q \quad (5)$$

The initial abstraction is empirically estimated as a linear fraction of the potential maximum retention:

$$I_a = \lambda \cdot S \quad (6)$$

where  $\lambda$  is the initial abstraction ratio, adopted in this research as  $\lambda = 0.20$  following the original SCS-CN formulation. Substituting Equation (6) into Equation (4) with  $\lambda = 0.20$  yields the operational form of the SCS-CN equation:

$$Q = \frac{(P - 0.2 S)^2}{(P + 0.8 S)} \text{ for } (P > I_a) \quad (7)$$

The retention parameter  $S$  is related to the dimensionless Curve Number through

$$S = \frac{25,400}{CN} - 254 \quad (8)$$

CN values range theoretically from 1 to 100 (HEC-HMS accepts a maximum operational value of 99). Estimating CN depends on multiple factors: soil granulometry, cover type, hydrological soil group, hydrological condition, antecedent moisture conditions (AMC), land use, and tillage practices, among others. Weighted CN values are typically reported under AMC-II reference conditions.

$$CN_{II} = \frac{\sum A_i \times CN_i}{\sum_1^n A_i} \quad (9)$$

where  $CN_{II}$  is the area-weighted Curve Number under AMC-II reference conditions,  $CN_i$  is the CN for spatial subdivision  $i$  (uniform land use and soil type),  $A_i$  is the drainage area

of subdivision  $i$ , and  $A = \sum A_i$  is the total catchment area [18]. Conversion to alternative moisture conditions (AMC-I, dry; AMC-III, wet) follows the standard relationships:

$$CN_I = \frac{CN_{II}}{(0.281 - 0.0128CN_{II})} \quad (10)$$

$$CN_{III} = \frac{CN_{II}}{(0.427 + 0.00573CN_{II})} \quad (11)$$

Lian et al. [44] investigated threshold  $CN$  values for different hydroclimatic regions, comparing field-derived values against those proposed by USDA, and emphasized that further research is required to obtain more accurate  $CN$  estimates for each region.

The Green-Ampt method (GA) is based on Darcy's Law and assumes a saturated upper layer, called the wetting front, above an unwetted zone of initial soil moisture content ( $\theta_i$ ) [45]:

$$F(t) = K_s \cdot t + \psi_s \theta_d \ln \left( 1 + \frac{F(t)}{\psi_s \theta_d} \right) \quad (12)$$

where  $F(t)$  is the cumulative infiltration depth (mm);  $K_s$  is the saturated hydraulic conductivity (mm/h);  $\psi_s$  is the wetting-front suction head (mm; greater in clayey than in sandy soils);  $\theta_d$  is the moisture deficit (dimensionless), defined as the difference between saturated moisture content ( $\theta_s$ ) and initial soil moisture ( $\theta_i$ ); and  $t$  is the time since the onset of ponding ( $h$ ). The equation is valid only under saturated conditions; prior to saturation, the infiltration rate is assumed equal to the rainfall intensity, as implemented in HEC-HMS [5,27].

The a priori estimation of the Green-Ampt parameters used in this research follows the workflow described in Section 2.3.1 (b):  $K_s$ ,  $\psi_s$ ,  $\theta_s$ ,  $\theta_i$ ,  $\theta_{fc}$ , and  $\theta_e$  were derived from tabulated sources indexed by soil textural class, locally supported by Bouyoucos textural analysis, in situ double-ring infiltrometer testing, and gravimetric field-capacity determination.

When direct field measurements of saturated hydraulic conductivity ( $K_s$ ) are unavailable, the empirical relationship based on clay, sand, and organic-matter content [46] can be applied:

$$K_s = 0.056C + 0.016S_a + 0.2311O_m - 0.0693 \quad (13)$$

where  $K_s$  is the saturated hydraulic conductivity of the soil (mm/h),  $C$  is the percentage of clay (%),  $S_a$  is the percentage of sand (%), and  $O_m$  is the organic-matter content of the soil (%). In the present study, field-derived  $K_s$  values from double-ring infiltrometer testing were compared against those obtained from Equation (13) as an independent cross-check.

HEC-HMS offers interoperability with HEC-RAS and ArcGIS, allowing seamless integration between different hydrological and hydraulic analyses, the simulation of complex systems with multiple outlets, the explicit representation of spatial rainfall distribution, and real-time forecasting capabilities. Direct runoff transformation is computed through the Clark unit hydrograph [43,47], based on the linear-reservoir storage equation:

$$S_v = R \cdot Q \quad (14)$$

where  $S_v$  is the storage volume ( $m^3$ ),  $R$  is the storage coefficient of the linear reservoir ( $h$ ), and  $Q$  is the outflow rate ( $m^3/s$ ). Applying continuity over a discrete time step  $\Delta t$  ( $I - Q = \Delta S_v / \Delta t$ ) yields the recursive expression for the outflow at time  $t_i$ :

$$Q_i = c \cdot \frac{I_{i-1} + I_i}{2} + (1 - c) \cdot Q_{i-1}, \text{ with } c = \frac{2 \cdot \Delta t}{2R + \Delta t} \quad (15)$$

where  $I_{i-1}$  and  $I_i$  are the inflow rates at times  $t_{i-1}$  and  $t_i$ ;  $Q_{i-1}$  and  $Q_i$  are the outflow rates at the same times;  $\Delta t$  is the time increment; and  $R$  is the storage coefficient defined above.

### 2.3.3. Sensitivity Analysis of the a Priori Parameterization

Complex hydrological models are subject to inherent limitations associated with the calibration and validation of multiple input parameters [23]. To address parametric uncertainty within the blind-validation framework adopted here (Section 2.3.1), a sensitivity analysis was performed by systematically varying key parameters around their a priori reference values, with a distinct perturbation protocol adopted for each infiltration method to reflect the way its parameters are derived:

For the SCS-CN method, a One-At-a-Time (OAT) perturbation was applied to the Curve Number (CN), which is the single parameter governing runoff generation in this formulation. All other model elements ( $T_c$ ,  $S_c$ , routing parameters) were held at their reference values reported in Table 2.

For the Green-Ampt method, a compound (simultaneous) perturbation was applied jointly to the four soil-hydraulic parameters of the formulation: saturated hydraulic conductivity ( $K_s$ ), wetting-front suction head ( $\psi_s$ ), initial moisture content ( $\theta_i$ , preserving  $\theta_i = \theta_{fc}$ ), and effective porosity ( $\theta_e$ ), at the same relative percentage. This compound design reflects the fact that the four GA parameters are jointly indexed by soil textural class in their tabulated derivation [38]; perturbing them independently would produce combinations that are physically inconsistent with any single textural class. All non-infiltration model elements ( $T_c$ ,  $S_c$ , routing parameters) were held at their reference values.

Three perturbation levels were tested in both methods:  $\pm 6\%$ ,  $\pm 10\%$ , and  $\pm 20\%$  of the reference values. The narrower ranges ( $\pm 6\%$  and  $\pm 10\%$ ) reflect the realistic measurement uncertainty associated with field-derived parameters (Bouyoucos textural analysis, double-ring infiltrometer testing, and gravimetric field-capacity determination, previously described), whereas the  $\pm 20\%$  range provides a broader test of model robustness, consistent with recommendations widely reported in the hydrological modeling literature [27].

The baseline reference values used for all perturbations were the sub-basin-scale CN and GA parameter values derived a priori (Section 2.3.1), consistent with the design decision  $\theta_i = \theta_{fc}$  measured gravimetrically. The sensitivity analysis was applied to all eight analysis windows described in Section 2.2.3, yielding 48 perturbed runs per method in addition to the unperturbed baseline runs. The response of each model to each perturbation was quantified through the mean absolute derivation (MAD, in  $\text{m}^3/\text{s}$ ) of the simulated peak discharge ( $Q_p$ ) with respect to the unperturbed baseline, computed across the eight analysis windows.

### 2.4. Statistical Measures

The agreement between simulated and observed peak discharges ( $Q_p$ ) was assessed through standard goodness-of-fit metrics [26,45], applied to the set of peak discharge values from the eight analysis time windows. Four metrics were computed: the Pearson correlation coefficient ( $R$ ), the root-mean-square error (RMSE), the mean absolute error (MAE), and the Nash–Sutcliffe efficiency (NSE):

$$R = \frac{\sum_{i=1}^n (Q_{s,i} - \overline{Q_s})(Q_{o,i} - \overline{Q_o})}{\sqrt{\sum_{i=1}^n (Q_{s,i} - \overline{Q_s})^2 \cdot \sum_{i=1}^n (Q_{o,i} - \overline{Q_o})^2}} \quad (16)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Q_{o,i} - Q_{s,i})^2} \quad (17)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Q_{o,i} - Q_{s,i}| \quad (18)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{o,i} - Q_{s,i})^2}{\sum_{i=1}^n (Q_{o,i} - \bar{Q}_o)^2} \quad (19)$$

where  $Q_{s,i}$  and  $Q_{o,i}$  are the simulated and observed peak discharges ( $\text{m}^3/\text{s}$ ) for window  $i$  ( $i = 1, \dots, 8$ ), respectively;  $\bar{Q}_s$  and  $\bar{Q}_o$  are the means of the eight simulated and observed peak-discharge values, respectively;  $n = 8$  is the number of analysis windows;  $R$  is the Pearson correlation coefficient (dimensionless, range  $-1$  to  $1$ );  $RMSE$  is the root-mean-square error ( $\text{m}^3/\text{s}$ );  $MAE$  is the mean absolute error ( $\text{m}^3/\text{s}$ ); and  $NSE$  is the Nash–Sutcliffe efficiency (dimensionless, range  $-\infty$  to  $1$ ).

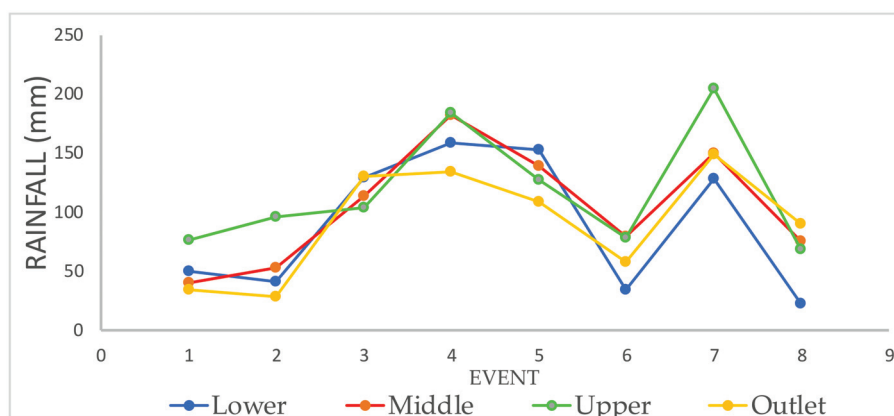
These metrics were computed only for the assessment of the a priori-parameterized models against the observed discharge series; no parameter of either SCS-CN or Green-Ampt was adjusted to optimize any of these metrics, consistent with the blind-validation framework declared.

Following the widely adopted criteria of Moriasi et al. [48], model performance was classified as very good ( $NSE > 0.75$ ), good ( $0.65 < NSE \leq 0.75$ ), satisfactory ( $0.50 < NSE \leq 0.65$ ), or unsatisfactory ( $NSE \leq 0.50$ ).

### 3. Results

#### 3.1. Determination of a Priori Input Parameters

Simulations of the eight analysis time windows (Table 1) were performed in HEC-HMS during the 2020–2025 rainy seasons (Figure 4). Considering the spatio-temporal variability of rainfall, models were developed for four sub-basins (Lower, Middle, Upper, and Outlet), where pluviographic monitoring was available. The SCS-CN and Green-Ampt (GA) methods were applied to estimate the rainfall–infiltration–runoff relationship, and a Clark unit hydrograph was used to transform rainfall excess into direct runoff at the basin outlet. Table 3 reports the particle-size distribution and soil texture classes at three depths within each sub-basin.



**Figure 4.** Spatio-temporal behavior of rainfall in the “Chilitas” basin during the 2020–2025 rainy seasons.

In the upper 0–20 cm horizon, sandy fractions predominate, favoring infiltration and reducing surface runoff, particularly in Kastanozem, Phaeozem, and Leptosol soils, whose permeability allows rapid percolation. In the Phaeozem/Kastanozem complex, the 0–20 cm horizon presents a loam-type texture with clay, sand, and silt fractions. At depths of 20–60 cm, clay content prevails, reducing infiltration rates and, depending on the intensity, generating greater surface runoff. The mobility of water towards the subsoil

therefore depends on the antecedent moisture state of the soil (between dry and saturated conditions); accordingly, soil composition and texture jointly control whether rainwater is retained in the upper layers, gradually percolates toward groundwater, or contributes to surface runoff and associated soil erosion.

**Table 3.** Particle-size distribution and USDA textural classification of soils sampled at three depths within each sub-basin. Percentages measured by Bouyoucos hydrometer method.

| Sub-Basin | Depth (cm) | Percentages by Soil Particle (%) |          |           | Textural Class  |
|-----------|------------|----------------------------------|----------|-----------|-----------------|
|           |            | Sand (S)                         | Clay (C) | Silt (Si) |                 |
| Lower     | 0–20       | 60.8                             | 21.4     | 17.8      | Sandy clay loam |
|           | 20–40      | 29.3                             | 59.4     | 11.3      | Clay            |
|           | 40–60      | 36.0                             | 57.4     | 6.6       | Clay            |
| Middle    | 0–20       | 68.2                             | 14.5     | 17.3      | Sandy loam      |
|           | 20–40      | 64.2                             | 16.5     | 19.3      | Sandy loam      |
|           | 40–60      | 64.8                             | 20.0     | 15.3      | Sandy loam      |
| Upper     | 0–20       | 40.2                             | 21.2     | 38.6      | Loam            |
|           | 20–40      | 20.9                             | 70.5     | 8.6       | Clay            |
|           | 40–60      | 43.5                             | 40.0     | 16.6      | Clay loam       |
| Outlet    | 0–20       | 82.4                             | 13.1     | 4.6       | Loamy sand      |
|           | 20–40      | 88.8                             | 8.0      | 3.3       | Sand            |
|           | 40–60      | 82.8                             | 6.7      | 10.6      | Loamy sand      |

Based on the clay, sand, and silt contents reported in Table 3, hydrologic soil groups (HSG) A, B, C, and D were assigned to the sub-basins, with corresponding infiltration rate ranges of >7.62, 3.81, 1.27, and <1.27 cm/h, respectively, following USDA criteria [37,49]. Field infiltration tests yielded the following saturated hydraulic conductivity ( $K_s$ ) value for the Leptosol soil at the Outlet sub-basin: 16.7 mm/h. For comparison, the Green-Ampt reference value proposed in the literature [37,49] for the corresponding USDA textural class is 29.9 mm/h, while Equation (13) yields an estimated 24 mm/h. These differences can be attributed to soil compaction, structural heterogeneity, and antecedent moisture contents; accordingly, the present research recommends prioritizing field-derived measurements when available. Table 4 reports the a priori-derived Curve Number (CN) and GA parameter sets per sub-basin and soil type.

The variation of CN across sub-basins reflects the combined influence of soil type, land use, predominant vegetation, and hydrodynamic properties, in agreement with previous reports [5,25]. Although a single area-weighted CN value of 83 could be assumed for the entire basin, this approach implicitly assumes a uniform hydrological response and yields averaged infiltration and peak-flow estimates that overlook the spatio-temporal heterogeneity of rainfall; comparable observations have been documented in [14,50].

**Table 4.** A priori-derived parameters for the SCS-CN and Green-Ampt methods per sub-basin and predominant soil type.

| Sub-Basin | Soil Type           | SCS-CN |                                                   | GA                                             |                                                |               |              |
|-----------|---------------------|--------|---------------------------------------------------|------------------------------------------------|------------------------------------------------|---------------|--------------|
|           |                     | CN     | $\theta_{fc}$ (cm <sup>3</sup> /cm <sup>3</sup> ) | $\theta_e$ (cm <sup>3</sup> /cm <sup>3</sup> ) | $\theta_s$ (cm <sup>3</sup> /cm <sup>3</sup> ) | $\psi_s$ (mm) | $K_s$ (mm/h) |
| Lower     | Kastanozem          | 79     | 0.354                                             | 0.370                                          | 0.449                                          | 324.1         | 0.7          |
| Middle    | Phaeozem            | 85     | 0.254                                             | 0.407                                          | 0.443                                          | 51.3          | 10.9         |
| Upper     | Phaeozem/Kastanozem | 90     | 0.324                                             | 0.412                                          | 0.472                                          | 30.7          | 26.0         |
| Outlet    | Leptosol            | 80     | 0.180                                             | 0.424                                          | 0.451                                          | 74.5          | 16.7         |



### 3.2. HEC-HMS Hydrological Simulations

Using the parameters of Table 4 and the geomorphological characterization of Table 2, simulations were performed in HEC-HMS to determine the rainfall–infiltration–runoff interaction and to estimate discharge for the eight analysis time windows (Table 1) recorded by the field instrumentation during the 2020–2025 rainy season (Figure 4).

For the analyzed period, the monthly spatio-temporal variability of rainfall ranged from 22.8 to 204.9 mm between sub-basins separated by approximately 5 km. In small basins, hydrological practice usually relies on a single gauge located within a 13 km radius. The present results demonstrate that this assumption is not valid for the “Chilitas” basin, since rainfall variation within distances as short as 5 km is significant, leading to differentiated infiltration and runoff volumes for the same storm analysis time window.

The reviewed scientific literature reports the Thiessen polygon method as a common approach to estimate the rainfall contributed by each gauge, given that it accounts for spatial gauge distribution and topographic relief [51]; however, when the inter-gauge distance is large, the method may introduce non-negligible interpolation errors. Limited hydroclimatological data thus constrain rainfall and water-availability estimations, with downstream implications for agricultural, industrial, and urban water management. The instrumented monitoring proposed in the present research contributes to mitigating this limitation. Table 5 summarizes the rainfall recorded per sub-basin, the simulated peak flows obtained with SCS-CN and Green-Ampt (GA) in HEC-HMS, and the discharges observed with the Levelogger at the basin outlet for eight analysis time windows.

**Table 5.** Recorded rainfall per sub-basin, simulated peak discharge by method (SCS-CN and Green-Ampt), and observed peak discharge at the basin outlet for the eight analysis time windows.

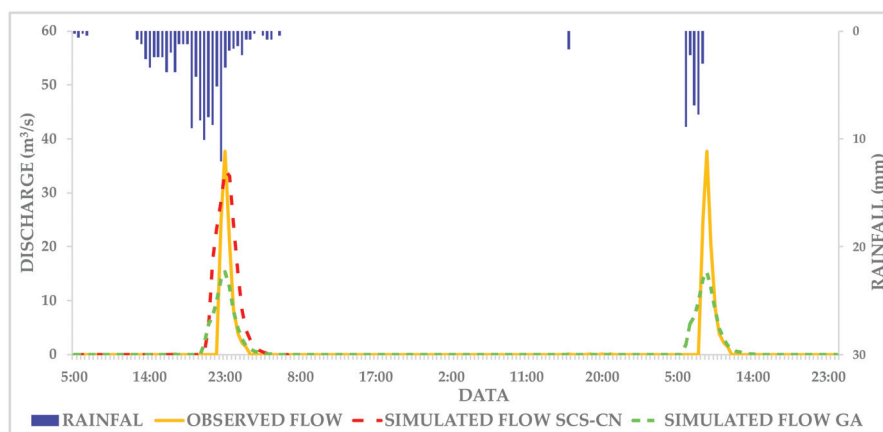
| Window | Data Event           | Sub-Basin Rainfall (mm) |        |       |        | Peak Discharge (m <sup>3</sup> /s) |       |          |
|--------|----------------------|-------------------------|--------|-------|--------|------------------------------------|-------|----------|
|        |                      | Lower                   | Middle | Upper | Outlet | SCS-CN                             | GA    | Observed |
| 1      | 29–30 August 2020    | 9.1                     | 13.0   | 26.8  | 14.7   | 13.6                               | 0.2   | 4.9      |
| 2      | 19–22 June 2021      | 31.8                    | 43.5   | 35.9  | 28.5   | 34.2                               | 20.1  | 37.76    |
| 3      | 14–17 September 2021 | 54                      | 75.2   | 60.6  | 50.7   | 92.0                               | 64.0  | 90.0     |
| 4      | 26–30 August 2021    | 51.2                    | 36.0   | 26.0  | 20.1   | 50.8                               | 95.7  | 48.0     |
| 5      | 30–31 August 2023    | 12.50                   | 29.80  | 21.0  | 31.0   | 29.6                               | 7.8   | 65.0     |
| 6      | 23–26 July 2024      | 50.4                    | 67.8   | 88.6  | 77.5   | 162.1                              | 28.5  | 125.0    |
| 7      | 21–30 June 2025      | 71.6                    | 75.2   | 90.6  | 68.8   | 193.2                              | 212.3 | 160.0    |
| 8      | 16–17 July 2025      | 15.8                    | 27.4   | 13.0  | 60.1   | 66.2                               | 46.1  | 44.82    |

Window 1 corresponds to the continuous rainfall episode recorded from 29 August 2020, 16:00 to 30 August 2020, 22:30 (duration 30.50 h; Table 1), with a 17.7 mm variation in rainfall depth between sub-basins (Table 5). The HEC-HMS simulation yielded sub-basin differentiated infiltration depths of 9.1, 13.0, 26.8, and 14.7 mm (Lower, Middle, Upper, and Outlet, respectively), generating an outlet peak discharge of 13.6 m<sup>3</sup>/s under SCS-CN. This demonstrates the spatio-temporal variation of rainfall and the differentiated runoff contribution by sub-basin [52,53], highlighting the importance of instrumentation for monitoring rainfall and recording its spatio-temporal variation. The corresponding Green-Ampt simulation produced an outlet peak discharge of only 0.2 m<sup>3</sup>/s, substantially underestimating the 4.9 m<sup>3</sup>/s recorded by the Levelogger. In contrast, the SCS-CN estimate of 13.6 m<sup>3</sup>/s overestimated the observation by a factor of 2.8. This large discrepancy in Window 1, characterized by low rainfall depth (9.1–26.8 mm across sub-basins), illustrates the higher sensitivity of GA to the a priori-specified initial soil-moisture conditions under low-intensity rainfall events, consistent with previous reports [24].

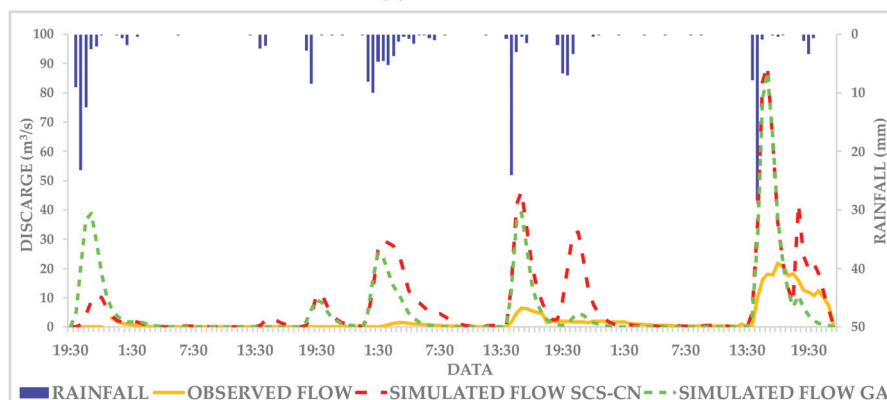


For Window 2 (2021), continuous rainfall produced an outlet hydrograph (Figure 5a) qualitatively similar to those derived from synthetic hydrographs, with an observed peak discharge of  $37.76 \text{ m}^3/\text{s}$ , a fast-rising limb (approximately 60 min) and a longer, smoother recession (approximately 4 h). This temporal asymmetry highlights the role of rainfall intensity in controlling both the magnitude and timing of peak discharge, even for short-duration storms. These findings reinforce the importance of considering the spatio-temporal variability of rainfall in hydrological studies, since averaged rainfall may overestimate runoff volumes and bias water demand planning. Figure 5a–e show the temporal behavior of recorded rainfall, simulated SCS-CN and GA flows, and Levelogger-observed discharges for Windows 2, 3, 6, 7, and 8.

Figure 6a–d present the simulated infiltrated and effective rainfall, together with the SCS-CN runoff response, for Window 3 (recorded rainfall 54.0, 75.2, 71.31, and 50.7 mm per sub-basin: Lower, Middle, Upper, and Outlet, respectively). Under spatio-temporally variable rainfall, the simulated infiltrated depths were 38.8, 35.7, 25.4, and 36.5 mm, with effective (excess) rainfall of 15.2, 39.3, 45.9, and 14.2 mm, respectively. Maximum runoff occurred in the sub-basin where Phaeozem/Kastanozem clayey soils predominate (Upper), generating a differentiated runoff volume of  $600\,824$ ;  $856\,825$ ;  $872\,480$ ; and  $294\,703 \text{ m}^3$  across sub-basins (Lower, Middle, Upper, and Outlet, respectively). Significant subsoil contributions occurred in the highest-elevation sub-basin through channel-bed infiltration along its dense drainage network rather than through the low-permeability clayey matrix, supporting its identification as a channel-mediated recharge zone and in agreement with previous reports [9].

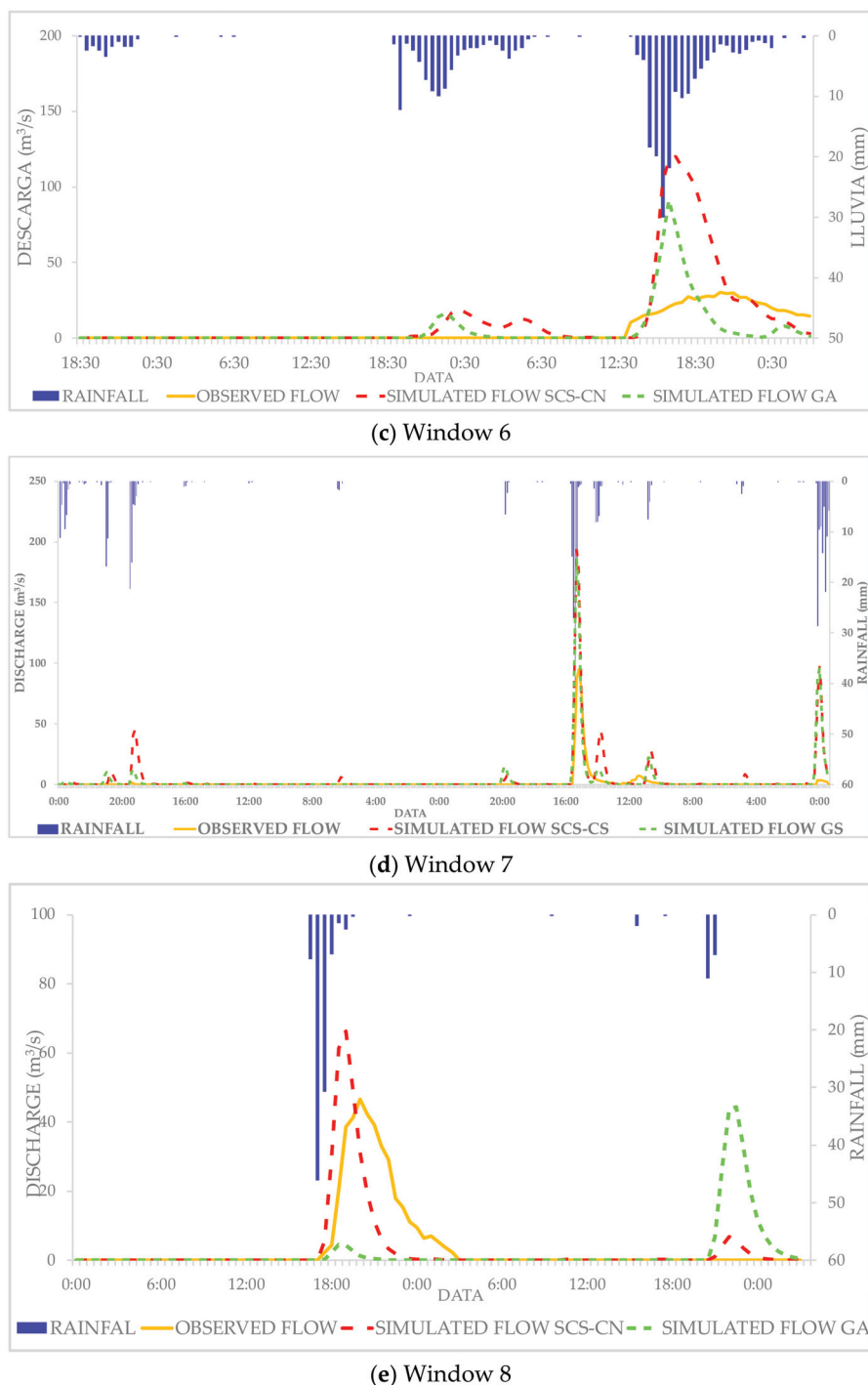


(a) Window 2



(b) Window 3

Figure 5. Cont.



**Figure 5.** Infiltration behavior, effective rainfall, and simulated and observed runoff for Windows 2, 3, 6, 7, and 8.

Table 6 reports the recorded, infiltrated, and effective rainfall depths for Windows 4–8, together with the SCS-CN simulated flows per sub-basin and the simulated and observed flows at the basin outlet. The spatio-temporal variation of rainfall for the different reported windows is observed, along with the differentiation of both surface and subsurface effective rainfall. This demonstrates the differentiated behavior of surface and subsurface runoff (aquifer recharge) in each sub-basin. If the calculations had been performed using an average value for the entire basin, much higher values would have been obtained, reporting larger water volumes that would not have occurred and that would have biased the

estimation of water demand for the population, potentially resulting in a supply deficit. This underscores the importance of the spatially distributed data and findings reported in the present research.

The outflow at the control point for the different windows results from a differentiated contribution per sub-basin; the sub-basin that contributes the most to the peak-flow response is the one with clay-dominated soils and a dense drainage network, attributable to the low hydraulic conductivity of the clayey layers and the high drainage-network density [8,23,44,45,51].

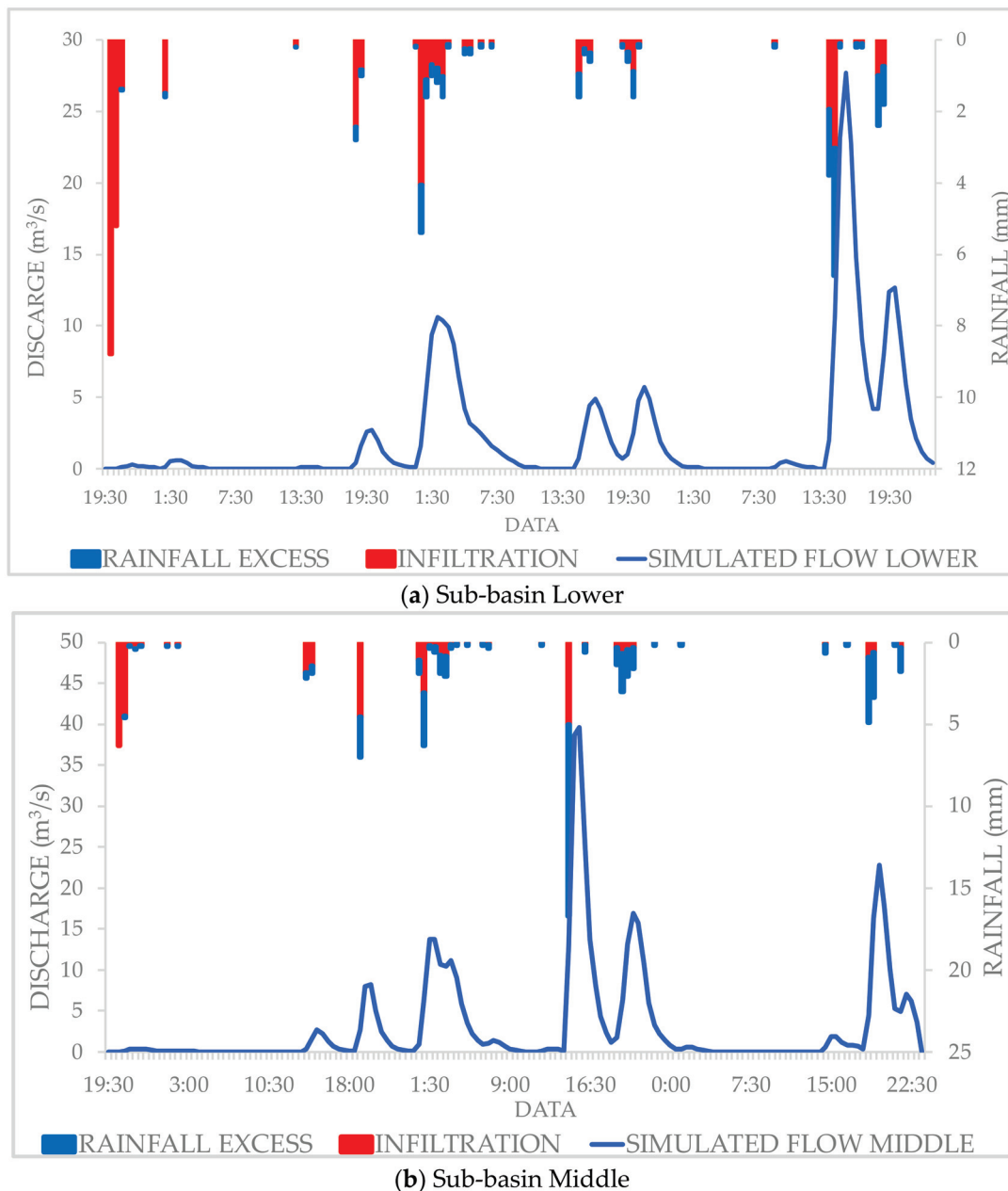
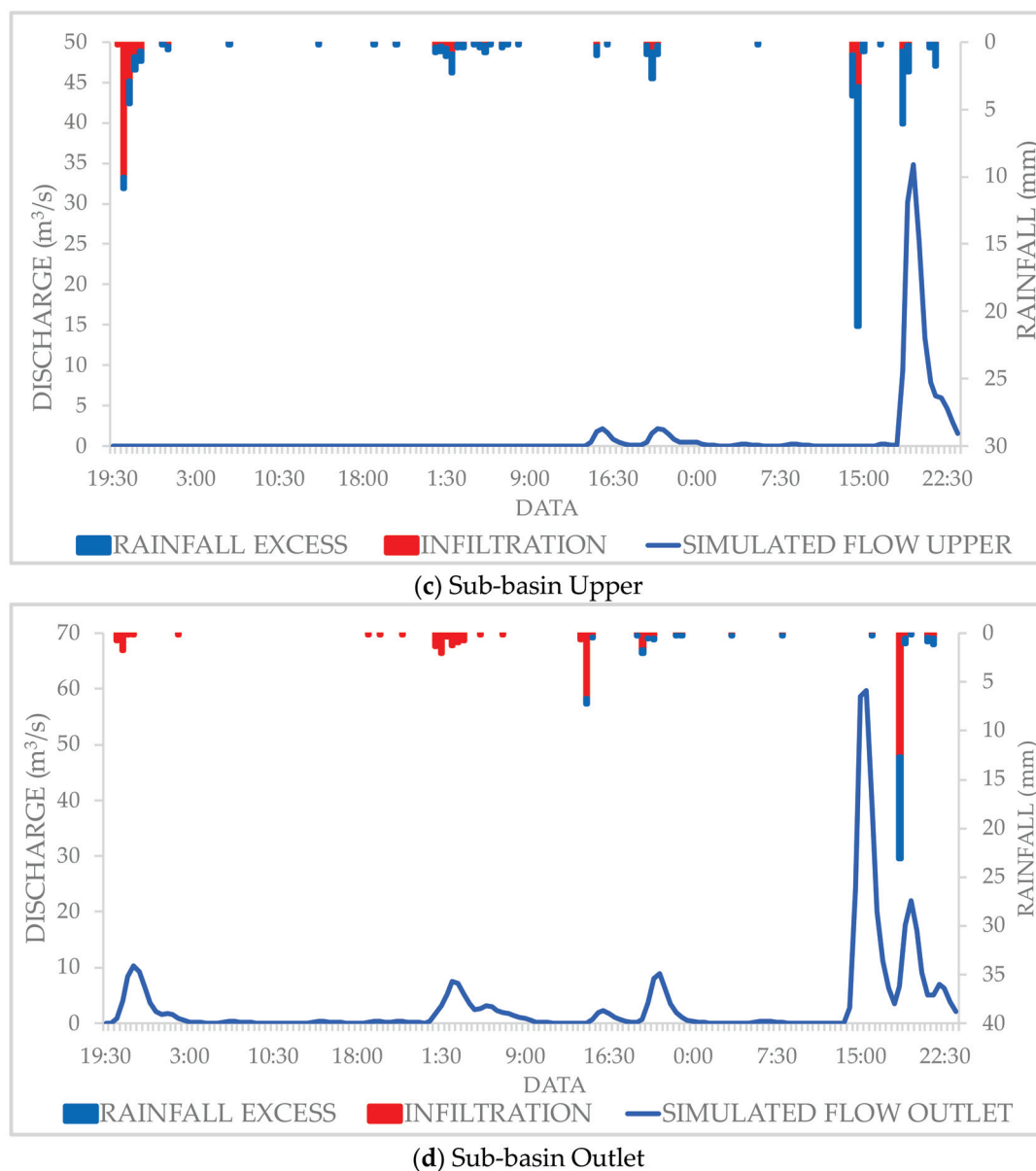


Figure 6. Cont.



**Figure 6.** Infiltrated rainfall, effective rainfall, and estimated flow rate by SCS-CN for Window 3.

**Table 6.** Recorded, infiltrated, and effective rainfall depths per sub-basin, together with SCS-CN simulated peak discharges per sub-basin and observed outflow at the basin outlet, for Windows 4–8.

| Window | Rainfall (mm) |      |      |             |       |       |           |       |       |       |        |       | Peak Discharge (m³/s) |         |       |      |       |
|--------|---------------|------|------|-------------|-------|-------|-----------|-------|-------|-------|--------|-------|-----------------------|---------|-------|------|-------|
|        | Registered    |      |      | Infiltrated |       |       | Effective |       |       | Lower | Middle | Upper | Outlet                | Outflow |       |      |       |
| 4      | 51.2          | 36.0 | 26.0 | 20.1        | 37.69 | 25.83 | 17.47     | 19.33 | 13.51 | 10.17 | 8.53   | 0.77  | 50.8                  | 32.7    | 12.2  | 0.8  | 50.8  |
| 5      | 12.5          | 29.8 | 21.0 | 31.0        | 12.50 | 23.19 | 15.59     | 26.91 | 0.00  | 6.61  | 5.41   | 4.09  | 0.0                   | 13.5    | 12.2  | 8.5  | 29.6  |
| 6      | 50.4          | 67.8 | 88.6 | 77.5        | 37.36 | 34.41 | 26.7      | 44.77 | 13.04 | 33.39 | 61.9   | 32.73 | 13.9                  | 34.1    | 58.6  | 25.4 | 162.1 |
| 7      | 71.6          | 75.2 | 90.6 | 68.8        | 44.73 | 35.7  | 26.83     | 42.49 | 26.87 | 39.50 | 63.77  | 26.31 | 61.9                  | 63.5    | 121.2 | 29.0 | 193.2 |
| 8      | 15.8          | 27.4 | 13   | 60.1        | 15.72 | 22.03 | 11.48     | 39.84 | 0.08  | 5.37  | 1.52   | 20.26 | 0.4                   | 9.0     | 3.2   | 54.0 | 66.2  |

### 3.3. Statistical Performance Assessment

The peak discharge ( $Q_p$ ) simulated by SCS-CN and Green-Ampt (GA) at the basin outlet was compared with the corresponding Levellogger records using the Pearson correlation coefficient ( $R$ ), the root-mean-square error ( $RMSE$ ), the mean absolute error ( $MAE$ ),

and the Nash–Sutcliffe efficiency ( $NSE$ ), computed over the set of peak-discharge values from the eight analysis windows. The pairwise comparison between the SCS-CN and GA simulations yielded  $R = 0.70$ , indicating a moderate linear correlation but substantial divergence in absolute peak-discharge values between the two methods (Table 7). Compared with observed discharges, SCS-CN yielded  $R = 0.95$  and  $NSE = 0.76$ , whereas GA yielded  $R = 0.70$  and  $NSE = 0.45$ . According to the criteria of Moriasi et al. [54], SCS-CN falls within the very good performance category ( $NSE > 0.75$ ), whereas GA falls within the unsatisfactory category ( $NSE \leq 0.50$ ). The high  $R$  value obtained for SCS-CN (0.95) corresponds to  $R^2 = 0.90$ , indicating that 90% of the between-window variance in observed peak discharge is explained by the model, while the residual 10% is attributable to the model formulation, parametric uncertainty, and measurement noise. The comparatively lower correlation for GA ( $R = 0.70$ , corresponding to  $R^2 = 0.49$ ) can be attributed to the higher sensitivity of the physically based formulation to the a priori-specified initial soil-moisture conditions ( $\theta_i = \theta_{fc}$ ), as further explored in the sensitivity analysis. Table 7 summarizes the statistical metrics for three pairwise comparisons.

**Table 7.** Goodness-of-fit metrics for the three pairwise comparisons between observed and simulated discharge series at the basin outlet, computed over all eight analysis time windows.

| Metric                         | Observed vs. SCS-CN | Observed vs. GA | SCS-CN vs. GA |
|--------------------------------|---------------------|-----------------|---------------|
| RMSE ( $\text{m}^3/\text{s}$ ) | 23.15               | 48.23           | 52.76         |
| MAE ( $\text{m}^3/\text{s}$ )  | 18.02               | 37.92           | 36.88         |
| R                              | 0.95                | 0.70            | 0.70          |
| NSE                            | 0.76                | 0.45            | 0.25          |

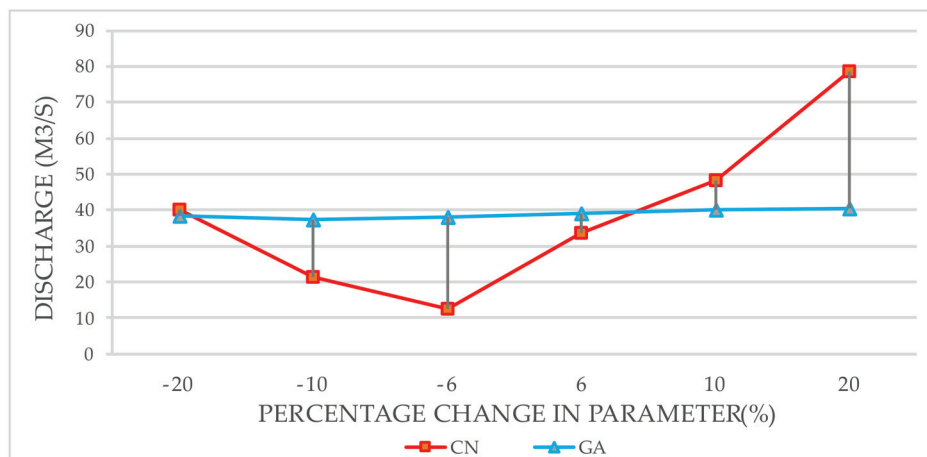
According to the criteria of Moriasi et al. [54], a Pearson correlation exceeding 0.7 is generally accepted as indicative of good agreement; both SCS-CN ( $R = 0.95$ ) and GA ( $R = 0.70$ ) meet this condition in the present research, in agreement with reports in similar semi-arid contexts. Regarding the  $NSE$  coefficient, values above 0.75 are classified as very good, values in the range 0.65–0.75 as good, values in the range 0.50–0.65 as satisfactory, and values below 0.50 as unsatisfactory [54]. The comparison between SCS-CN and GA versus observed yields  $NSE$  of 0.76 and 0.45, respectively, placing SCS-CN in the very good category and Green-Ampt in the unsatisfactory category. The pairwise SCS-CN versus GA comparison reports  $R = 0.70$  and  $NSE = 0.25$ , indicating a moderate linear correlation but substantial divergence in absolute peak-discharge values between the two methods; the mean absolute deviation between the SCS-CN and GA simulated series across the eight windows was  $36.9 \text{ m}^3/\text{s}$  (Table 7). The  $MAE$  values obtained in the present research ( $18.02\text{--}37.92 \text{ m}^3/\text{s}$ ) fall within the range of previously reported studies in the literature.

### 3.4. Sensitivity Analysis

Under the blind-validation framework adopted in this study, the robustness of the two infiltration formulations against parametric uncertainty was assessed through the sensitivity-analysis protocol (Section 2.3.3)—an OAT perturbation of the Curve Number (CN) for SCS-CN—and a compound (simultaneous) perturbation of the four soil-hydraulic parameters ( $K_s$ ,  $\psi_s$ ,  $\theta_i = \theta_{fc}$ ,  $\theta_e$ ) for Green-Ampt, both at  $\pm 6\%$ ,  $\pm 10\%$ , and  $\pm 20\%$  of the a priori reference values.

Figure 7 shows the behavior of the simulated peak discharge ( $Q_p$ ) to parameter perturbations for each perturbation level, averaged across the eight analysis time windows and expressed as the mean absolute deviation ( $\text{m}^3/\text{s}$ ) between simulated and observed peak-discharge values. These perturbation ranges were selected to reflect the realistic mea-

surement uncertainty of field-derived parameters ( $\pm 6\%$  and  $\pm 10\%$ ) [27] and the broader robustness range widely recommended in the hydrological modeling literature ( $\pm 20\%$ ).



**Figure 7.** Mean absolute deviation ( $\text{m}^3/\text{s}$ ) between simulated and observed peak discharge across the eight analysis time windows, versus parameter perturbation levels ( $\pm 6\%$ ,  $\pm 10\%$  and  $\pm 20\%$ ) for SCS-CN and Green-Ampt.

The SCS-CN model exhibited a pronounced directional and monotonic sensitivity of the simulated peak discharge ( $Q_p$ ) to the perturbation of the Curve Number (Figure 7). Positive CN perturbations progressively amplified the mean absolute deviation from the observed peak discharge, reaching approximately  $33.7 \text{ m}^3/\text{s}$ ,  $48.2 \text{ m}^3/\text{s}$ , and  $78.5 \text{ m}^3/\text{s}$  at  $+6\%$ ,  $+10\%$ , and  $+20\%$ , respectively. Negative perturbations yielded  $12.6 \text{ m}^3/\text{s}$  at  $-6\%$ ,  $21.5 \text{ m}^3/\text{s}$  at  $-10\%$ , and  $40.2 \text{ m}^3/\text{s}$  at  $-20\%$ ; the minimum deviation over the whole perturbation range therefore occurred at  $-6\%$  rather than at the unperturbed baseline, indicating a slight offset of the a priori CN above the empirical optimum for this basin. This behavior is consistent with the mathematical structure of the SCS-CN formulation, in which the potential maximum retention  $S$  decreases as CN increases, thereby amplifying the runoff response for a given rainfall depth. The minimum-deviation point at  $-6\%$  suggests that the a priori CN values (Table 4) are slightly above the empirical optimum for this basin; nevertheless, the unperturbed baseline mean deviation ( $MAE = 18.02 \text{ m}^3/\text{s}$ ) remains in the very-good performance range ( $NSE = 0.76$ , Table 7), demonstrating the operational suitability of the a priori estimation approach.

The Green-Ampt model exhibited a markedly different response pattern (Figure 7). The simulated  $Q_p$  mean absolute deviation from the observed values remained nearly constant, in the narrow range of  $37.3 \text{ m}^3/\text{s}$  to  $40.5 \text{ m}^3/\text{s}$  across all six perturbation levels, with mean deviations of  $38.1 \text{ m}^3/\text{s}$ ;  $37.3 \text{ m}^3/\text{s}$ ;  $38.5 \text{ m}^3/\text{s}$  at  $-6\%$ ,  $-10\%$ , and  $-20\%$ ; and  $39.0 \text{ m}^3/\text{s}$ ,  $40.1 \text{ m}^3/\text{s}$ , and  $40.5 \text{ m}^3/\text{s}$  at  $+6\%$ ,  $+10\%$ , and  $+20\%$ , respectively. The compound perturbation was designed to preserve the physical consistency of the parameter set with a single textural class, reflecting the joint indexed derivation of the four GA parameters from tabulated sources [38].

The near-invariance of the GA response (all values clustered around the unperturbed baseline MAE of  $37.92 \text{ m}^3/\text{s}$  reported in Table 7) across the six perturbation levels indicates that the systematic underperformance of Green-Ampt observed in the pairwise comparison against the Levellogger records ( $NSE = 0.45$ ,  $R = 0.70$ ) is not correctable by parametric adjustments within the  $\pm 20\%$  range. This behavior suggests that the deviation of GA from the observed peak discharge is dominated by: (i) the physical assumption of a homogeneous saturated wetting front in soils with pronounced textural stratification (Table 3); (ii) the



a priori initial-moisture specification ( $\theta_i = \theta_{fc}$ ), which may be inadequate for the range of antecedent conditions occurring across the eight windows; and (iii) the tabulated derivation of the four GA parameters from soil textural class alone [38], which cannot fully capture the influence of soil structure, compaction, and macroporosity in the study basin.

In relative terms, the response of SCS-CN to the perturbation of the Curve Number was substantially greater than that of Green-Ampt to its compound soil-hydraulic parameterization (amplitude of mean deviation across the six perturbation levels: 65.9 m<sup>3</sup>/s for SCS-CN vs. 3.2 m<sup>3</sup>/s for Green-Ampt). This finding supports two complementary interpretations: the SCS-CN method is more directly customizable through its single Curve Number parameter, offering a clear operational lever for basin-specific adaptation; and Green-Ampt exhibits a structural rigidity in its response that limits its parametric adaptability within the range of a priori-derivable uncertainty, particularly under the blind-validation framework adopted here, a finding relevant for the transferability of both methods to ungauged semi-arid basins.

## 4. Discussion

### 4.1. Spatio-Temporal Rainfall Variability in Semi-Arid Instrumented Basins

This research examined the rainfall–infiltration–runoff relationship in a semi-arid watershed, combining HEC-HMS simulation with field-observed data from an instrumented basin. The instrumentation of the basin (four rain-gauge stations with a mean inter-gauge distance of approximately 5 km and one hydrometric station equipped with a Levellogger at the outlet) allowed the spatio-temporal variability of rainfall to be resolved at the sub-basin scale, a level of detail rarely reported in semi-arid basins of Mexico. This level of spatial resolution is central to understanding water availability and quantifying the volumes required to meet demand, a prerequisite for sustainable water-resource management.

The monthly rainfall variability of 22.8–204.9 mm observed between sub-basins separated by approximately 5 km (Figure 4) confirms that the assumption of spatial uniformity of rainfall, still common in lumped hydrological models, is not tenable in semi-arid basins where convective storms dominate. This finding is consistent with previous studies in semi-arid watersheds [29,44,55] and with Derdour et al. [56], who reported similar rainfall heterogeneity in a semi-arid basin analyzed with HEC-HMS, as well as with Lian et al. [44], who described the spatial patterns of rainfall and their hydrological consequences under comparable climatic regimes.

The novelty of the present research lies in the explicit sub-basin-scale coupling between spatially distributed rainfall observations recorded at four independent gauges under a one-to-one gauge–sub-basin assignment scheme, and a semi-distributed HEC-HMS implementation with sub-basin-specific parameterization of both infiltration methods. This design enables direct attribution of the observed peak-discharge variability to the spatial heterogeneity of rainfall inputs, rather than to lumped-basin assumptions that would otherwise mask the sub-basin-scale processes documented here.

### 4.2. Comparative Performance of SCS-CN and Green-Ampt Within HEC-HMS

The comparative assessment of SCS-CN and Green-Ampt (GA) yielded asymmetric statistical performances against the Levellogger records (Table 7): SCS-CN reached NSE = 0.76 and R = 0.95, falling within the very good performance category of Moriasi et al. [54], whereas Green-Ampt reached NSE = 0.45 and R = 0.70, falling within the unsatisfactory category of the same classification. Although the initial working hypothesis that the physically based GA would outperform SCS-CN under spatio-temporally variable rainfall was based on the theoretical richness of the Green-Ampt formulation, the results indicate



that at the sub-basin scale and for the eight analysis time windows analyzed, SCS-CN provides substantially higher correlation and efficiency with observed discharges, and the two simulated discharge series differ by a mean absolute deviation of  $36.9 \text{ m}^3/\text{s}$  (Table 7).

Three complementary explanations account for this result:

1. The GA model is highly sensitive to the initial soil-moisture content ( $\theta_i$ ), which in semi-arid environments is subject to rapid temporal changes between consecutive storm events. In the present research,  $\theta_i$  was fixed at the sub-basin-scale field capacity  $\theta_{fc}$ , a decision motivated by the blind-validation framework and the absence of continuous soil-moisture monitoring, but this uniform initial condition cannot capture the actual antecedent moisture state, which varies from window to window. Even when GA parameters are derived from careful laboratory characterization, uncertainty in  $\theta_i$  can dominate the GA error [1,46].
2. The SCS-CN method, despite its empirical nature, implicitly captures some of the basin-scale hydrological response through the Curve Number weighting procedure [29,44], which in this case was performed at the sub-basin level, partially compensating for its lack of explicit physical basis.
3. The GA equation is strictly valid under saturated conditions [31], whereas several of the analyzed windows involved low-depth, low-intensity rainfall that did not produce sustained saturation of the soil profile (Windows 1, 5, and 8 in particular), a regime under which the physical assumptions of GA are only partially met, thereby limiting its operational advantage in the present setting.

The sensitivity analysis (Figure 7) provides a complementary structural explanation. SCS-CN was highly sensitive to Curve Number perturbations, with the mean absolute deviation from observations varying across a  $65.9 \text{ m}^3/\text{s}$  amplitude between  $-20\%$  and  $+20\%$ ; this sensitivity is directional and monotonic and offers a clear operational lever for basin-specific adaptation. In contrast, Green-Ampt was nearly insensitive to compound perturbations of its four soil-hydraulic parameters within the same  $\pm 20\%$  range (amplitude  $3.2 \text{ m}^3/\text{s}$ ), and its deviation from observations remained anchored around  $38 \text{ m}^3/\text{s}$  regardless of parameter adjustment. This near-invariance indicates that the systematic underperformance of GA in the present study is not attributable to inaccurate parameter values within reasonable a priori uncertainty but rather to structural limitations of the GA formulation under blind validation, including the assumption of a homogeneous saturated wetting front in soils with pronounced textural stratification (Table 3).

These findings do not disqualify Green-Ampt as a methodologically richer framework for prolonged storms in homogeneous soils or for applications in which continuous soil-moisture monitoring supports a data-informed  $\theta_i$  specification and iterative calibration is feasible; its operational superiority over SCS-CN depends critically on data availability and on storm characteristics. Comparable patterns have been reported by Belina et al. [27] who documented that, in semi-arid basins with sparse moisture monitoring, the physically based Green-Ampt method does not necessarily outperform the empirical SCS-CN under limited-data conditions.

#### 4.3. Curve Number Selection, Sensitivity, and Parameter Uncertainty

The selection of Curve Number values is one of the most decisive parametric choices for the SCS-CN method [44,57]. The results of the sensitivity analysis (Figure 7) confirm that the CN is a highly responsive parameter: perturbations of  $\pm 6\%$ ,  $\pm 10\%$ , and  $\pm 20\%$  around the a priori reference values (Table 4) produced a broad range of variations in the mean absolute deviation between simulated and observed peak discharge, from  $12.6 \text{ m}^3/\text{s}$  (at  $-6\%$ ) to  $78.5 \text{ m}^3/\text{s}$  (at  $+20\%$ ), averaged across all eight analysis time windows.

This high sensitivity reinforces the recommendation of Lian et al. [44] and Segura-Peña et al. [25], who advocate the assignment of differentiated CN values according to hydrological soil group, land use, and antecedent moisture condition rather than a single area-weighted basin value. In the present research, applying a single area-weighted CN of 83 would have masked the differentiated runoff response of each sub-basin, leading to systematic biases in water-availability estimates and hydrograph timing.

In contrast, the sensitivity analysis of Green-Ampt revealed a markedly different pattern (Figure 7): the mean absolute deviation from observed peak discharge remained nearly constant in the 37.3–40.5 m<sup>3</sup>/s range across all six perturbation levels ( $\pm 6\%$ ,  $\pm 10\%$ ,  $\pm 20\%$ ) of the compound soil-hydraulic parameter set ( $K_s$ ,  $\psi_s$ ,  $\theta_i$ ,  $\theta_e$ ). This near-invariance indicates that, while the initial moisture content  $\theta_i$  is nominally the parameter to which GA is most sensitive in theoretical analyses of individual events [25], the systematic underperformance of GA observed at the basin scale under blind-validation conditions is not correctable by parametric adjustments within a reasonable a priori uncertainty envelope.

Regarding GA, previous studies in semi-arid basins have identified  $\theta_i$  as the most sensitive individual parameter of the Green-Ampt formulation [58]. However, the compound-perturbation results reported here suggest that, within the constrained a priori parameterization of the four GA parameters indexed by soil textural class [38], the operational uncertainty of GA is dominated by structural factors, including the assumption of a homogeneous saturated wetting front in soils with pronounced textural stratification (Table 3) and the uniform specification  $\theta_i = \theta_{fc}$  across windows with different antecedent conditions, rather than by residual parametric imprecision.

The practical implication is that, in basins where high-frequency soil-moisture monitoring is unavailable, SCS-CN offers a more robust operational performance than GA, despite its empirical nature, and the sub-basin-specific weighting of the Curve Number (rather than a single area-weighted basin value) is a critical design choice for capturing the spatio-temporal heterogeneity of runoff generation in semi-arid catchments.

#### 4.4. Statistical Performance and Reproducibility

The Pearson correlation coefficient ( $R$ ) ranges from  $-1$  to  $1$ , with values  $> 0.7$  typically considered acceptable in hydrological modeling [45,49]. The values obtained in the present research (0.70–0.95, Table 7) meet this threshold for both infiltration methods and are comparable with those reported by Yilma and Kebede [5] and EL-Hames [57] for HEC-HMS applications in semi-arid basins, although the asymmetry between SCS-CN ( $R = 0.95$ ) and Green-Ampt ( $R = 0.70$ ) is more pronounced than in most previously reported comparative studies. The Nash–Sutcliffe efficiency ( $NSE$ ), which ranges from  $-\infty$  to  $1.0$  [13,45], yielded values of 0.76 for SCS-CN and 0.45 for Green-Ampt against the Levelogger records (Table 7), classifying SCS-CN in the very good ( $NSE > 0.75$ ) performance category and Green-Ampt in the unsatisfactory ( $NSE \leq 0.50$ ) category of the Moriasi et al. criteria [54].

The mean absolute error ( $MAE$ ) ranged from 18.02 (SCS-CN) to 37.92 (Green-Ampt) m<sup>3</sup>/s against the Levelogger records, with an inter-model  $MAE$  of 36.88 m<sup>3</sup>/s between the SCS-CN and Green-Ampt simulations, within the range of previously reported values for HEC-HMS in semi-arid basins [27]. The lower  $MAE$  achieved by SCS-CN (18.02 m<sup>3</sup>/s) supports the argument that sub-basin-scale Curve Number weighting, informed by spatially distributed rainfall observations, captures the runoff response of the basin more faithfully than a lumped implementation would.

Compared with Montiel et al. [24], who reported  $MAE$  in the range of 0.20–0.70 m<sup>3</sup>/s for a smaller basin of 15.46 km<sup>2</sup>, the higher  $MAE$  values obtained in the present  $\sim 100$  km<sup>2</sup>

basin are consistent with the expected scaling of absolute error with basin size and peak-discharge magnitude, since *MAE* is not normalized by discharge amplitude.

The reproducibility of the reported metrics is supported by the fully documented a priori parameterization of both methods (Table 4), the blind-validation framework in which no parameter of either method was adjusted against observed discharge, and the description of the data-processing pipeline.

#### 4.5. Practical Implications for Water Management

The results of this research have direct implications for water-resource planning in semi-arid regions. The demonstration that rainfall heterogeneity at the 5 km scale produces differentiated runoff and recharge volumes per sub-basin implies that water-availability estimates based on single area-weighted CN values can systematically bias available water, either overestimating or underestimating depending on the rainfall distribution, with downstream consequences for agricultural, urban, and ecological water allocations. The identification of the upper sub-basin as a preferential recharge zone, operating through channel-bed infiltration rather than diffuse matrix percolation, provides actionable information for groundwater protection and land-use planning, in agreement with previous reports for similar hydrological settings [18,59].

The methodological approach presented here, combining instrumented monitoring, semi-distributed HEC-HMS simulation, and dual-method (SCS-CN and GA) a priori parameterization under a blind-validation framework, is transferable to other semi-arid basins, where data scarcity remains a structural constraint for sustainable water management [26,27]. The extension of the instrumented monitoring to additional hydrometric control points within the “Chilitas” basin is identified as a priority direction for future research in order to further constrain the sub-basin-scale runoff response and enable model–observation intercomparison at intermediate spatial scales.

The use of HEC-HMS, now widely adopted for estimating flows and runoff volumes, provides operational value at two complementary levels: estimating flows that may affect the population by generating floods, as well as estimating the volume available for water supply required for different human activities during the dry season [59], and supporting decision-making on land use and water management at the sub-basin scale. The dual-method approach adopted here, combining an empirical parameterization (SCS-CN) and a physically based formulation (Green-Ampt) under identical field-observed initial conditions, enables the operational assessment of both formulations under the same blind-validation framework, providing a robust foundation for the transferability of the methodology to comparable semi-arid contexts.

#### 4.6. Limitations

Four limitations should be acknowledged: the goodness-of-fit metrics were computed over  $n = 8$  peak-discharge values, which limits the statistical precision of individual estimates; the initial soil-moisture content required by GA was set uniformly to  $\theta_i = \theta_{fc}$  across all windows, a simplification that likely contributes to its systematic underperformance; observed discharge was recorded at a single hydrometric control point at the basin outlet, precluding direct model–observation intercomparison at the sub-basin scale; and the a priori parameterization adopted to preserve the blind-validation framework involved no calibration against observed discharge. Consequently, the reported performance should be interpreted as representative of ungauged or newly instrumented semi-arid basins and may improve where site-specific calibration is feasible.

## 5. Conclusions

A structural constraint in many developing countries is the low density of rain-gauge monitoring networks, which limits the availability of high-temporal-resolution rainfall data. In semi-arid regions, the operational assumption typically applied is that a single meteorological station represents an influence area of approximately 500 km<sup>2</sup> (a ~13 km radius under the idealized circular basin assumption), implying spatial uniformity of rainfall and, consequently, imprecise runoff estimates at the sub-basin scale.

The present research evaluated the comparative performance of SCS-CN and Green-Ampt (GA) infiltration methods implemented in HEC-HMS for rainfall-infiltration-runoff modeling under spatio-temporal rainfall variability in an instrumented semi-arid basin of approximately 100 km<sup>2</sup> in central Mexico during the 2020–2025 rainy seasons, with a monitoring network operational since October 2019. According to the currently available literature, this is one of the few multi-year, sub-basin-scale, instrumented comparisons of both methods in a semi-arid Mexican basin. Based on this, the following conclusions were drawn:

- (1) **Geomorphological characterization.** The integration of a 5 m resolution DEM with GIS-based geoprocessing and field verification enabled the identification of four hydrologically distinct sub-basins (Lower, Middle, Upper, Outlet), with concentration times ranging from 101.6 to 133.2 min. Sandy fractions dominate the upper soil horizon (0–20 cm) in the Kastanozem, Phaeozem, and Leptosol soil types, while clay fractions prevail at depths of 20–60 cm, conditioning a markedly heterogeneous infiltration response across the basin.
- (2) **Spatio-temporal rainfall variability.** Monthly rainfall depths recorded by the four-station rain-gauge network with a mean inter-gauge distance of approximately 5 km varied from 22.8 to 204.9 mm between sub-basins, demonstrating that the assumption of spatial uniformity, still common in hydrology, is not tenable for the spatially variable convective storms that dominate semi-arid regions. This heterogeneity translated into differentiated effective rainfall, infiltration depths, and runoff volumes per sub-basin, with the upper, clay-dominated sub-basin acting as a preferential recharge zone (through channel-bed infiltration along its dense drainage network). These findings confirm that area-averaged rainfall inputs systematically bias water-availability estimates in semi-arid basins.
- (3) **Inter-method comparative performance.** The a priori Curve Number values varied from 79 to 90 across sub-basins (a 20% range), while the initial moisture content specified as field capacity for the GA method varied from 0.354 to 0.180 cm<sup>3</sup>/cm<sup>3</sup> (Lower to Outlet, Table 4), favoring infiltration during low-intensity rainfall. The pairwise comparison between the SCS-CN and GA simulated peak-discharge series yielded a Pearson correlation of 0.70 and a Nash–Sutcliffe efficiency (*NSE*) of 0.25 (Table 7), values that fall within the unsatisfactory category of the Moriasi et al. criteria [54], and a mean absolute deviation of 36.88 m<sup>3</sup>/s between the two simulated series. The initial working hypothesis that the physically based GA method would outperform SCS-CN under spatio-temporally variable rainfall was therefore not confirmed: SCS-CN systematically produced closer approximations to the observed peak discharges. This outcome is attributable to three main factors: the high sensitivity of GA to the antecedent moisture condition (assumed uniform as  $\theta_i = \theta_{fc}$  across all windows), the partial compensation offered by the empirical CN weighting at the sub-basin level, and the predominance of low-depth, low-intensity convective storms that did not sustain the saturated conditions required by the GA formulation.

- (4) Model–observation agreement. The comparison between simulations and Levellogger records at the basin outlet showed that SCS-CN yielded  $R = 0.95$  and  $NSE = 0.76$  ( $MAE = 18.02 \text{ m}^3/\text{s}$ ), classifying its performance in the very good ( $NSE > 0.75$ ) category of the Moriasi criteria [54]. In contrast, Green-Ampt yielded  $R = 0.70$  and  $NSE = 0.45$  ( $MAE = 37.92 \text{ m}^3/\text{s}$ ), falling within the unsatisfactory ( $NSE \leq 0.50$ ) category. These results indicate that SCS-CN can be reliably applied to peak-discharge estimation in semi-arid instrumented basins, provided that spatially and temporally distributed rainfall observations are available at the sub-basin scale, reinforcing the value of the instrumented monitoring approach adopted in this research.
- (5) Sensitivity and structural interpretation. The sensitivity analysis (Figure 7) revealed an asymmetric response between methods: SCS-CN was highly sensitive to Curve Number perturbations (mean-deviation amplitude  $65.9 \text{ m}^3/\text{s}$  across  $\pm 6\%$ ,  $\pm 10\%$ , and  $\pm 20\%$  perturbations), whereas Green-Ampt was nearly insensitive to compound perturbations of its four soil-hydraulic parameters (amplitude  $3.2 \text{ m}^3/\text{s}$ ). This near-invariance indicates that the systematic underperformance of GA under blind validation is not correctable by parametric adjustment within reasonable a priori uncertainty, reflecting structural limitations of the physically based formulation rather than parametric imprecision.

Future research should focus on the extension of the instrumented monitoring network to additional hydrometric control points within the “Chilitas” basin, enabling model–observation intercomparison at intermediate spatial scales; on the continuation of high-resolution rainfall monitoring to further characterize the spatio-temporal variability of precipitation events, particularly those associated with extreme oceanic phenomena that govern the water balance in semi-arid regions; and on the replication of the blind-validation framework in comparable Mexican and international semi-arid basins to consolidate the transferability of the methodology and the generalizability of the SCS-CN and Green-Ampt performance asymmetry reported here.

**Author Contributions:** Conceptualization, J.G.-T. and E.C.L.; methodology, J.G.-T. and E.C.L.; software, E.C.L., S.D.-H. and A.I.V.G.; validation, D.A.C.S., A.R.R.C. and E.C.L.; formal analysis, H.E.J.-F., M.I.R. and R.U.S.A.; investigation, E.C.L., J.G.-T., H.E.J.-F., D.A.C.S. and S.D.-H.; resources, E.C.L., J.G.-T., H.E.J.-F., A.I.V.G. and D.A.C.S.; writing—original draft preparation, J.G.-T., C.O.R.R. and H.E.J.-F.; writing—review and editing, J.G.-T., C.O.R.R. and H.E.J.-F.; visualization, J.G.-T. and E.C.L.; supervision, D.A.C.S. and A.R.R.C.; project administration, J.G.-T. and E.C.L.; funding acquisition, J.G.-T. and E.C.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data cannot be made public; Readers should contact the corresponding author for details.

**Acknowledgments:** Esthela Campos Lara gratefully acknowledges the doctoral scholarship granted by the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI, formerly CONACYT) (Grant No. 1101889). The authors also acknowledge the Universidad Autónoma de Zacatecas (UAZ) for providing institutional support for this research.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Huang, J.; Yu, H.; Guan, X.; Wang, G.; Guo, R. Accelerated dryland expansion under climate change. *Nat. Clim. Chang.* **2016**, *6*, 166–171. [CrossRef]
2. Calvin, K.; Dasgupta, D.; Krinner, G.; Mukherji, A.; Thorne, P.W.; Trisos, C.; Romero, J.; Aldunce, P.; Barrett, K.; Blanco, G.; et al. *Climate Change 2023: Synthesis Report, Summary for Policymakers. Contribution of Working Groups I, II and III to the Sixth Assessment*



- Report of the Intergovernmental Panel on Climate Change*; Core Writing Team, Lee, H., Romero, J., Eds.; IPCC: Geneva, Switzerland, 2023. [CrossRef]
3. Ríos, I.H.; Cruz-Pérez, N.; Chirivella-Guerra, J.I.; García-Gil, A.; Rodríguez-Alcántara, J.S.; Rodríguez-Martín, J.; Marazuela, M.Á.; Santamarta, J.C. Proposed recharge of island aquifer by deep wells with regenerated water in Gran Canaria (Spain). *Groundw. Sustain. Dev.* **2023**, *22*, 100959. [CrossRef]
  4. Sechu, G.L.; Nilsson, B.; Iversen, B.V.; Møller, A.B.; Greve, M.B.; Trolborg, L.; Greve, M.H. Mapping groundwater-surface water interactions on a national scale for the stream network in Denmark. *J. Hydrol. Reg. Stud.* **2022**, *40*, 101015. [CrossRef]
  5. Yilma, Z.L.; Kebede, H.H. Simulation of the rainfall-runoff relationship using an HEC-HMS hydrological model for Dabus Subbasin, Blue Nile Basin, Ethiopia. *H2Open J.* **2023**, *6*, 331–342. [CrossRef]
  6. Lewandowski, J.; Meinikmann, K.; Krause, S. Groundwater-surface water interactions: Recent advances and interdisciplinary challenges. *Water* **2020**, *12*, 296. [CrossRef]
  7. Tassew, B.; Belete, M.; Miegel, K. Application of HEC-HMS Model for Flow Simulation in the Lake Tana Basin: The Case of Gilgel Abay Catchment, Upper Blue Nile Basin, Ethiopia. *Hydrology* **2019**, *6*, 21. [CrossRef]
  8. Woessner, W.W. *Groundwater-Surface Water Exchange*; The Groundwater Project: Guelph, ON, Canada, 2020.
  9. Patil, M.; Saha, A.; Pingale, S.M.; Rathore, D.S.; Goyal, V.C. Identification of potential zones on the estimation of direct runoff and soil erosion for an ungauged watershed based on remote sensing and GIS techniques. *Int. J. Eng. Geosci.* **2023**, *8*, 224–238. [CrossRef]
  10. Rahvareh, M.; Motamedvaziri, B.; Moghaddamnia, A.; Moridi, A. Modeling runoff management strategies under climate change scenarios using hydrological simulation in the Zarrineh River Basin, Iran. *J. Water Clim. Change* **2023**, *14*, 2205–2226. [CrossRef]
  11. Acworth, R.I.; Rau, G.C.; Cuthbert, M.O.; Leggett, K.; Andersen, M.S. Runoff and focused groundwater-recharge response to flooding rains in the arid zone of Australia. *Hydrogeol. J.* **2021**, *29*, 737–764. [CrossRef]
  12. Beven, K. The era of infiltration. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 851–866. [CrossRef]
  13. Guo, Z. Estimating Method of Maximum Infiltration Depth and Soil Water Supply. *Sci. Rep.* **2020**, *10*, 9726. [CrossRef] [PubMed]
  14. Thomas, A.D.; Ofori, A.E.; Emmanuel, A.; De-Graft, A.J.; Ayine, A.G.; Asare, A.; Alexander, A. Comparison and Estimation of Four Infiltration Models. *Open J. Soil Sci.* **2020**, *10*, 45–57. [CrossRef]
  15. Xie, C.; Cai, S.; Yu, B.; Yan, L.; Liang, A.; Che, S. The effects of tree root density on water infiltration in urban soil based on a Ground Penetrating Radar in Shanghai, China. *Urban For. Urban Green.* **2020**, *50*, 126648. [CrossRef]
  16. Mwendwa, S. Revisiting soil texture analysis: Practices towards a more accurate Bouyoucos method. *Heliyon* **2022**, *8*, e09395. [CrossRef] [PubMed]
  17. Dvk, N.R. Soil has a volume: Volumetric versus gravimetric expression of soil nutrient contents in understanding soil-plant interrelationships. *Acad. Lett.* **2021**, 1736. [CrossRef]
  18. Patil, M.; Saha, A.; Karwariya, S.; Pingale, S.M.; Chandra Goyal, V.; Singh Rathore, D.; Behera, N. Assessment of rainfall recharge using rainfall infiltration factor method and empirical equations. *Discov. Nat.* **2019**, *13*, 1–8. [CrossRef]
  19. Viji, R.; Prasanna, P.R.; Ilangoan, R. Modified SCS-CN and Green-Ampt Methods in Surface Runoff Modelling for the Kundahpallam Watershed, Nilgiris, Western Ghats, India. *Aquat. Procedia* **2015**, *4*, 677–684. [CrossRef]
  20. Safeeq, M.; Fares, A. Groundwater and Surface Water Interactions in Relation to Natural and Anthropogenic Environmental Changes. In *Emerging Issues in Groundwater Resources*; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 289–326. [CrossRef] [PubMed]
  21. Yang, J.; Yu, Z.; Yi, P.; Frape, S.K.; Gong, M.; Zhang, Y. Evaluation of surface water and groundwater interactions in the upstream of Kui river and Yunlong lake, Xuzhou, China. *J. Hydrol.* **2020**, *583*, 124549. [CrossRef]
  22. Yang, M.; Zhang, Y.; Pan, X. Improving the Horton infiltration equation by considering soil moisture variation. *J. Hydrol.* **2020**, *586*, 124864. [CrossRef]
  23. López, J.J.; González, M.; Scaini, A.; Goñi, M.; Valdenebro, J.; Gimena, F. Characterization of the HEC-HMS model to the river Arga's watershed in Pamplona and its application to five significant floods. *Obras Proy.* **2012**, *12*, 15–30. [CrossRef]
  24. Gonzaga, R.M.; Prado Hernández, J.V.; Vázquez Peña, M.A.; Ibáñez Castillo, L.A.; Ramírez, F.P. Evaluation of surface runoff from the numerical curve and green-ampt models in the chapingo River Basin, Mexico. *Terra Latinoam.* **2019**, *37*, 291–301. [CrossRef]
  25. Segura-Peña, J.D.; Segura-Peña, S.; Penagos Barreto, H.S.; García-Corredor, D.F.; Gómez-Castañeda, D.C. Determination of the Hydrological Parameter of the Curve Number (NC) for the Las Cañas Stream in the Siscunsi Paramo by means of Geographic Information Systems. *Lond. J. Res. Sci. Nat. Form.* **2023**, *23*, 55–65.
  26. Gichamo, T.; Nourani, V.; Gökçekuş, H.; Gelete, G. Ensemble of artificial intelligence and physically based models for rainfall-runoff modeling in the upper Blue Nile Basin. *Hydrol. Res.* **2024**, *55*, 976–1000. [CrossRef]
  27. Belina, Y.; Kebede, A.; Masinde, M. Comparative analysis of HEC-HMS and machine learning models for rainfall-runoff prediction in the upper Baro watershed, Ethiopia. *Hydrol. Res.* **2024**, *55*, 873–889. [CrossRef]

28. Azizian, A. Comparison of salt experiments and empirical time of concentration equations. *Proc. Inst. Civ. Eng. Water Manag.* **2019**, *172*, 109–122. [CrossRef]
29. Peleg, N.; Ban, N.; Gibson, M.J.; Chen, A.S.; Paschalis, A.; Burlando, P.; Leitão, J.P. Mapping storm spatial profiles for flood impact assessments. *Adv. Water Resour.* **2022**, *166*, 104258. [CrossRef]
30. Chang, C.H.; Lee, K.T. Analysis of geomorphologic and hydrological characteristics in watershed saturated areas using topographic-index threshold and geomorphology-based runoff model. *Hydrol. Process* **2008**, *22*, 802–812. [CrossRef]
31. Wang, H.; Jiao, Y.; Hu, B.X.; Li, F.; Li, D. Study on Interaction between Surface Water and Groundwater in Typical Reach of Xiaoqing River Based on WEP-L Model. *Water* **2023**, *15*, 492. [CrossRef]
32. Yin, H.; Wang, F.; Zhang, X.; Zhang, Y.; Chen, J.; Xia, R.; Jin, J. Rainfall-runoff modeling using long short-term memory based step-sequence framework. *J. Hydrol.* **2022**, *610*, 127901. [CrossRef]
33. Instituto Nacional de Estadística y Geografía (INEGI). SIATL. 2026. Available online: [https://antares.inegi.org.mx/analisis/red\\_hidro/siatl/](https://antares.inegi.org.mx/analisis/red_hidro/siatl/) (accessed on 14 May 2026).
34. Instituto Nacional de Estadística y Geografía (INEGI). Marco Geoestadístico Municipal 2005 Versión 1.0 (Censo de Población y Vivienda 2005). Available online: <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825292850> (accessed on 10 May 2026).
35. Turnipseed, D.P.; Sauer, V.B. *Discharge Measurements at Gaging Stations*; U.S. Geological Survey Techniques and Methods, Book 3, Chapter A8; U.S. Geological Survey: Reston, VA, USA, 2010; Volume 3.
36. *Datos Estadísticos Y Cartográficos*; Aguascalientes, Mexico. 2020. Available online: <https://www.inegi.org.mx/> (accessed on 14 May 2026).
37. U.S. Department of Agriculture, Soil Conservation Service. Urban Hydrology for Small Watersheds (Technical Release 55). 1986. Available online: <https://nationalstormwater.com/wp-content/uploads/2020/07/Urban-Hydrology-for-Small-Watersheds-TR-55.pdf> (accessed on 14 May 2026).
38. Rawls, W.J.; Asce, M.; Brakensiek, D.L.; Miller, N. Green-ampt infiltration parameters from soils data. *J. Hydraul. Eng.* **1983**, *109*, 62–70. [CrossRef]
39. Tügel, F.; Hassan, A.; Hou, J.; Hinkelmann, R. Applicability of Literature Values for Green–Ampt Parameters to Account for Infiltration in Hydrodynamic Rainfall–Runoff Simulations in Ungauged Basins. *Environ. Model. Assess.* **2022**, *27*, 205–231. [CrossRef]
40. Bouyoucos, G.J. Hydrometer Method Improved for Making Particle Size Analyses of Soils. *Agron. J.* **1962**, *54*, 464–465. [CrossRef]
41. Bouwer, H. Intake rate: Cylinder infiltrometer. In *Methods of Soil Analysis. Part 1: Physical and Mineralogical Methods*, 2nd ed.; Klute, A., Ed.; Agronomy: Madison, WI, USA, 1986; Volume 9, pp. 825–844.
42. Burt, R. *Soil Survey Laboratory Methods Manual*; National Soil Survey Center: Lincoln, NE, USA, 2004.
43. Sultan, D.; Tsunekawa, A.; Tsubo, M.; Haregeweyn, N.; Adgo, E.; Meshesha, D.T.; Fenta, A.A.; Ebabu, K.; Berihun, M.L.; Setargie, T.A. Evaluation of lag time and time of concentration estimation methods in small tropical watersheds in Ethiopia. *J. Hydrol. Reg. Stud.* **2022**, *40*, 101025. [CrossRef]
44. Lian, H.; Yen, H.; Huang, J.C.; Feng, Q.; Qin, L.; Bashir, M.A.; Wu, S.; Zhu, A.-X.; Luo, J.; Di, H.; et al. CN-China: Revised runoff curve number by using rainfall-runoff events data in China. *Water Res.* **2020**, *177*, 115767. [CrossRef] [PubMed]
45. Parnas, F.E.Å.; Abdalla, E.M.H.; Muthanna, T.M. Evaluating three commonly used infiltration methods for permeable surfaces in urban areas using the SWMM and STORM. *Hydrol. Res.* **2021**, *52*, 160–175. [CrossRef]
46. Ortega-Gaucin, D.; Bartolón, J.D.I.C.; Bahena, H.V.C. Drought vulnerability indices in Mexico. *Water* **2018**, *10*, 1671. [CrossRef]
47. Lee, J.; Yoo, C. Determination of Clark unit hydrograph parameters for estimating probable maximum flood. *Hydrol. Res.* **2023**, *54*, 120. [CrossRef]
48. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **2007**, *50*, 885–900. [CrossRef]
49. Xiang, S.; Huang, S.; Peng, H.; Ye, J.; Ma, Y.; Qiu, H.; Guo, J.; Nie, H.; Chen, H. Flood simulation using HEC-HMS model based on antecedent soil moisture changes in the Fenshui River Basin. *Hydrol. Res.* **2026**, *57*, 542–565. [CrossRef]
50. Kousari, M.R.; Malekinezhad, H.; Ahani, H.; Asadi Zarch, M.A. Sensitivity analysis and impact quantification of the main factors affecting peak discharge in the SCS curve number method: An analysis of Iranian watersheds. *Quat. Int.* **2010**, *226*, 66–74. [CrossRef]
51. Eslamian, S. Book: Handbook of Engineering Hydrology (Three-Volume Set). In *Handbook of Engineering Hydrology*; Eslamian, S., Ed.; CRC Press: Boca Raton, FL, USA, 2014; pp. 23–40.
52. Patle, G.T.; Sikar, T.T.; Rawat, K.S.; Singh, S.K. Estimation of infiltration rate from soil properties using regression model for cultivated land. *Geol. Ecol. Landsc.* **2019**, *3*, 1–13. [CrossRef]



53. Zimmermann, E. Bayesian approach to daily rainfall modelling to estimate monthly net infiltration using the Thornthwaite water budget and Curve Number methods. *Hydrogeol. J.* **2006**, *14*, 648–656. [CrossRef]
54. Moriasi, D.N.; Gitau, M.W.; Pai, N.; Daggupati, P. Hydrologic and water quality models: Performance measures and evaluation criteria. *Trans. ASABE* **2015**, *58*, 1763–1785. [CrossRef]
55. Zakaria Lukwasa, A.; Ayal, D.Y.; Zeleke, T.T.; Beketie, K.T. Spatio-temporal rainfall variability and its linkage with large scale climate oscillations over the Republic of South Sudan. *Clim. Serv.* **2022**, *28*, 100322. [CrossRef]
56. Derdour, A.; Bouanani, A.; Babahamed, K. Modelling rainfall runoff relations using HEC-HMS in a semi-arid region: Case study in Ain Sefra watershed, Ksour Mountains (SW Algeria). *J. Water Land Dev.* **2018**, *36*, 45–55. [CrossRef]
57. El-Hames, A.S. An empirical method for peak discharge prediction in ungauged arid and semi-arid region catchments based on morphological parameters and SCS curve number. *J. Hydrol.* **2012**, *456–457*, 94–100. [CrossRef]
58. Odey, G.; Cho, Y. Event-Based vs. Continuous Hydrological Modeling with HEC-HMS: A Review of Use Cases, Methodologies, and Performance Metrics. *Hydrology* **2025**, *12*, 39. [CrossRef]
59. Munyejuru, I.F.; Stagge, J.H. Multi-Objective Calibration of a Pre-Modern Nile Hydrologic Model Using Recovered Records. *Hydrology* **2026**, *13*, 114. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## Article

# Assessing Flood Susceptibility Using Machine Learning in Arid Regions

Mostafa Mashal <sup>1,2,\*</sup>, Doaa Amin <sup>2</sup>, Mona A. Hagras <sup>1</sup> and Ashraf M. Elmoustafa <sup>1</sup>

<sup>1</sup> Irrigation and Hydraulics Department, Faculty of Engineering, Ain Shams University, Cairo 11517, Egypt; mona\_hagras@eng.asu.edu.eg (M.A.H.); ashraf\_elmoustafa@eng.asu.edu.eg (A.M.E.)

<sup>2</sup> Water Resources Research Institute (WRI), National Water Research Center (NWRC), Ministry of Water Resources and Irrigation (MWRI), Qalyubia 13621, Egypt; doaa\_amin74@yahoo.com

\* Correspondence: mostafa.mashal2@yahoo.com; Tel.: +20-100-6567083

## Abstract

Flash floods are among the most destructive natural hazards, often causing substantial loss of life and severe damage to infrastructure and property. Predicting flood-prone areas remains challenging because flood generation is controlled by complex interactions among topographic, hydrological, climatic, and environmental factors. In this study, six machine learning algorithms—Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree Classifier (DTC), AdaBoost, and Artificial Neural Network (ANN)—were developed to predict flash-flood inundation locations using satellite-derived flood inventories from two major rainfall events in Wadi El-Darb and Wadi El-Allaqi, Egypt. Model performance was evaluated using accuracy, precision, recall, and F1-score. During model development, Random Forest and Decision Tree Classifier achieved the highest prediction accuracy (94%), followed by AdaBoost and ANN (92%), while Logistic Regression (89%) and SVM (88%) also produced satisfactory results. To evaluate model generalization, the trained models were independently validated using a rainfall event in Wadi Hodein (Egypt) and a major flash-flood event that occurred in Oman during April 2024. The external validation showed that AdaBoost achieved the highest predictive performance in both validation basins, with accuracies of 87% for Wadi Hodein and 83% for Oman, providing encouraging initial evidence of applicability across hydrologically similar arid watersheds. While AdaBoost and Logistic Regression maintained satisfactory performance during external validation, other algorithms exhibited noticeable reductions in recall and F1-score, particularly in the Oman case study, indicating variability in model generalization across independent watersheds. These findings suggest that the proposed framework may support flood susceptibility assessment in ungauged arid environments with comparable hydrological characteristics, although further validation across a wider range of climatic and geological settings is needed. Overall, the results highlight the value of integrating satellite remote sensing with machine learning to support flood hazard assessment, disaster preparedness, early warning systems, and flood risk management in data-scarce regions.

**Keywords:** machine learning; flood hazard; flood susceptibility

## 1. Introduction

Floods are among the most frequent and destructive natural disasters worldwide, causing severe impacts on human life, infrastructure, and ecosystems [1–3]. Recent global

assessments report that flood events have inundated more than 2.23 million km<sup>2</sup> and affected over 290 million people in the last few decades [4]. Information on floods—such as the spatial and temporal extent of inundation, flood depth, vulnerable areas, high-risk zones, and safe locations—is essential for effective disaster management and emergency response [5]. These figures highlight the growing need for accurate and reliable methods to predict flood inundation under extreme rainfall events.

Climate change is expected to intensify the frequency and magnitude of extreme rainfall events in many regions in the world, increasing the likelihood of flash floods and large-scale inundation [6]. Assessing future flood hazard under such conditions is particularly challenging, as it requires integrating rainfall projections with complex topographic, hydrologic, and land-surface characteristics across large spatial domains [7].

Traditional deterministic models become computationally expensive at these scales, especially when forced with outputs from multiple climate scenarios. Therefore, the development of machine learning models that can incorporate diverse rainfall events alongside topographic and environmental predictors offers a promising pathway for efficiently simulating and mapping future flood hazards [8,9].

Flood hazard assessment plays a crucial role in supporting early warning systems and reducing flood-related impacts. Traditionally, several approaches have been applied, ranging from statistical analyses based on historical flood data [10] to uncertainty-based methods such as Artificial Neural Networks and fuzzy logic [11]. Remote sensing techniques have also become a key tool for detecting and mapping inundated areas [12]. In addition, hydrological and hydrodynamic models—such as MIKE, Info Works ICM, and LISFLOOD-FP—are widely used to simulate the spatial distribution and evolution of floods [13,14]. However, these models often require high-resolution input data, extensive calibration, and are computationally expensive when applied to large regions. As a result, machine learning approaches have gained attention as efficient alternatives, offering the ability to integrate diverse inputs and rapidly predict flood-prone locations.

Advances in computer science and artificial intelligence have greatly expanded the use of machine learning in flood prediction. These algorithms enhance the efficiency of hazard assessment, address complex nonlinear relationships, and often provide more objective results compared to traditional methods. Owing to their flexibility and predictive capability, machine learning models have shown strong potential for widespread application in flood hazard studies. Broadly, machine learning techniques can be divided into supervised, unsupervised, and reinforcement learning. Supervised learning relies on labeled datasets to learn patterns and predict new outcomes, while unsupervised learning groups data into clusters without prior labels. Reinforcement learning, on the other hand, focuses on optimizing decision-making through iterative interaction with the environment. Several machine learning algorithms have been widely applied in flood risk assessment, each with unique strengths in pattern recognition and prediction. These include Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) [15], Random Forest (RF) [16], Logistic Regression (LR) [17], Support Vector Machines (SVM) ([18]), and Particle Swarm Optimization (PSO) [19]. These models generally work by identifying relationships between input indicator factors and flood risk outcomes, enabling predictions under different environmental and hydrological scenarios [20]. In this study, six machine learning algorithms, Random Forest, Logistic Regression, Support Vector Machine, Decision Tree Classifier, AdaBoost, and ANN, were selected for flood hazard assessment, and their results were systematically compared.

Recent advances in remote sensing, cloud-computing platforms, and machine learning have significantly enhanced flood susceptibility assessment by enabling the integration of multi-source geospatial datasets with advanced predictive algorithms. Recent studies have

demonstrated the effectiveness of combining Sentinel-1 SAR imagery, Landsat data, and the Sentinel Application Platform (SNAP) for flood assessment under different environmental conditions, highlighting the value of satellite-based approaches for flood monitoring and hazard mapping [21]. Similarly, integrating satellite imagery with machine learning algorithms has substantially improved flood susceptibility mapping by combining remotely sensed environmental variables with robust predictive models and comprehensive performance evaluation [22]. More recent studies have also emphasized systematic comparisons of machine learning algorithms and the integration of multiple environmental conditioning factors to improve the accuracy, reliability, and interpretability of flood susceptibility models [23]. Despite these advances, relatively few studies have evaluated the spatial transferability of machine learning models across independent watersheds, particularly in ungauged arid environments. Addressing this limitation is one of the primary objectives of the present study.

Despite the significant progress in flood hazard assessment, several limitations remain in the existing literature. Physically based hydrological and hydraulic models, while capable of providing detailed flood simulations, require extensive input data, calibration, and computational resources, which are often unavailable in arid and ungauged regions. On the other hand, many machine learning studies rely on locally trained datasets and are typically developed for specific watersheds, limiting their transferability to other regions with different hydrological and geomorphological characteristics. In addition, although remote sensing has been widely used for flood detection, its integration with machine learning models for predictive flood hazard assessment in data-scarce environments remains limited. Many existing approaches focus either on flood mapping or susceptibility analysis and are commonly developed for individual watersheds. Consequently, there remains a need to investigate the extent to which machine learning models trained using multi-source satellite datasets can generalize to independent watersheds with comparable hydrological and environmental characteristics.

To address these gaps, the main objectives of this study are summarized as follows:

- To develop and evaluate a machine learning framework for flood susceptibility assessment in arid and data-scarce regions using multi-source satellite data.
- To integrate satellite-derived flood inundation maps, satellite-based rainfall data, and topographic/environmental conditioning factors for model training and evaluation.
- To compare the performance of multiple supervised and unsupervised machine learning algorithms for flood susceptibility mapping.
- To investigate whether training the models using two hydrologically distinct basins (Wadi El-Darb and Wadi El-Allaqi) improves their ability to generalize to independent case studies.
- To evaluate model transferability by testing the trained models on independent and unseen case studies (Wadi Hodein and a flood event in Oman) without retraining.
- To investigate the potential of satellite-driven machine learning approaches for supporting flood susceptibility mapping in ungauged regions.

The novelty of this study lies in the development and evaluation of a machine learning framework for flood susceptibility assessment in arid and data-scarce environments through the integration of multi-source satellite datasets, including flood inundation maps and event-based rainfall information. Unlike many previous studies that focus on single-watershed analyses, this research combines data from two hydrologically distinct basins to investigate whether broader training conditions can improve model generalization. Furthermore, the proposed framework is evaluated using two independent flood events without retraining, providing an initial assessment of model transferability under different environmental conditions. The external validation demonstrates encouraging results

while also highlighting that transferability varies among machine learning algorithms and requires further evaluation across a wider range of climatic and geomorphological settings. Overall, the study demonstrates the potential of satellite-driven machine learning approaches for supporting flood susceptibility assessment in data-scarce environments.

## 2. Materials and Methods

### 2.1. Study Area

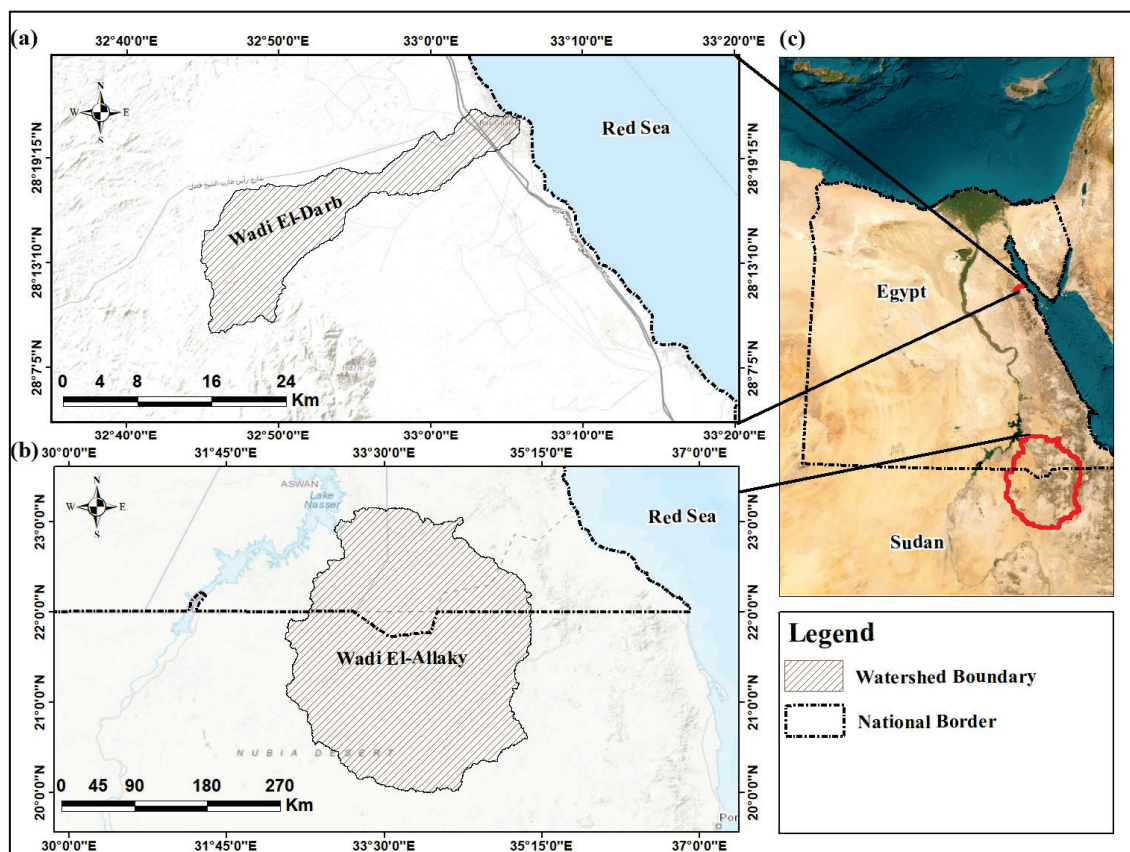
This study focuses on two distinct watersheds in Egypt that were selected based on their hydrological importance and the availability of suitable satellite imagery for flood analysis. The first site is Wadi El-Darb, located in the Eastern Desert, which drains eastward toward the Red Sea and directly threatens Ras Gharib city, one of the most vulnerable urban areas to flash floods along the Red Sea coast. Wadi El-Darb extends between latitudes  $28^{\circ}08' \text{ N}$  and  $28^{\circ}24' \text{ N}$  and longitudes  $32^{\circ}20' \text{ E}$  and  $32^{\circ}44' \text{ E}$ , and is well known for its susceptibility to intense flash flooding. The second site is Wadi El-Allaqi, one of the largest catchments in the Eastern Desert, extending partially into Sudan and draining westward into Lake Nasser, upstream of the High Aswan Dam. Wadi El-Allaqi spans the region between latitudes  $20^{\circ}00' \text{ N}$  and  $23^{\circ}10' \text{ N}$  and longitudes  $32^{\circ}30' \text{ E}$  and  $35^{\circ}10' \text{ E}$ , and, similar to Wadi El-Darb, is highly prone to flash flood events.

The rationale for selecting these two basins lies in their contrasting drainage directions, geomorphological features, and hydrological settings. Wadi El-Darb represents a west–east drainage system, where floods pose direct risks to coastal urban areas, while Wadi El-Allaqi illustrates an east–west drainage system, where floodwaters contribute to Lake Nasser. This contrast was deliberately chosen to ensure that the machine learning models are trained on diverse hydrological conditions, enabling them to capture different flood-generating processes and to enhance their robustness for wider application. In both basins, suitable satellite images capturing flood events were available, which facilitated the extraction of inundated locations used for model training and validation. Figure 1 shows the geographic location and boundaries of the two case study areas.

The selection of Wadi El-Darb and Wadi El-Allaqi as case studies is based on both their hydrological significance and data availability. These basins are located in arid and semi-arid regions where flash floods are typically triggered by short-duration, high-intensity rainfall events. The geomorphological characteristics of these wadis, including steep slopes, sparse vegetation cover, and low infiltration capacity, promote rapid surface runoff and increase flood susceptibility. Wadi El-Darb, in particular, has experienced severe flash flood events, such as the Ras Gharib flood on 28 October 2016, which caused significant damage to infrastructure and settlements. Similarly, Wadi El-Allaqi is a large and hydrologically active basin that responds quickly to intense rainfall events, such as the event of 7–9 August 2024, generating noticeable runoff and flood inundation. In addition to their hydrological characteristics, the availability of high-quality satellite imagery corresponding to these flood events played a key role in their selection. This ensured the ability to accurately extract flood inundation maps required for training and validating the machine learning models.

Although Wadi El-Darb, Wadi El-Allaqi, Wadi Hodein, and the Oman validation basin differ in their local geological composition and lithological characteristics, they share several important hydrological and geomorphological features typical of arid environments. All watersheds are characterized by sparse vegetation, ephemeral stream networks, limited infiltration during intense rainfall, steep mountain fronts, and rapid runoff responses that generate flash floods. These common hydrological characteristics provide a suitable basis for evaluating the transferability of the developed machine learning models while acknowledging that local geological variability may influence runoff generation to some extent.





**Figure 1.** (a) Location map of Wadi El-Darb watershed in Egypt along the Red Sea coast; (b) location map of Wadi El-Allaqi watershed in southeastern Egypt; (c) regional overview showing the study area location within Egypt and its proximity to Sudan and the Red Sea.

## 2.2. Data Description

This part describes the steps in processing the acquired data for the machine learning model implementation. All variables were derived from all raster datasets acquired, processed, and subsequently prepared for modeling using geospatial techniques and analysis.

### 2.2.1. Flood Events Description

The flood events used in this study were selected based on the occurrence of significant rainfall that generated observable surface runoff in the study areas, along with the availability of corresponding satellite imagery. Two major rainfall-triggered flood events were considered. The first event occurred on 28 October 2016 in Wadi El-Darb (Ras Gharib region), which resulted in severe flash flooding affecting the area. The second event corresponds to a multi-day rainfall event that took place from 7 to 9 August 2024 in Wadi El-Allaqi, leading to noticeable surface runoff and flood inundation.

These events were selected to represent different temporal and hydrological conditions, ensuring that the dataset captures variability in flood generation processes. For each event, satellite imagery acquired immediately after the rainfall was used to delineate the flood extent, which was then used to extract flooded and non-flooded samples for model development.

### 2.2.2. Rainfall Data Description

Rainfall data used in this study were obtained exclusively from satellite-based products provided by the Center for Hydrometeorology and Remote Sensing (CHRS), as no ground-



based rain gauge data were available within the study areas. Therefore, no merging or bias correction between different rainfall sources was required.

The rainfall variable was considered on an event basis, where the total rainfall associated with each flood-triggering event was used as an input to the machine learning models. The CHRS dataset has a spatial resolution of approximately  $0.1^\circ$  ( $\sim 10$  km), which is suitable for representing regional rainfall patterns in arid environments.

Rainfall data were downloaded as raster layers corresponding to the selected flood events. To integrate rainfall into the dataset, the rainfall value at each sample point was extracted from the raster data using spatial interpolation techniques within the GIS environment. This ensured that each point in the dataset was associated with a representative rainfall value corresponding to the triggering event.

### 2.2.3. Dataset Description

The dataset used for model development was constructed by combining samples from both Wadi El-Darb and Wadi El-Allaqi. This integration was intentionally performed to expose the machine learning models to diverse hydrological conditions, rainfall characteristics, and geomorphological features, thereby improving their ability to generalize across different environments.

The combined dataset was randomly divided into training (70%) and testing (30%) subsets, ensuring that both subsets included representative samples from the two watersheds. This approach enables the models to learn from heterogeneous conditions while maintaining a consistent evaluation framework. Figure 2. Flowchart of the proposed methodology for flood susceptibility assessment, illustrating the processes of data collection, preprocessing, training dataset construction, machine learning model development, validation, and flood susceptibility mapping.

### 2.2.4. Derivation of Flash Flood Triggering Factors

Flash floods result from the combined influence of topographic, hydrological, and environmental factors, making the careful selection of input variables essential for reliable machine learning predictions [24]. In this study, static triggering factors were derived mainly from global datasets, guided by basin characteristics and insights from previous research. Rainfall is the primary driver of floods; however, runoff also depends on watershed conditions such as catchment area, topography, and land use/land cover. The selected factors include elevation, slope degree, slope length, topographic wetness index (TWI), distance to main streams, and curve number, etc., with most derived from Shuttle Radar Topographic Mission (SRTM) DEM data.

Although lithology, bedrock permeability, and geological structures influence runoff generation, these variables were beyond the scope of the current satellite-based framework because consistent geological datasets were unavailable for all study and validation basins. Instead, infiltration-related characteristics were represented indirectly through the Curve Number (CN), which incorporates Hydrologic Soil Groups and land-use information. Therefore, CN should be regarded as a partial representation of infiltration characteristics rather than a direct representation of geological conditions.

The data processing starts with the derivation of flash flood-triggering factors after processing the SRTM DEM using ArcMap version 10.8.2 software. The majority of flash flood triggering factors, such as elevation, topographic wetness index (TWI), slope length, slope degree, and distance from the main stream, were extracted for each pixel from the DEM using geospatial techniques. The data sources of these input factors were outlined in Table 1. Here is an example of some of the influencing factors and how to calculate them:

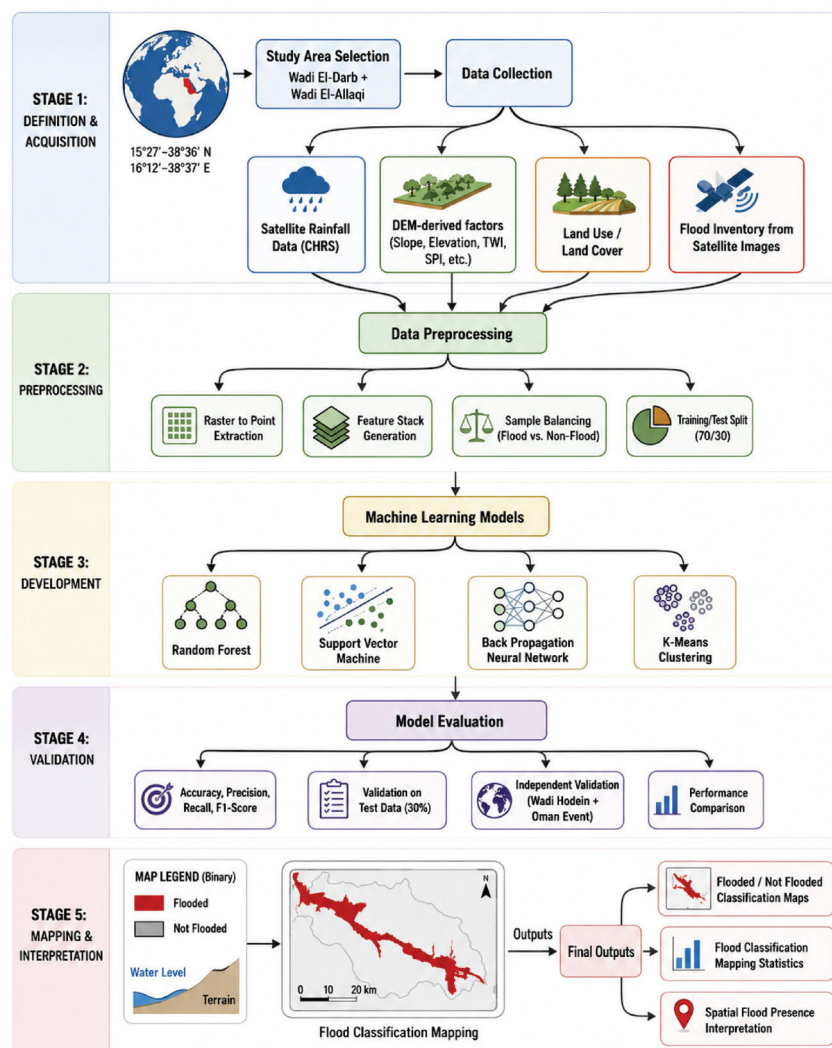


Figure 2. Proposed methodology for flood susceptibility assessment.

Table 1. Input factor selection and data source.

| No. | Factor Name             | Unit   | Data Source                                                                                                                                                                                                     |
|-----|-------------------------|--------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | Elevation               | m      | <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> (Accessed on 5 January 2026)                                                                                                      |
| 2   | Distance to main stream | m      | Derived from elevation data                                                                                                                                                                                     |
| 3   | Curve number (CN)       | None   | Derived from: 1. WRRI Soil Type Maps (Source A) and 2. Land Use/Land Cover Data (Source B). Derived using: The USDA-SCS Curve Number Look-up Tables to combine the Hydrologic Soil Groups and Land Use classes. |
| 4   | TWI                     | None   | Derived from elevation data                                                                                                                                                                                     |
| 5   | Slope Degree            | Degree | Derived from elevation data                                                                                                                                                                                     |
| 6   | Precipitation           | mm     | <a href="https://chrdata.eng.uci.edu/">https://chrdata.eng.uci.edu/</a> (Accessed on 5 January 2026)                                                                                                            |
| 7   | Slope Length            | m      | Derived from elevation data                                                                                                                                                                                     |

Table 1. Cont.

| No. | Factor Name       | Unit               | Data Source                 |
|-----|-------------------|--------------------|-----------------------------|
| 8   | Flow Accumulation | Count of cells     | Derived from elevation data |
| 9   | Aspect            | Degrees            | Derived from elevation data |
| 10  | Stream Density    | Km/km <sup>2</sup> | Derived from elevation data |
| 11  | SPI               | none               | Derived from elevation data |
| 12  | Plan Curvature    | m <sup>-1</sup>    | Derived from elevation data |
| 13  | Profile Curvature | m <sup>-1</sup>    | Derived from elevation data |
| 14  | Curvature         | m <sup>-1</sup>    | Derived from elevation data |

### ▪ Slope

The slope is one of the key factors that trigger flash floods; areas with higher slope angles will increase the water's velocity. This means that the speed of the flow of flooded waters is controlled by slope, and as a result, it is considered an important factor that can lead to flash floods [24]. In addition to this, an increased slope led to surface runoff and a reduction in infiltration. The slope (in degrees) of the study area was derived from the SRTM DEM using the spatial analyst tool in the ArcMap GIS environment.

### ▪ Topographic Wetness Index (TWI)

This is a factor that has a direct impact on flood occurrence. It is a spatial representation of the variation of basin wetness. It measures the tendency to accumulate water at any point in the basin [25]. Each pixel of the TWI raster data depicts the amount of water stored in the location represented by such a pixel.

In order to calculate the TWI for the study area, the following equation was used after using the map algebra tool raster calculator in ArcMap for the calculation [26].

$$TWI = \ln \left( \frac{(\alpha)}{\tan(\beta)} \right) \quad (1)$$

where

$\alpha$  indicates the upslope area of a specific catchment area draining through a particular point;

$\beta$  represents the slope angle (in degrees) at such a point [26]. Areas with high values of TWI are highly correlated with flooding.

### • Curvature

Curvature is the degree of curvature of a curve. Negative curvature means concave downward, positive curvature means convex upward, and a curvature of 0 means the surface is flat. Therefore, the smaller the curvature value, the more prone to flooding. In this paper, the curvature analysis tool in ArcGIS version 10.8.2 software was used to obtain the curvature distribution of the study area.

### • Stream Power Index (SPI)

The Stream Power Index (SPI) is used to quantify the erosive and destructive potential of flowing water. It is calculated based on two key parameters: catchment area (flow accumulation) and slope, as expressed in Equation (2).

$$SPI = (\alpha) * \tan(\beta) \quad (2)$$

where  $\alpha$ : the catchment area,  $\beta$ : Slope.

It should be noted that several of the conditioning factors used in this study are derived from the same DEM dataset and may exhibit some degree of correlation. However, the selection of these variables is consistent with previous studies in flood susceptibility assessment, where DEM-derived indices such as slope, curvature, topographic wetness index (TWI), stream power index (SPI), and flow accumulation are widely used due to their strong physical relevance to hydrological processes.

Each of these factors captures a different aspect of terrain influence on flood generation, including runoff velocity, flow convergence, water accumulation potential, and surface morphology. Therefore, they were retained to preserve the physical interpretability of the model inputs.

Flood samples derived from satellite imagery are naturally concentrated near drainage networks, which may influence the importance of variables such as distance to stream. Although efforts were made to include non-flooded samples from diverse locations, this spatial clustering may still affect model interpretation.

Furthermore, the machine learning algorithms employed in this study, particularly tree-based models such as Random Forest and Decision Tree Classifier, are known to be relatively robust to multicollinearity and can effectively handle correlated predictors. In addition, factor importance analysis was conducted to evaluate the contribution of each variable to the model performance.

Although geological conditions influence runoff generation through lithology, bedrock permeability, weathering characteristics, and structural features, these variables were not explicitly incorporated into the machine learning models because consistent geological datasets with comparable spatial resolution were unavailable for all study and validation basins. Instead, geological effects were considered only indirectly through the Curve Number (CN), which integrates Hydrologic Soil Groups and land-use characteristics that influence infiltration and runoff potential. Therefore, the CN should be interpreted as a partial representation of infiltration-related properties rather than a direct substitute for geological information.

To assess potential feature redundancy within the 14 flood conditioning factors, a multicollinearity diagnostic was conducted using the Variance Inflation Factor (VIF). As presented in Table 2, all primary hydro-geomorphological and meteorological factors (Factors 1 to 11) exhibited exceptional statistical independence, with VIF values strictly below 2.5, well within the conservative academic threshold of 5.0. Severe collinearity was mathematically observed only among the curvature components (Plan, Profile, and Total Curvatures), which is physically expected due to their definitive geometric relationships. However, all 14 parameters were strictly retained in the final modeling framework. This decision is justified by the deployment of advanced, non-linear machine learning algorithms (such as Random Forest, AdaBoost, and Support Vector Machines), which are structurally robust against multicollinearity and capable of processing overlapping topographic constraints without losing prediction stability.

**Table 2.** Multicollinearity analysis of the 14 flood conditioning factors using the Variance Inflation Factor (VIF).

| Flood Conditioning Factor | VIF  |
|---------------------------|------|
| Elevation                 | 1.33 |
| Distance to main stream   | 1.62 |
| Curve number (CN)         | 1    |
| TWI                       | 1.54 |
| Slope Degree              | 1.57 |

Table 2. Cont.

| Flood Conditioning Factor | VIF                   |
|---------------------------|-----------------------|
| Precipitation             | 1.03                  |
| Slope Length              | 1.16                  |
| Flow Accumulation         | 2.36                  |
| Aspect                    | 1.05                  |
| Stream Density            | 1.62                  |
| SPI                       | 2.34                  |
| Plan Curvature            | $4.5 \times 10^{10}$  |
| Profile Curvature         | $2.41 \times 10^{11}$ |
| Curvature                 | $3.48 \times 10^{11}$ |

Although several DEM-derived variables exhibited relatively high VIF values, they were retained because the objective of this study was predictive flood susceptibility mapping rather than statistical coefficient inference. Many machine learning algorithms are capable of handling correlated predictors during model training, although multicollinearity may influence the interpretation of feature importance rankings. Consequently, the reported feature importance values should be interpreted as model-specific measures of predictive contribution rather than evidence of independent hydrological influence. Because several terrain attributes are derived from the same digital elevation model, they inevitably share overlapping information, and their relative importance may vary among machine learning algorithms.

### 2.3. Research Methodology

This study utilizes multiple datasets including remote sensing data, geological, and topographic data to map historical flood occurrences, develop susceptibility models, and evaluate their accuracy. As a part of the machine learning data requirements, gridded rainfall data for the defined flash flood events historical dates was obtained and incorporated into the analysis, as it is a crucial factor influencing the occurrence of flood events, the temporal and spatial patterns of rainfall were analyzed to understand its role in triggering the selected flash flood occurrences.

The study's key requirement was to associate all factors that lead to flash floods together as input data for the machine learning model. In the previous section, all flash-flood conditioning factors were calculated, but we need to link the calculated pixel Factors with the associated rainfall that possibly triggered the flash floods within that location.

To associate the historical locations of flash floods with their corresponding rainfall events, the following steps were followed.

The dates of the flash flood events and the corresponding dates of the rainfall data that match these dates were first selected and compiled.

Rainfall data were obtained from satellite-based products (CHRS) due to the absence of ground-based rain gauge measurements in the study areas.

Rainfall data were added to the dataset, which was used as input data for the machine learning.

A database consisting of all the spatiotemporal features extracted from the past flood events, combined with the different selected factors and their corresponding target class for each pixel, was created. This database served as input for the classification modeling

as described before. The database was prepared for further processing before the actual implementation of the machine learning models.

### ▪ **Overview of Machine Learning Algorithms**

This study addresses a binary classification task, where flood-prone and non-flooded locations were distinguished using machine learning. To identify the most suitable approach, several algorithms—including Random Forest, Support Vector Machine, Decision Tree Classifier, Logistic Regression, AdaBoost, and Artificial Neural Networks—were comparatively evaluated. These models were selected for their proven accuracy, efficiency, and robustness against overfitting in similar applications. The following sections outline some of their theoretical basis and operational mechanisms, providing the foundation for evaluating their predictive performance in this research.

### ▪ **Random Forest**

Random Forest is a widely used ensemble learning algorithm that combines the results of multiple decision trees built through bootstrap aggregation and random feature selection [27]. One of its key advantages is the ability to evaluate the relative importance of input variables, offering insights into data relationships [28]. Decision trees, the building blocks of Random Forest, partition data using criteria such as Gini impurity and information entropy, both aimed at reducing classification uncertainty. Random Forest is highly effective in handling large datasets with complex, non-linear interactions between variables without requiring assumptions about their relationships. It is also robust, efficient, and capable of producing accurate results even with default parameter settings. These strengths make Random Forest one of the most versatile and reliable machine learning techniques for predictive modeling.

### ▪ **Logistic regression**

Binary Logistic Regression (LR) is a supervised classification method used to predict dichotomous outcomes ('0' or '1'). It employs the logistic function, ranging from 0 to 1, to estimate the probability that a data point belongs to the positive class [29]. The function is defined as in Equation (3):

$$P(x) = \frac{1}{1 + e^{-(\beta_0 - \beta_1 x)}} \quad (3)$$

where

$P(x)$  is interpreted as the probability of the dependent variable  $Y$  equaling a success/case rather than a failure/non-case;

$e$  is the base of the natural logarithm (approximately 2.71828);

$\beta_0$  and  $\beta_1$  are the regression coefficients, which are estimated from the training data.

Logistic Regression is widely used for binary classification due to its simplicity, probabilistic interpretation, and ability to capture non-linear relationships between predictors and the outcome variable.

### ▪ **Artificial Neural Networks (ANNs)**

Artificial Neural Networks (ANNs) are non-linear machine learning models inspired by the structure of the human brain, designed to capture complex interactions between input and output variables [30]. A typical ANN consists of an input layer, one or more hidden layers, and an output layer, where hidden layers apply non-linear transformations to extract patterns from the data [31]. The learning process is based on assigning weights to inputs and adjusting them iteratively to minimize prediction errors. The general form of the ANN output can be expressed as follows:

$$O(v) = f\left(\omega_0 + \sum_{r=1}^k \omega_r x_r\right) = f\left(\omega_0 + W^T v\right) \quad (4)$$



where

$O(v)$  is the output of the neural network;

$\omega_0$  is the bias (intercept) term;

$k$  is the total number of flood conditioning factors;

$\omega_k$  the flood conditioning factors' weight coefficient;

$f$  specifies the activation function in the hidden and output layers, ranging from 0 to 1.

The function is calculated by Equation (5)

$$W^T v = (\omega_1 v_1 + \omega_2 v_2 + \dots + \omega_k v_k) \quad (5)$$

where

$\omega$  is the vector containing all the weights without the intercept;

$v$  is the vector of all covariates;

$W^T v$  is a scaled product of the weights and input vectors.

This iterative training process enables ANNs to learn underlying patterns and enhance predictive performance, making them highly effective for flood hazard modeling.

#### 2.4. Flood Inundation Mapping from Satellite

Flood inundation maps were derived from satellite imagery and used as reference data for training and validating the machine learning models. In this study, multispectral Sentinel-2 imagery with a spatial resolution of 10 m was utilized due to its high spatial detail and suitability for surface water detection. The images were selected immediately after the rainfall and runoff events to capture the maximum spatial extent of flood inundation.

Preprocessing steps included atmospheric correction and cloud masking to ensure the reliability of the spectral information. Flooded areas were identified using a combination of spectral water indices, including the Normalized Difference Water Index (NDWI), Modified Normalized Difference Water Index (MNDWI), and Soil Moisture Index (SMI). These indices were employed to enhance the detection of open water and saturated soils, which are characteristic of flood conditions, especially in arid and semi-arid regions.

Threshold-based classification was applied to each index to distinguish flooded from non-flooded pixels. The outputs of the different indices were then integrated to improve classification accuracy and reduce misclassification caused by soil background or vegetation effects. The accuracy of the extracted flood maps was evaluated through visual interpretation and comparison with known flood-affected areas and high-resolution imagery where available.

The final flood inundation maps were used to generate the binary target variable required for machine learning model development, where flooded locations were assigned a value of 1 and non-flooded locations a value of 0.

The performance of machine learning models is strongly influenced by the selection of hyperparameters, which control model complexity, learning behavior, and generalization capability. In this study, commonly used hyperparameters were adopted for each algorithm based on their relevance in controlling model structure and prediction performance. These parameters were selected to ensure stable and reliable results while avoiding overfitting. A summary of the key hyperparameters used for each machine learning model is presented in Table 3.

Hyperparameter optimization was performed using a combination of Grid Search and empirical tuning. For DTC, Grid Search with cross-validation was applied, whereas for RF, SVM, LR, AdaBoost, and ANN, parameter values were selected based on literature recommendations and preliminary experimental evaluation.

**Table 3.** Hyperparameter settings and tuning strategy for the machine learning models.

| Model               | Hyperparameter       | Explored Values/Strategy            | Final Selected Value            |
|---------------------|----------------------|-------------------------------------|---------------------------------|
| RF                  | n_estimators         | 50, 100, 150, 200                   | 100                             |
|                     | max_depth            | 5, 10, 15, 20                       | 10                              |
|                     | min_samples_split    | 2, 5, 10                            | 10                              |
|                     | min_samples_leaf     | 1, 3, 5                             | 5                               |
|                     | class_weight         | None, balanced                      | balanced                        |
| Logistic Regression | penalty              | Literature-guided selection         | l2                              |
| SVM                 | Regularization (C)   | Preliminary experimental evaluation | Default value                   |
|                     | kernel               | linear, rbf                         | rbf                             |
|                     | C                    | 0.1, 0.3, 0.5, 1.0                  | 0.3                             |
|                     | gamma                | 0.01, 0.05, 0.1, 0.5                | 0.1                             |
|                     | class_weight         | None, balanced, custom weights      | {0:1.5, 1:1}                    |
| DTC                 | max_depth            | 2, 4, 6, 8, 10                      | 10                              |
|                     | min_samples_split    | 2, 4, 6, 8, 10                      | 7                               |
|                     | min_samples_leaf     | 2, 4, 6, 8, 10                      | 6                               |
|                     | class_weight         | None, balanced                      | balanced                        |
| AdaBoost            | n_estimators         | 50, 100, 150, 200                   | 150                             |
|                     | learning_rate        | 0.1, 0.3, 0.5, 1.0                  | 0.5                             |
|                     | base estimator depth | 4, 6, 8, 10                         | 10                              |
| ANN                 | Hidden layers        | 1–3 hidden layers                   | 2 hidden layers                 |
|                     | Neurons per layer    | 3–10 neurons                        | 6, 3                            |
|                     | Optimizer            | Adam, SGD                           | Adam                            |
|                     | Epochs               | 100, 150, 200                       | 200                             |
|                     | Activation function  | ReLU, Sigmoid                       | ReLU (hidden), Sigmoid (output) |

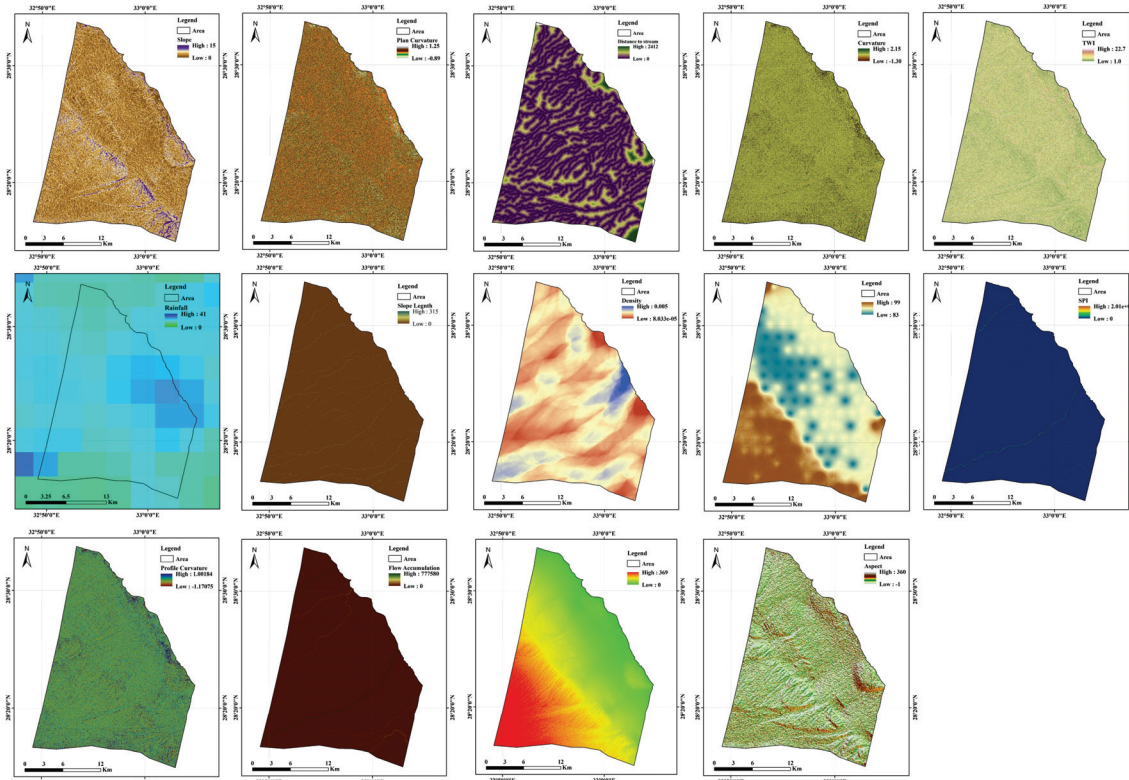
### 2.5. Flood Hazard Assessment Model

The flood hazard assessment in this study was carried out on two contrasting case studies, Wadi El-Darb and Wadi El-Allaqi, both of which have experienced runoff events triggered by extreme rainfall. Multiple significant rainfall events across Egypt were first identified, and satellite images corresponding to these events were downloaded and used as reference (observed) data. The spatial extent of the analysis was defined according to the coverage of the available satellite imagery rather than being restricted solely to the watershed boundaries, ensuring that the reference data captured the maximum possible flooded area during each event. These images were then used to verify whether specific data points were flooded or not.

A total of approximately 1600 sample points were extracted from the study areas, representing both flooded and non-flooded locations. Flooded samples were identified from the satellite-derived inundation maps, while non-flooded samples were selected from areas outside the detected flood extent. Care was taken to ensure that the dataset represents the spatial variability of the study areas and includes a balanced representation of both classes.

The dataset was constructed using samples extracted from the two selected flood events (28 October 2016 in Wadi El-Darb and 7–9 August 2024 in Wadi El-Allaqi), ensuring that the machine learning models were trained on diverse flood conditions.

To construct the training dataset, fourteen flood-influencing factors were selected: Topographic Wetness Index (TWI), Curve Number (CN), distance to stream, slope degree, precipitation, slope length, elevation, curvature, profile curvature, aspect, plan curvature, drainage density, Stream Power Index (SPI), and flow accumulation, as shown in Figures 3 and 4.

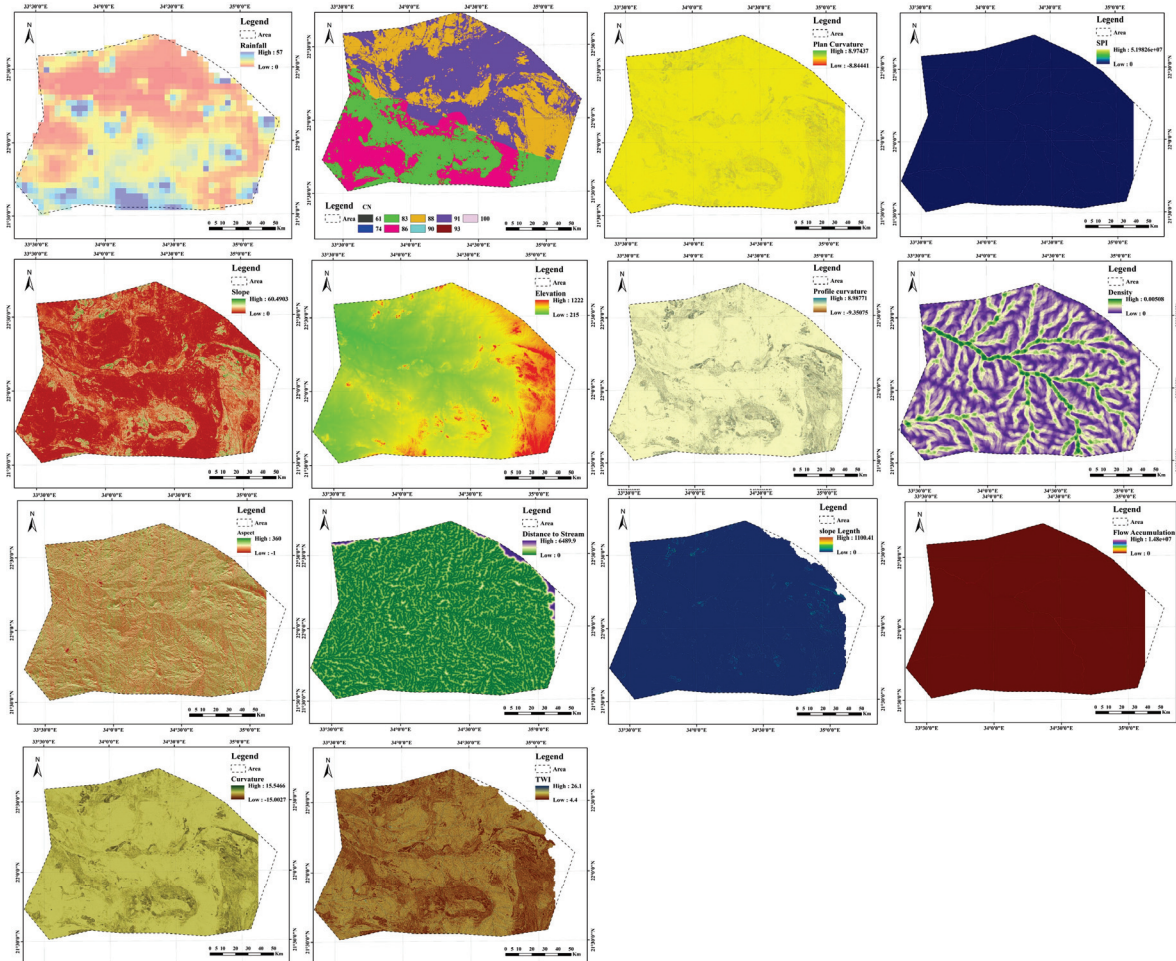


**Figure 3.** Explanatory factors of Wadi El-Darb. Spatial distribution of the conditioning factors used for flood susceptibility modeling. Each raster layer represents a predictor variable from which values were extracted at the sampled flood and non-flood locations. The extracted values were combined to create the feature dataset used for machine learning model training and testing.

Although lithology, bedrock permeability, and geological structures can significantly influence runoff generation, these variables were beyond the scope of the current satellite-based framework because consistent geological datasets with comparable spatial resolution were not available for all study and validation basins. Instead, infiltration-related characteristics were represented indirectly through the Curve Number (CN), which incorporates Hydrologic Soil Groups and land-use information. Therefore, CN reflects part of the hydrological response associated with infiltration capacity but should not be considered a direct representation of geological conditions.

The flooded and non-flooded points were assigned values of 1 and 0, respectively. The dataset was then randomly split into training (70%) and testing (30%) subsets, corresponding to approximately 1120 samples for training and 480 samples for testing. The split was performed while maintaining a representative distribution of flooded and non-flooded samples in both subsets to ensure robust model evaluation, ensuring consistency by applying the same data partitioning across all machine learning models. Several supervised machine

learning algorithms were applied, including Random Forest, Support Vector Machine, Decision Tree Classifier, Artificial Neural Network, AdaBoost Classifier, and Logistic Regression. Model performance was evaluated by comparing predicted flood-prone locations with the observed inundation maps derived from satellite imagery of the selected rainfall events.



**Figure 4.** Explanatory factors of Wadi El-Allaky. Spatial distribution of the conditioning factors used for flood susceptibility modeling. Each raster layer represents a predictor variable from which values were extracted at the sampled flood and non-flood locations. The extracted values were combined to create the feature dataset used for machine learning model training and testing.

This workflow enabled the systematic testing of multiple algorithms and input factors across two contrasting watersheds, providing insights into the predictive capacity of machine learning approaches for flood hazard mapping in arid regions of Egypt.

The machine learning methodology adopted in this study follows a systematic sequence of steps to ensure the robustness and reliability of the developed models. First, all training data were loaded and compiled into a unified dataset, containing both the flood-influencing factors and the binary target variable representing flooded and non-flooded locations. The dataset was then split into two subsets: 70% for training and 30% for testing, ensuring a balanced distribution of flooded and non-flooded points.

Pre-processing of the data was performed prior to model training. This included handling missing values, which were filled with zeros to maintain consistency and prevent bias in model learning. Following pre-processing, the models were trained using selected



algorithms, namely Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), AdaBoost, Decision Tree Classifier (DTC), and Logistic Regression (LR).

After training, the models were evaluated on both the training and testing subsets to assess their predictive performance. To further enhance model accuracy, fine-tuning of hyperparameters was conducted. For instance, in the case of the Decision Tree Classifier, optimal settings such as `max_depth = 4`, `min_samples_leaf = 2`, `min_samples_split = 6`, and `random_state = 42` were tested. Similarly, for the Support Vector Machine, the classifier was tuned with parameters including `C = 30`, `degree = 2`, `kernel = 'poly'`, and `random_state = 42`.

Finally, the testing accuracy of each model was evaluated after fine-tuning, and the comparative performance of all algorithms was analyzed to identify the most suitable machine learning approach for flood hazard prediction.

It should be noted that the spatial extent of the analysis was defined based on the coverage of available satellite imagery corresponding to the selected flood events. In some cases, the satellite image footprint was either larger or smaller than the exact watershed boundaries. This approach was adopted to maximize the use of observed flood information derived from remote sensing, particularly in data-scarce regions.

Despite this, the modeling framework maintains hydrological consistency, as all conditioning factors were derived based on terrain and watershed characteristics, including flow direction, flow accumulation, drainage networks, and distance to the main stream. These variables inherently represent the hydrological behavior of the basin and control runoff generation and flood processes.

Therefore, while the spatial extent was influenced by data availability, the analysis remains physically meaningful and representative of the hydrological system.

### 3. Results

#### 3.1. Model Evaluation Method

In this study, the performance of the machine learning models was evaluated using four widely adopted metrics: accuracy (ACC), precision (P), recall (R), and F1-score (F). Accuracy measures the overall proportion of correctly classified points, including both flooded and non-flooded locations. Precision evaluates the ability of the model to correctly identify flood-prone points among all points predicted as flooded. Recall measures the proportion of actual flooded points that were correctly identified by the model. The F1-score, which is the harmonic mean of precision and recall, provides a balanced measure of the model's ability to detect flood-prone locations.

The metrics were calculated according to the following equations:

$$Accuracy(ACC) = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$Precision(P) = \frac{TP}{TP + FP} \quad (7)$$

$$Recall(R) = \frac{TP}{TP + FN} \quad (8)$$

$$F1 - score(F) = 2 * \frac{P * R}{P + R} \quad (9)$$

where TP (true positives) are correctly predicted flooded points, FP (false positives) are non-flooded points incorrectly predicted as flooded, TN (true negatives) are correctly predicted non-flooded points, and FN (false negatives) are flooded points incorrectly predicted as non-flooded.

This framework ensured a robust and objective evaluation of the predictive capabilities of the applied machine learning algorithms.

The evaluation of six machine learning algorithms—Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree Classifier (DTC), AdaBoost, and Artificial Neural Networks (ANNs)—was carried out using four metrics: precision, recall, F1-score, and overall accuracy.

### 3.2. Models Comparison

Among the models, Random Forest and Decision Tree Classifier achieved the highest accuracy (0.94), with both models demonstrating balanced precision and recall across the flooded and non-flooded classes. RF, in particular, maintained high stability (precision = 0.96, recall = 0.93 for flooded points), confirming its robustness in handling complex datasets.

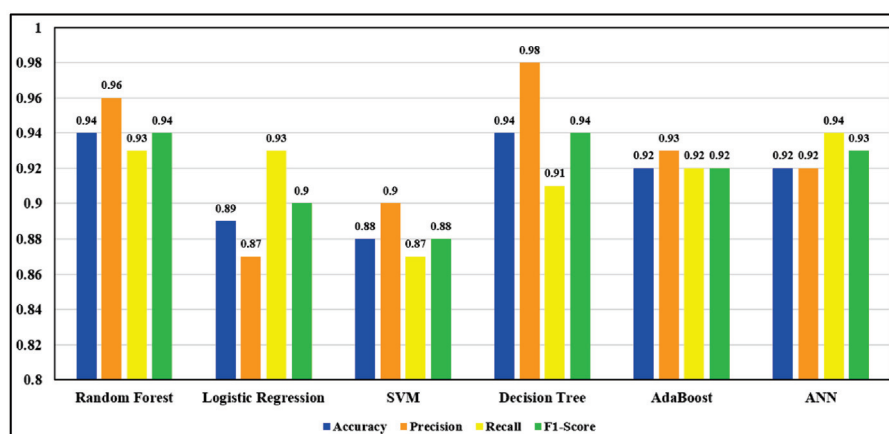
The ANN model also performed strongly (accuracy = 0.92), with an F1-score of 0.927, indicating reliable detection of flood-prone areas. Similarly, AdaBoost achieved consistent results (accuracy = 0.92), though its metrics were slightly lower than those of ANN.

Logistic Regression and SVM yielded slightly lower accuracies (0.89 and 0.88, respectively), yet they still provided competitive results and demonstrated good generalization ability. LR was particularly effective at capturing flooded points (recall = 0.93), while SVM maintained balance across classes.

Overall, based on accuracy and stability across all metrics, Random Forest emerges as the most suitable model for flood hazard prediction, followed closely by the Decision Tree Classifier and ANN. These models demonstrate strong potential for application in large-scale flood prediction studies.

To present these findings more effectively, the following figure, number 5, shows a comparison between all the accuracy algorithms.

It is important to note that all performance metrics presented in this study, including accuracy, fine-tuning, recall, and the F1 metric (Figure 5), were calculated based on the test dataset, which represents 30% of the total data. The remaining 70% of the dataset was used to train the machine learning models.



**Figure 5.** Model performance comparison: Bar chart comparing accuracy, precision, recall, and F1-score across all algorithms.

The total dataset comprises approximately 1600 sample points, including both flooded and unflooded locations. A balanced representation of both categories was ensured during dataset construction to minimize potential bias and guarantee a reliable assessment of model performance.



To better understand the role of each conditioning factor in flood susceptibility prediction, the relative importance and direction of influence of all input variables were analyzed across the applied machine learning models. Each algorithm evaluates the contribution of predictors differently, but together they provide a comprehensive view of the mechanisms driving flooding.

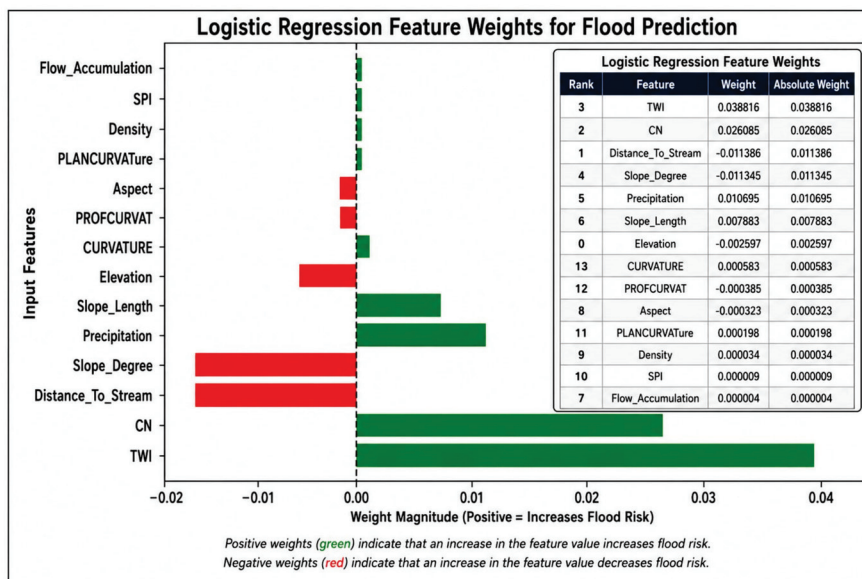
The feature importance analysis revealed that some conditioning factors were consistently assigned higher predictive importance by several machine learning algorithms, whereas others received relatively lower importance. These rankings provide insight into how the models utilized the available predictors during flood susceptibility estimation, rather than establishing causal hydrological mechanisms. For example, Support Vector Machines (SVMs) assigned a strong negative weight to distance to main stream, indicating that shorter distances were consistently associated with higher predicted flood susceptibility within the SVM model, while larger distances reduced it. Similarly, density showed a high positive weight, indicating that density received relatively high predictive importance in distinguishing flooded from non-flooded locations. Logistic Regression assigned relatively high positive coefficients to slope degree and precipitation, indicating a positive association with predicted flood susceptibility within the fitted model, whereas several curvature-related variables received lower or negative coefficients within the model. Random Forest and Decision Tree classifiers consistently highlighted distance to stream and curve number (CN) as dominant predictive variables, while topographic indices such as slope length and curvature played smaller, sometimes negative roles. Artificial Neural Networks (ANNs) distributed influence across multiple variables, but reinforced the importance of distance to stream and CN as consistently important predictors across the trained models.

Overall, the models not only identified the most influential predictors (distance to stream, CN, density, and elevation) but also indicated whether each variable exhibited a positive or negative association with the predicted flood susceptibility within the corresponding machine learning model. This dual interpretation adds depth to the flood hazard analysis, highlighting how different machine learning algorithms assigned varying predictive importance and association directions to the conditioning factors. The following sub-figures (Figure 6) illustrate these findings, showing the factor importance and the direction of their influence for each machine learning model.

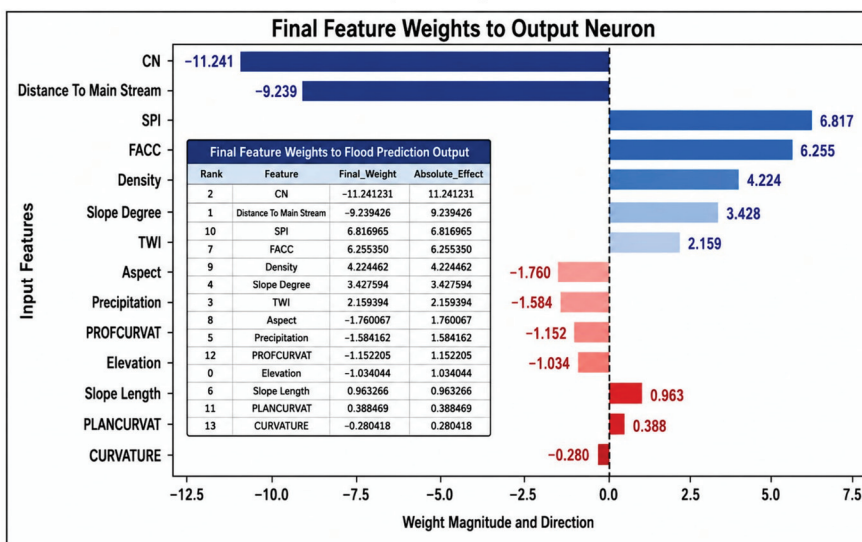
It is important to note that the feature importance measures derived from different machine learning algorithms cannot be directly compared in a quantitative sense, as they rely on different underlying mechanisms. For example, regression-based models (such as Logistic Regression and Support Vector Machines) provide coefficients indicating the direction of influence of each variable, while tree-based models (such as Random Forests and decision trees) estimate the relative importance of variables based on their contribution to decision splits.

It should be noted that feature importance measures differ across machine learning algorithms, and direct comparison between models should be treated qualitatively. Consequently, the reported importance rankings reflect general trends rather than absolute or directly comparable values.

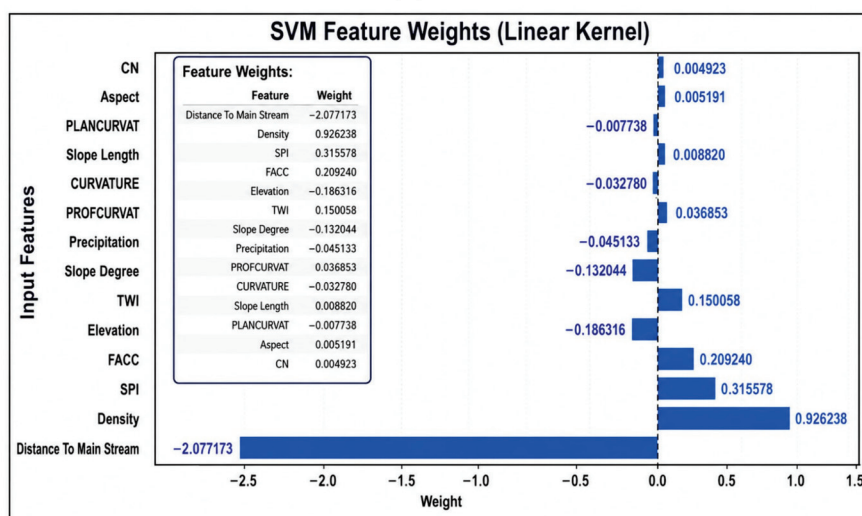
Furthermore, because several conditioning factors are derived from the same digital elevation model, some degree of multicollinearity is unavoidable despite the inclusion of the VIF analysis. Consequently, the relative importance assigned to individual variables may partly reflect shared predictive information among correlated terrain attributes and the mathematical structure of each machine learning algorithm. Therefore, the reported feature importance rankings should be interpreted primarily as model-specific measures of predictive contribution rather than independent evidence of the physical influence of individual factors on flood occurrence.



(A) Logistic Regression

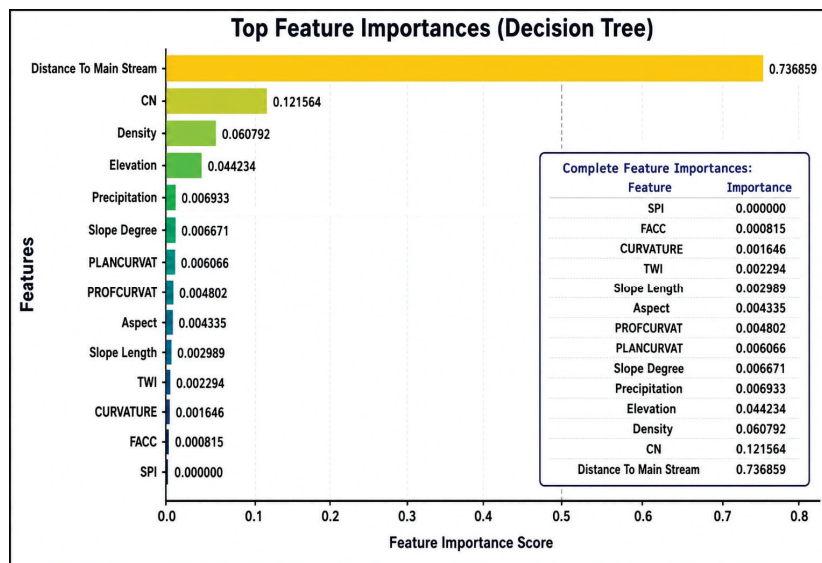


(B) ANN

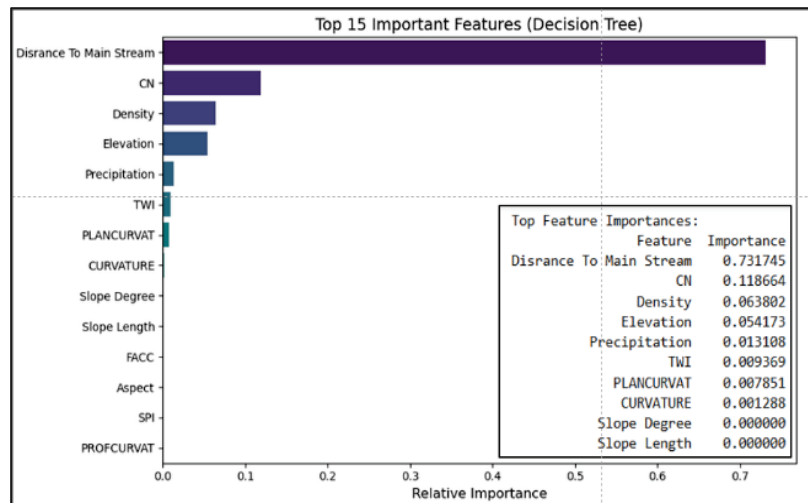


(C) SVM

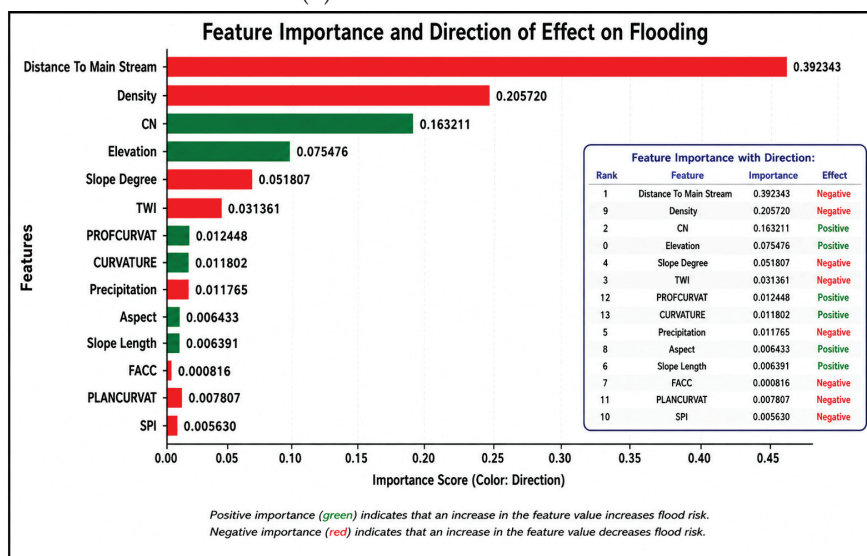
Figure 6. Cont.



(D) AdaBoost



(E) Decision Tree Classifier



(F) Random Forest

**Figure 6.** Factor importance and the direction of their influence across the implemented machine learning models: (A) Logistic Regression, (B) Artificial Neural Network (ANN), (C) Support Vector Machine (SVM), (D) AdaBoost, (E) Decision Tree Classifier, and (F) Random Forest.

To complement the graphical outputs, Table 4 summarizes the dominant role of each conditioning factor across the different machine learning models. The table shows whether each variable generally exhibited a positive or negative association with predicted flood susceptibility. This comparison highlights the consistency of certain predictors (e.g., distance to stream and curve number) and the variable behavior of others (e.g., curvature indices). This tabular synthesis reinforces that distance to stream, curve number, density, and elevation are the most consistently important predictors identified by the evaluated machine learning models across models, whereas topographic curvature variables generally received lower predictive importance or negative associations depending on the algorithm.

**Table 4.** Influence of each input factor across the evaluated machine learning algorithms.

| Factor             | SVM      | Logistic Regression | Random Forest/DTC | ANN      | General Predictive Interpretation                          |
|--------------------|----------|---------------------|-------------------|----------|------------------------------------------------------------|
| Distance to Stream | Negative | Positive            | Positive          | Positive | Consistently important predictor                           |
| Curve Number (CN)  | Positive | Positive            | Positive          | Positive | Frequently assigned high predictive importance             |
| Density            | Positive | Neutral/Positive    | Positive          | Positive | Generally, positively associated with predicted flooding   |
| Elevation          | Negative | Mixed               | Positive          | Positive | Positive association varies among models                   |
| Slope Degree       | Negative | Positive            | Mixed             | Mixed    | Association depends on the algorithm                       |
| Precipitation      | Negative | Positive            | Positive          | Mixed    | Frequently associated with higher predicted susceptibility |
| TWI                | Positive | Negative            | Mixed             | Mixed    | Role varies by algorithm                                   |
| Slope Length       | Positive | Negative            | Negative          | Negative | Longer slopes may spread runoff                            |
| Curvature          | Negative | Negative            | Mixed             | Mixed    | Often mitigating effect                                    |
| Plan Curvature     | Negative | Neutral             | Negative          | Mixed    | Limited but usually reducing                               |
| Profile Curvature  | Negative | Neutral             | Neutral           | Mixed    | Minor influence                                            |
| SPI                | Positive | Neutral             | Neutral           | Positive | Frequently assigned positive importance                    |
| Flow Accumulation  | Positive | Neutral             | Neutral           | Positive | Frequently assigned positive importance                    |
| Aspect             | Positive | Positive            | Positive          | Negative | Orientation influences runoff patterns                     |

Accordingly, the feature importance analysis is intended to improve understanding of model behavior and facilitate comparison among different machine learning algorithms rather than to establish independent hydrological causality or quantify the physical contribution of individual conditioning factors.

To further evaluate the spatial performance of the developed models, flood susceptibility maps were generated for the Ras Ghareb case study using a sample of different machine learning algorithms. These maps illustrate the spatial distribution of flood-prone areas classified into different susceptibility levels, providing a visual comparison of model outputs.



For the Ras Ghareb area, generated using different machine learning algorithms. The maps show the spatial distribution of flood susceptibility classes ranging from low to very high, as shown in Figure 7.

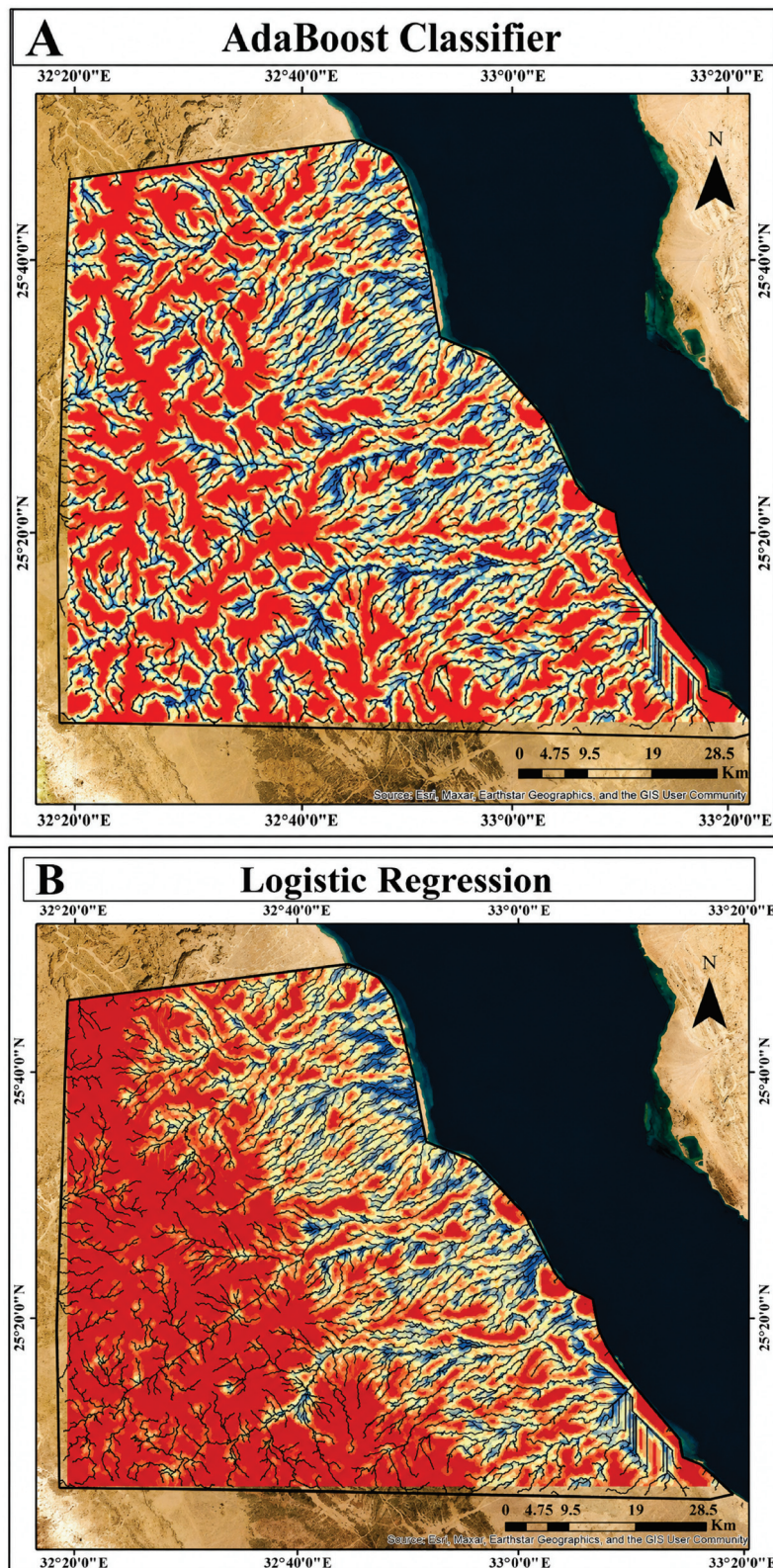
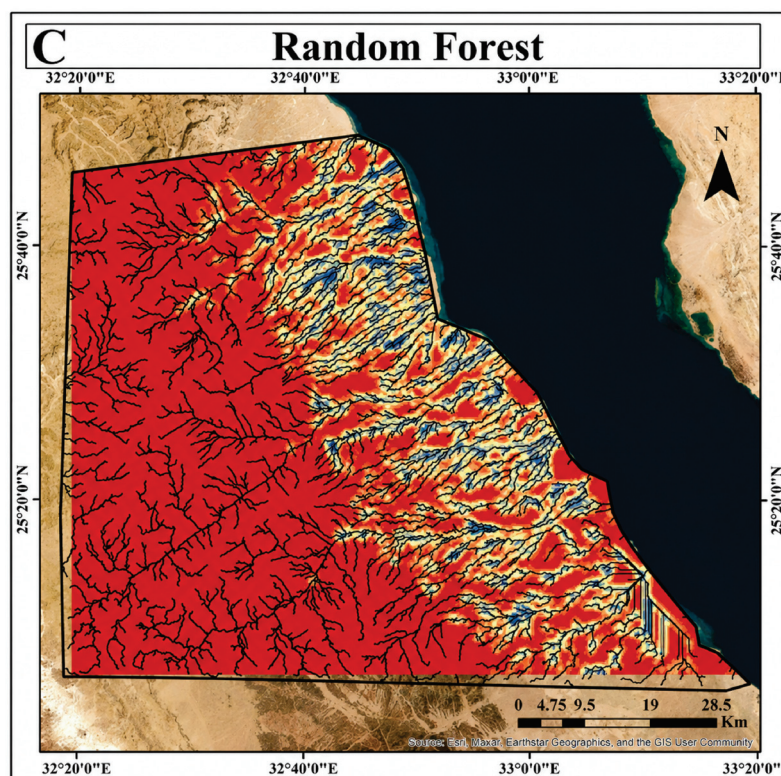


Figure 7. Cont.



**Figure 7.** Flood susceptibility maps of the Ras Ghareb area generated using three different machine learning algorithms: (A) AdaBoost Classifier, (B) Logistic Regression, and (C) Random Forest. Blue areas represent predicted flooded (flood-susceptible) locations, whereas red areas represent predicted non-flooded (non-susceptible) locations.

#### 4. Discussion

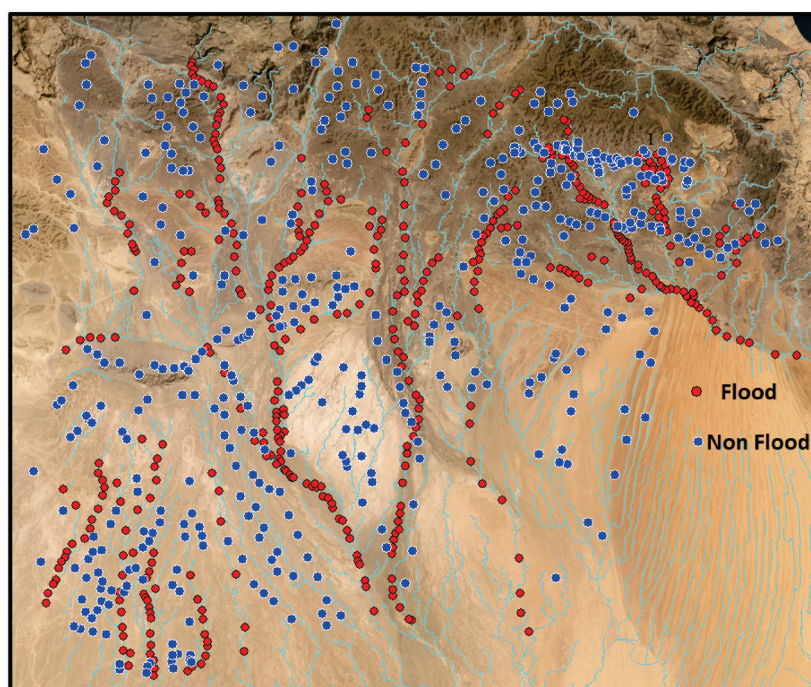
To further evaluate the robustness and generalizability of the developed machine learning models, an additional case study was introduced outside the primary training domain. After identifying the algorithms that achieved the highest predictive performance among the six tested models, the selected methods were first applied to a new watershed, Wadi Hodein, located in southeastern Egypt, using an extreme rainfall event that generated noticeable surface runoff as a test case. Furthermore, the models were evaluated on a major rainfall and flash-flood event that occurred across parts of Oman from 14 to 15 April 2024, during which more than 90 mm of rain was recorded in Al Mudhaibi (North Al Sharqiyah Governorate), and widespread inundation affected wadis such as Wadi Bani Khalid and Samad Al Shan. For both events, satellite-derived flood inundation maps were employed as observed reference data to validate the model predictions. This dual evaluation provided an initial assessment of model transferability across distinct climatic and geomorphological settings. The results showed that while several machine learning algorithms maintained satisfactory predictive performance, others experienced noticeable reductions in accuracy, indicating that transferability varies among algorithms and is influenced by regional environmental conditions. These findings highlight both the potential and the limitations of applying satellite-driven machine learning approaches beyond the calibration watersheds.

This dataset was not part of the model training process and thus represented unseen data, providing an initial assessment of model transferability across different watersheds. The performance of each algorithm was tested against this independent dataset to assess how well models trained on Wadi El-Darb and Wadi El-Allaqi could predict flood-prone locations in a hydrologically and topographically different environment. The accuracy



metrics derived from this analysis were used to evaluate the extent to which the selected algorithms could generalize beyond the initial case studies. Rather than confirming universal applicability, the external validation provides evidence of the models' ability to generalize under comparable environmental conditions while also revealing differences in performance among algorithms.

For the Oman case study, a total of 889 sample points were used to evaluate the model performance, including 439 flooded points and 450 non-flooded points, as shown in Figure 8. The samples were spatially distributed across the study area to ensure balanced representation of both classes and to reduce spatial bias. Flooded points were primarily concentrated along drainage pathways, while non-flooded points were selected from areas away from the main channels and elevated terrains to ensure a representative dataset for validation.



**Figure 8.** Spatial distribution of the validation sample points used in the Oman case study. Blue points indicate flooded locations, whereas red points indicate non-flooded locations.

This validation step strengthens the practical significance of the research by providing initial evidence that the trained models may be applicable to other ungauged basins with comparable hydrological characteristics. At the same time, the variability observed among model performances indicates that additional validation is required before broader transferability can be established.

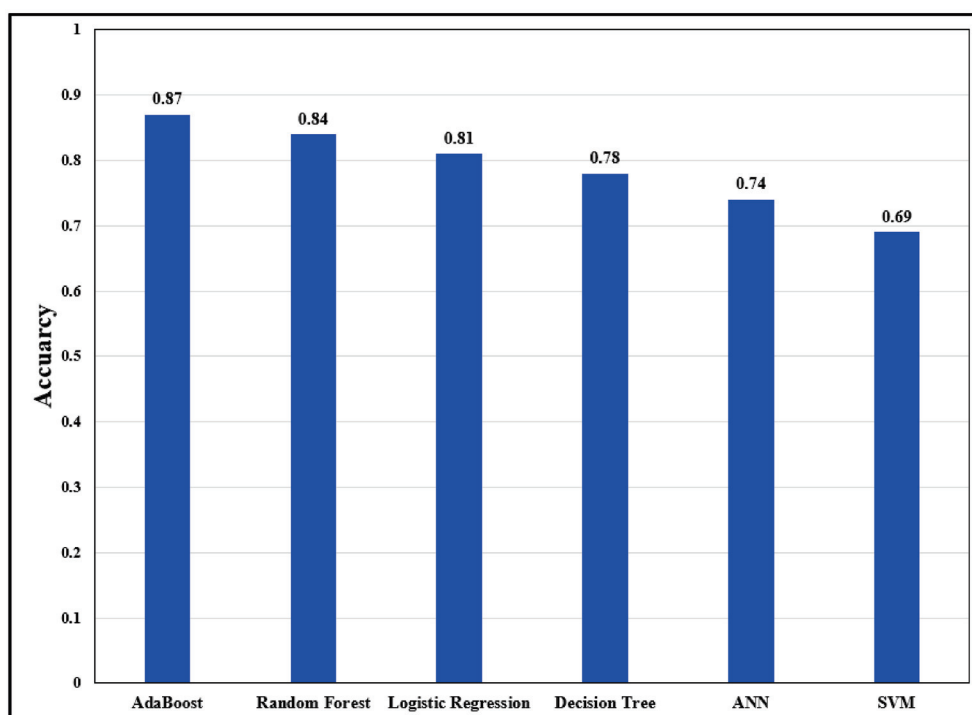
Although the validation watersheds differ from the training basins in their local geological settings, they exhibit similar hydrological behavior characteristic of arid environments, including ephemeral drainage networks, limited vegetation cover, and rapid runoff following high-intensity rainfall. The satisfactory predictive performance obtained in these independent basins suggests that the selected topographic, hydrological, and rainfall predictors capture the dominant controls on flood generation despite variations in geological conditions. Nevertheless, incorporating explicit geological variables, such as lithology or bedrock permeability, could further improve model transferability and represents an important direction for future research.

It is important to note that the models were applied directly to these external case studies without any retraining. This approach represents a spatial transferability test, where the models trained on Wadi El-Darb and Wadi El-Allaqi were used to predict flood-prone areas in entirely different watersheds. The independent validation results from Wadi Hodein and Oman provide encouraging evidence of the generalization capability of the trained models. However, the observed variability in predictive performance among the tested algorithms indicates that transferability should not be considered uniform across all machine learning methods. Instead, the results suggest that successful application depends on both the selected algorithm and the similarity between the calibration and target watersheds.

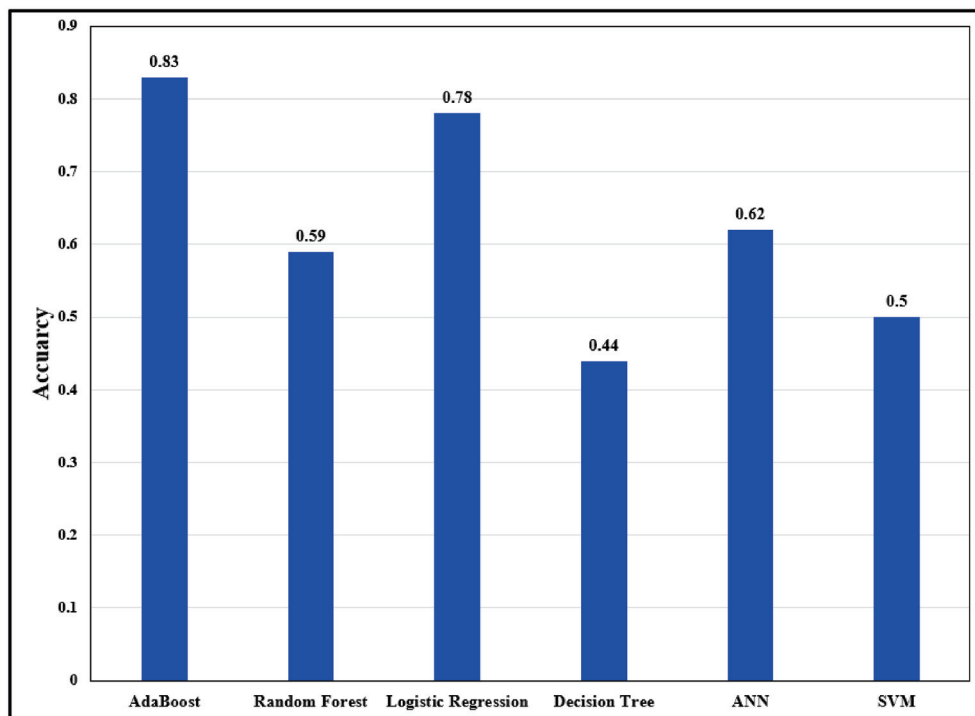
Although Wadi El-Darb, Wadi El-Allaqi, Wadi Hodein, and the Oman validation basin exhibit differences in lithology and local geological composition, they share several important hydrological characteristics typical of arid environments. All four watersheds are characterized by ephemeral drainage systems, sparse vegetation cover, steep mountainous terrain, and flash floods generated primarily by short-duration, high-intensity rainfall events. These common hydrological characteristics promote rapid runoff generation despite local variations in bedrock type and permeability. Consequently, the satisfactory performance of the trained models in Wadi Hodein and Oman suggests that the selected topographic, hydrological, and rainfall-related predictors capture the dominant controls on flood generation across these arid watersheds. Nevertheless, the inclusion of explicit geological variables, such as lithology or bedrock permeability, may further improve model transferability and should be considered in future studies.

Although a random 70/30 train–test split was adopted, it is acknowledged that spatial autocorrelation may exist due to the clustered nature of flood occurrences. This may introduce some degree of spatial dependence between training and testing samples.

The independent validation results showed accuracies ranging from 69–87% for Wadi Hodein (Figure 9) and 44–83% for the Oman flood event (Figure 10), demonstrating varying levels of model transferability beyond the training domain.



**Figure 9.** Comparison of model accuracy for each machine learning algorithm in Wadi Hodein.



**Figure 10.** Comparison between models' accuracy for each algorithm in the Oman flood case study (April 2024 rainfall event).

The quantitative evaluation of the external validation experiments is summarized in Table 5 using four performance metrics: accuracy, precision, recall, and F1-score. Overall, the results demonstrate that the predictive performance of the machine learning models varies across the two independent validation basins, reflecting differences in their environmental and hydrological characteristics.

**Table 5.** Detailed quantitative overall accuracy (%) of the six machine learning algorithms across the validation settings.

| Case Study          | Metric    | AdaBoost | Random Forest | Logistic Regression | Decision Tree | ANN | SVM  |
|---------------------|-----------|----------|---------------|---------------------|---------------|-----|------|
| Wadi Hodein (Egypt) | Accuracy  | 87%      | 84%           | 81%                 | 78%           | 46% | 46%  |
|                     | Precision | 98%      | 100%          | 98%                 | 91%           | 46% | 45%  |
|                     | Recall    | 80%      | 73%           | 70%                 | 69%           | 50% | 50%  |
|                     | F1-Score  | 88%      | 85%           | 82%                 | 79%           | 48% | 47%  |
| Oman                | Accuracy  | 83%      | 59%           | 79%                 | 45%           | 63% | 51%  |
|                     | Precision | 100%     | 100%          | 100%                | 34%           | 99% | 100% |
|                     | Recall    | 67%      | 17%           | 57%                 | 13%           | 25% | 0%   |
|                     | F1-Score  | 80%      | 29%           | 73%                 | 19%           | 39% | 1%   |

For Wadi Hodein, AdaBoost achieved the best overall performance, with an accuracy of 87%, precision of 98%, recall of 80%, and an F1-score of 88%, indicating a well-balanced ability to correctly identify both flooded and non-flooded locations. Random Forest also performed strongly, achieving an accuracy of 84% and a high F1-score of 85%, followed by Logistic Regression and Decision Tree Classifier, which maintained satisfactory performance across all evaluation metrics. In contrast, ANN and SVM showed substantially

lower performance, with accuracies of 46% and comparatively low precision, recall, and F1-scores, suggesting limited generalization capability for this independent watershed.

The external validation results for the Oman case study revealed greater variability among the evaluated algorithms. AdaBoost again demonstrated the most robust performance, achieving the highest accuracy (83%) and an F1-score of 80%, confirming its superior transferability to a geographically distinct watershed. Logistic Regression also maintained satisfactory predictive capability, with an accuracy of 79% and an F1-score of 73%. Although Random Forest achieved perfect precision (100%), its recall decreased substantially to 17%, resulting in a relatively low F1-score (29%). This indicates that the model produced very few false-positive predictions but failed to identify a considerable proportion of the actual flooded locations. Similarly, Decision Tree Classifier, ANN, and SVM exhibited lower recall and F1-score values; these findings indicate that while AdaBoost and Logistic Regression maintained satisfactory predictive performance during the Oman validation, other algorithms exhibited substantial reductions in recall and F1-score, highlighting important differences in model transferability across independent watersheds.

Overall, the external validation demonstrates that ensemble-based approaches, particularly AdaBoost, provide the most reliable spatial transferability across independent arid watersheds. The observed reduction in model performance in the Oman case study, compared with Wadi Hodein, likely reflects differences in geological setting, terrain characteristics, and hydrological response between the training and validation basins. Nevertheless, the satisfactory performance of the best-performing models indicates that the selected topographic, hydrological, and rainfall-related predictors successfully capture the dominant controls on flood generation and support the applicability of the proposed framework for flood susceptibility assessment in ungauged arid environments.

The observed transferability of the proposed framework is consistent with recent studies highlighting the importance of external validation for assessing the robustness of machine learning models beyond their calibration domains [32]. Previous research has demonstrated that machine learning models can successfully generalize to independent flood events when the dominant hydrological processes are adequately represented, while also emphasizing that predictive performance is influenced by regional environmental variability [32]. Likewise, recent remote sensing-based flood assessment studies have shown that integrating satellite-derived environmental variables with advanced machine learning algorithms improves the reliability and operational applicability of flood susceptibility mapping across different geographical settings [21,23]. The present findings are in agreement with these observations, as the AdaBoost model consistently achieved the highest predictive performance in both the Wadi Hodein and Oman validation basins despite differences in geological and environmental conditions. This suggests that the selected conditioning factors effectively capture the dominant controls governing flash-flood generation in arid environments, while reinforcing the importance of external validation for evaluating model robustness and practical applicability.

The independent validation conducted in Oman provides valuable insight into the generalization capability of the developed models. While several algorithms maintained satisfactory predictive performance, others exhibited substantial reductions in accuracy compared with their performance in the calibration watersheds. These differences indicate that model transferability is influenced by regional variations in topography, hydrological response, rainfall characteristics, and environmental conditions, as well as by the inherent characteristics of individual machine learning algorithms. Consequently, the present findings should be interpreted as providing encouraging initial evidence of transferability under comparable environmental conditions, rather than demonstrating universal applicability.

bility. Additional validation across a broader range of watersheds and climatic settings is therefore recommended to further assess the robustness and generalization capability of the proposed framework.

Although the proposed machine learning models demonstrated strong performance in predicting flood inundation areas, several limitations should be acknowledged. First, the study relies primarily on satellite-derived inundation maps from a limited number of historical rainfall events, which may introduce uncertainties due to sensor resolution, cloud contamination, and potential classification errors in reference data. Second, the training dataset is constrained by the availability of labeled flood events, which may limit the ability of the models to fully capture the complex hydrological variability across different climatic and geomorphological conditions. Third, although the independent validation demonstrated encouraging generalization capability, the observed differences in model performance between Wadi Hodein and Oman indicate that transferability varies among machine learning algorithms and environmental settings. Therefore, additional testing across a broader range of watersheds remains necessary before the proposed framework can be considered broadly applicable. Further testing across a wider range of watersheds is still required to ensure stronger spatial generalization and robustness under extreme or rare hydrological conditions. Future research should focus on integrating multi-source data, including high-resolution remote sensing, real-time rainfall measurements, and physically based hydrological variables, to improve model reliability. In addition, emerging deep learning architectures, such as convolutional neural networks (CNNs) and hybrid physics-informed machine learning models, present promising directions for enhancing spatial feature extraction and improving predictive accuracy. The incorporation of real-time data assimilation and transfer learning techniques could further support the development of scalable, transferable flood prediction systems capable of supporting early warning systems and disaster risk reduction strategies in data-scarce regions.

The reported model performance should therefore be interpreted with caution, as the use of random data splitting may lead to optimistic estimates of accuracy. Future work could incorporate spatial cross-validation or block-based validation techniques to better account for spatial dependence.

## 5. Conclusions

This study demonstrated the potential of machine learning for accurately predicting flood inundation locations in data-scarce environments. By combining rainfall events with remote sensing imagery, two contrasting watersheds—Wadi El-Darb, which drains eastward toward Ras Gharib, and Wadi El-Allaqi, which drains westward into Lake Nasser—were used as case studies to construct robust training datasets. Fourteen flood-influencing factors derived from global datasets and DEM-based analyses were integrated as inputs, and six machine learning algorithms (RF, LR, SVM, DTC, AdaBoost, and ANN) were tested.

The results showed that Random Forest (RF) and Decision Tree Classifier (DTC) achieved the highest accuracy (94%), closely followed by ANN and AdaBoost (92%), while Logistic Regression (89%) and SVM (88%) provided slightly lower yet reliable performance. Evaluation metrics (accuracy, precision, recall, and F1-score) confirmed that ensemble-based and tree-based models offered more stable and consistent results across both case studies.

To further assess model transferability and robustness, the top-performing algorithms were validated on two independent, unseen rainfall events outside the training domain. The first was in Wadi Hodein, southeastern Egypt, which experienced a localized extreme rainfall event that produced significant surface runoff. The second was the major flash-



flood event that occurred across parts of Oman from 14–15 April 2024, during which over 90 mm of rain was recorded in Al Mudhaibi (North Al Sharqiyah Governorate), causing widespread inundation across wadis such as Wadi Bani Khalid and Samad Al Shan. In both cases, satellite-derived flood extent maps were used as observational references.

The external validation results demonstrated that several models, particularly Adaboost and Logistic Regression, maintained satisfactory predictive performance when applied to independent flood events outside the training watersheds, whereas other algorithms exhibited substantial reductions in recall and F1-score, especially during the Oman validation. These findings indicate that the proposed framework has promising potential for application in comparable hydrological settings, while also highlighting that model transferability may vary according to regional environmental and geomorphological characteristics.

Overall, this research demonstrates the effectiveness of machine learning techniques for flood hazard assessment in arid and semi-arid regions, particularly where hydrometeorological observations are limited. The proposed framework achieved reliable performance in the investigated case studies and showed encouraging results during independent external validation, suggesting its potential applicability to watersheds with comparable hydrological and environmental characteristics.

While the results are encouraging, the extension of the methodology to large-scale domains such as the Nile Basin or the Arabian Peninsula should be considered as a potential future research direction, as it requires further validation across a wider range of climatic and geomorphological conditions.

In principle, the framework could be integrated with rainfall forecast data to support near-real-time flood prediction, since rainfall represents the primary dynamic driver of flood generation, while most terrain-related factors remain relatively stable over time. However, such an operational implementation was not explored in the present study and remains a direction for future work.

These findings contribute to improving flood susceptibility mapping in data-limited arid and semi-arid regions and provide a comparative assessment of machine learning performance across different hydrological settings. The results also help to better understand the influence of key conditioning factors on flood occurrence and highlight the potential of satellite-based datasets for supporting data-driven flood hazard studies. Future work may further explore extensions toward operational applications and early warning systems.

The application of the trained models to independent case studies provides an initial indication of their potential applicability to new regions; however, further validation across diverse environments is required to fully assess model transferability.

**Author Contributions:** M.M. Conceptualization, methodology, data analysis, writing—original draft. D.A., M.A.H. and A.M.E. Supervision, review, and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data supporting the findings of this study are available from publicly accessible sources. Digital Elevation Model (DEM) data were obtained from the USGS EarthExplorer (<https://earthexplorer.usgs.gov/>, 3 July 2026). Precipitation data were obtained from the Climate Hazards Center InfraRed Precipitation with Station data (CHIRPS) (<https://www.chc.ucsb.edu/data/chirps>, 3 July 2026). Land use/land cover and soil datasets were obtained from the sources cited in the manuscript. The processed flood inventory, derived conditioning factors, and machine learning datasets generated during this study are available from the corresponding author upon reasonable request.

**Conflicts of Interest:** The authors declare that they have no known financial or personal relationships that could have appeared to influence the work reported in this paper. On behalf of all authors, the corresponding author confirms that there is no conflict of interest.

## References

1. Ward, P.J.; Jongman, B.; Aerts, J.C.J.H.; Bates, P.D.; Botzen, W.J.W.; Loaiza, A.D.; Hallegatte, S.; Kind, J.M.; Kwadijk, J.; Scussolini, P.; et al. A global framework for future costs and benefits of river-flood protection in urban areas. *Nat. Clim. Change* **2017**, *7*, 642–646. [CrossRef]
2. Woodruff, J.D.; Irish, J.L.; Camargo, S.J. Coastal flooding by tropical cyclones and sea-level rise. *Nature* **2013**, *504*, 44–52. [CrossRef] [PubMed]
3. Lian, J.; Xu, H.; Xu, K.; Ma, C. Optimal management of the flooding risk caused by the joint occurrence of extreme rainfall and high tide level in a coastal city. *Nat. Hazards* **2017**, *89*, 183–200. [CrossRef]
4. Tellman, B.; Sullivan, J.A.; Kuhn, C.; Kettner, A.J.; Doyle, C.S.; Brakenridge, G.R.; Erickson, T.A.; Slayback, D.A. Satellite imaging reveals increased proportion of population exposed to floods. *Nature* **2021**, *596*, 80–86. [CrossRef] [PubMed]
5. Panhalkar, S.S.; Jarag, A.P. Flood risk assessment of Panchganga River (Kolhapur district, Maharashtra) using GIS-based multicriteria decision technique. *Curr. Sci.* **2017**, *112*, 785–793. [CrossRef]
6. Trenberth, K.E. Changes in precipitation with climate change. *Clim. Res.* **2011**, *47*, 123–138. [CrossRef]
7. Alfieri, L.; Bisselink, B.; Dottori, F.; Naumann, G.; de Roo, A.; Salamon, P.; Wyser, K.; Feyen, L. Global projections of river flood risk in a warmer world. *Earth's Future* **2017**, *5*, 171–182. [CrossRef]
8. Mosavi, A.; Ozturk, P.; Chau, K.-W. Flood Prediction Using Machine Learning Models: Literature Review. *Water* **2018**, *10*, 1536. [CrossRef]
9. Costache, R. Flash-flood Potential Index mapping using weights of evidence, decision Trees models and their novel hybrid integration. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 1375–1402. [CrossRef]
10. Benito, G.; Thorndycraft, V.R. *Systematic, Palaeoflood and Historical Data for the Improvement of Flood Risk Estimation: Methodological Guidelines*; Consejo Superior de Investigaciones Científicas, Centro de Ciencias Medioambientales: Madrid, Spain, 2004.
11. Apel, H.; Thielen, A.H.; Merz, B.; Blöschl, G. Flood risk assessment and associated uncertainty. *Nat. Hazards Earth Syst. Sci.* **2004**, *4*, 295–308. [CrossRef]
12. Burke, J.J.; Pricope, N.G.; Blum, J. Thermal imagery-derived surface inundation modeling to assess flood risk in a flood-pulsed Savannah watershed in Botswana and Namibia. *Remote Sens.* **2016**, *8*, 676. [CrossRef]
13. Donnelly, J.; Abolfathi, S.; Pearson, J.; Chatrabgoun, O.; Daneshkhan, A. Gaussian process emulation of spatio-temporal outputs of a 2D inland flood model. *Water Res.* **2022**, *225*, 119100. [CrossRef] [PubMed]
14. Li, G.; Shao, W.; Su, X.; Li, Y.; Zhang, Y.; Song, T. Urban Flood Hazard Assessment Based on Machine Learning Model. *Water Resour. Manag.* **2025**, *39*, 1953–1970. [CrossRef]
15. Sushanth, K.; Mishra, A.; Mukhopadhyay, P.; Singh, R. Real-time streamflow forecasting in a reservoir-regulated river basin using explainable machine learning and conceptual reservoir module. *Sci. Total Environ.* **2023**, *861*, 160680. [CrossRef] [PubMed]
16. Lee, S.; Kim, J.C.; Jung, H.S.; Lee, M.J.; Lee, S. Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomat. Nat. Hazards Risk* **2017**, *8*, 1185–1203. [CrossRef]
17. Tehrany, M.S.; Pradhan, B.; Jebur, M.N. Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *J. Hydrol.* **2013**, *504*, 69–79. [CrossRef]
18. Yin, G.; Yoshikane, T.; Yamamoto, K.; Kubota, T.; Yoshimura, K. A support vector machine-based method for improving real-time hourly precipitation forecast in Japan. *J. Hydrol.* **2022**, *612*, 128125. [CrossRef]
19. Razavi-Termeh, S.V.; Sadeghi-Niaraki, A.; Jelokhani-Niaraki, M.; Choi, S.-M. Flood susceptibility mapping using optimized deep learning models: A non-structural framework. *Appl. Water Sci.* **2025**, *15*, 208. [CrossRef]
20. Zhao, G.; Pang, B.; Xu, Z.; Yue, J.; Tu, T. Mapping flood susceptibility in mountainous areas on a national scale in China. *Sci. Total Environ.* **2018**, *615*, 1133–1142. [CrossRef] [PubMed]
21. Ahmad, M.N.; Shao, Z.; Javed, A. Mapping impervious surface area increase and urban pluvial flooding using Sentinel Application Platform (SNAP) and remote sensing data. *Environ. Sci. Pollut. Res. Int.* **2023**, *30*, 125741–125758. [CrossRef] [PubMed]
22. Wedajo, G.K.; Lemma, T.D.; Fufa, T.; Gamba, P. Integrating Satellite Images and Machine Learning for Flood Prediction and Susceptibility Mapping for the Case of Amibara, Awash Basin, Ethiopia. *Remote Sens.* **2024**, *16*, 2163. [CrossRef]
23. Li, Z.; Tian, J.; Zhu, Y.; Chen, D.; Ji, Q.; Sun, D. A Study on Flood Susceptibility Mapping in the Poyang Lake Basin Based on Machine Learning Model Comparison and SHapley Additive exPlanations Interpretation. *Water* **2025**, *17*, 2955. [CrossRef]

24. Tien Bui, D.; Hoang, N.D.; Martínez-Álvarez, F.; Ngo, P.-T.T.; Hoa, P.V.; Pham, T.D.; Samui, P.; Costache, R. A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Sci. Total Environ.* **2020**, *701*, 134413. [CrossRef] [PubMed]
25. Shahabi, H.; Shirzadi, A.; Ghaderi, K.; Omidvar, E.; Al-Ansari, N.; Clague, J.J.; Geertsema, M.; Khosravi, K.; Amini, A.; Bahrami, S.; et al. Flood detection and susceptibility mapping using Sentinel-1 remote sensing data and a machine learning approach: Hybrid intelligence of bagging ensemble based on K-Nearest Neighbor classifier. *Remote Sens.* **2020**, *12*, 266. [CrossRef]
26. Costache, R.; Pham, Q.B.; Sharifi, E.; Linh, N.T.T.; Abba, S.; Vojtek, M.; Vojteková, J.; Nhi, P.T.T.; Khoi, D.N. Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and GIS techniques. *Remote Sens.* **2020**, *12*, 106. [CrossRef]
27. Band, S.S.; Janizadeh, S.; Pal, S.C.; Saha, A.; Chakraborty, R.; Melesse, A.M.; Mosavi, A. Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. *Remote Sens.* **2020**, *12*, 3568. [CrossRef]
28. La Cava, W.; Bauer, C.; Moore, J.H.; Pendergrass, S.A. Interpretation of machine learning predictions for patient outcomes in electronic health records. *arXiv* **2019**, arXiv:1903.12074. [CrossRef]
29. Pradhan, B. Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. *J. Spat. Hydrol.* **2009**, *9*, 1–18.
30. Günther, F.; Fritsch, S. neuralnet: Training of Neural Networks. *R J.* **2010**, *2*, 30–38. Available online: <https://journal.r-project.org/articles/RJ-2010-006/index.html> (accessed on 3 July 2026). [CrossRef]
31. Zhang, Z.; Laakso, T.; Wang, Z.; Pulkkinen, S.; Ahopelto, S.; Virrantaus, K.; Li, Y.; Cai, X.; Zhang, C.; Vahala, R.; et al. Comparative study of AI-based methods-Application of analyzing inflow and infiltration in sanitary sewer subcatchments. *Sustainability* **2020**, *12*, 6254. [CrossRef]
32. Pakdehi, M.; Ahmadisharaf, E.; Nazari, B.; Cho, E. Transferability of machine-learning-based modeling frameworks across flood events for hindcasting maximum river water depths in coastal watersheds. *Nat. Hazards Earth Syst. Sci.* **2024**, *24*, 3537–3559. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## Article

# A Large-Scale Evaluation of SWOT-Derived Water Surface Elevations: Precision Drivers and Strategies to Enhance Data Availability

Thiago Lappicy <sup>1,2,\*</sup>, Daniel Beltrão <sup>1,2</sup>, Luana Oliveira Sales <sup>1</sup>, Tati Almeida <sup>1</sup>, Guilherme Gomes Pessoa <sup>1</sup>, Saulo Souza <sup>3</sup>, Renato Prata de Moraes Frasson <sup>2</sup> and Rejane Ennes Cicerelli <sup>1</sup>

<sup>1</sup> Postgraduate Program in Applied Geosciences and Geodynamics (PPGAG), University of Brasilia (UnB), Brasilia 70910-900, DF, Brazil; beltrao.daniel@aluno.unb.br (D.B.); luana.sales@aluno.unb.br (L.O.S.); tati\_almeida@unb.br (T.A.); guilherme.pessoa@unb.br (G.G.P.); rejaneig@unb.br (R.E.C.)

<sup>2</sup> Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, CA 91109, USA; renato.prata.de.moraes.frasson@jpl.nasa.gov

<sup>3</sup> National Water and Sanitation Agency (ANA), Brasilia 70610-200, DF, Brazil; saulo.souza@ana.gov.br

\* Correspondence: thiago.lappicy@aluno.unb.br

## Highlights

### What are the main findings?

- SWOT-derived WSE shows good agreement with in situ observations across 132 Brazilian lakes: Flag = 0 reduces the 68th percentile errors to below 12 cm but retains only 22% of observations, while Flag = 1 is affected by outliers (68th percentile errors below 21 cm).
- A Random Forest analysis identifies cross-track distance and lake geometry as the dominant drivers of WSE precision across three SWOT products (vector and raster products).

### What are the implications of the main findings?

- The SQRTL filter combines all Flag = 0 with cross-track constrained Flag = 1 observations, more than tripling usable data relative to Flag = 0 and achieving comparable precision (68th percentile of errors below 16 cm).
- This framework is transferable to other regions, showing potential to expand systematic lake monitoring for (un)gauged lakes.

## Abstract

High-quality water surface elevation (WSE) measurements are critical in hydrological applications, yet no systematic evaluation of the Surface Water and Ocean Topography (SWOT) mission exists for Brazil's diverse lake systems, where satellite observations are essential given limited in situ monitoring. We evaluated WSE from SWOT over 132 Brazilian lakes, comparing LakeSP, Raster\_250m, and Raster\_100m products against field measurements over a 20-month period. The 68th percentile errors were under 29 cm for the full dataset, below 12 cm for Flag = 0, and below 21 cm for Flag = 1, indicating good agreement but also the presence of outliers and the need for data screening. A Random Forest analysis identified quality flags, lake geometry, and cross-track distance as key drivers of WSE precision. Flag = 0 is overly restrictive, retaining only 22% of observations, while Flag = 1 contains anomalous data. The SWOT Quality-Range Threshold for Lakes (SQRTL) filter combines Flag = 0 with cross-track constrained Flag = 1 observations. SQRTL more than triples data availability relative to Flag = 0, maintaining comparable precision (68th percentile below 16 cm) and reducing median revisit from 88–123 days to 16–18 days for raster products

and from 25 to 14 days for LakeSP. These results provide the first large-scale SWOT WSE evaluation over Brazilian lakes and a transferable filtering framework applicable wherever SWOT and field observations overlap, with potential to extend monitoring to over 100,000 water bodies in the SWOT Prior Lake Database.

**Keywords:** SWOT; water surface elevation (WSE); inland waters; lake monitoring; quality filtering; cross-track distance; Brazil; satellite altimetry; random forest

## 1. Introduction

Lakes and reservoirs are essential components of water resource systems, supporting water supply, irrigation, livestock watering, hydropower generation, and recreation. Globally, estimates of the number of lakes range from tens to hundreds of millions [1–3], making comprehensive in situ monitoring impractical and motivating the use of satellite observations to assess inland water dynamics at broad spatial scales [4].

In Brazil, the National Water and Sanitation Agency (ANA) has mapped 236,302 lakes, of which 147,830 are larger than 0.01 km<sup>2</sup> and 58,622 exceed the Surface Water and Ocean Topography (SWOT) detection limit of 0.0625 km<sup>2</sup> [5]. Satellite-based inventories can provide a complementary perspective on these systems. The Prior Lake Database (PLD) [6] identifies over 6 million lakes globally, including 109,332 in Brazil fully covered by SWOT overpasses, representing nearly half of all lakes mapped nationally.

In addition, Brazil has 3361 officially mapped artificial reservoirs, with an estimated total storage capacity of 630.2 billion m<sup>3</sup>, of which 92.7% is associated with hydropower generation [5]. However, only 714 reservoirs (21%) are systematically monitored at the federal level with publicly available data [5,7], limiting the assessment of spatial and temporal variability in water levels and storage. This imbalance between the extent of inland waters and observational coverage highlights the importance of satellite observations for characterizing lake and reservoir dynamics at large scales [8–11].

Recent advances in satellite altimetry have improved the remote estimation of water surface elevation (WSE), with missions such as Sentinel-3, Sentinel-6/Jason-CS, ICESat-2, SARAL/AltiKa, and, most recently, SWOT [12–17]. Using lidar and radar remote sensing technologies, these missions have expanded the availability and quality of hydrological observations. SWOT differs from previous altimetry missions by enabling wide-swath radar observations of lakes and rivers, providing spatially distributed estimates of WSE, river discharge, and lake storage dynamics [12].

These capabilities create new opportunities for lake and reservoir monitoring [18–22], hydrological modeling/data assimilation [23,24], river and discharge-related analyses [25–28], flood and river-network studies [17,29,30], and broader large-scale hydrological assessments [31–34].

Studies using SWOT data report the promising accuracy of SWOT-derived WSE over inland waters, with errors below 20 cm after applying quality flags [35,36]. Flag-based filtering has been applied across different SWOT products, including pixel cloud, vector, and raster datasets [21,35–42], with many studies combining it with ancillary variables from the SWOT products, such as the backscatter coefficient *sig0* [37], WSE uncertainty estimates [22,38], dark water fraction [20,39,40], and cross-track distance [39], as well as with statistical outlier removal methods [21,22,38,40–42]. However, most of these studies focus on specific regions or individual products, limiting direct comparison across SWOT products and leaving open questions regarding the potential to generalize the filtering strategies.



Thus, relying solely on quality flags, either by retaining noisy observations or discarding a substantial fraction of the available observations, has often proven insufficient, leaving the trade-off between improved accuracy and data availability unresolved. Recent studies have explored alternatives to strict quality-flag screening [28], but no broadly adopted strategy has yet emerged for balancing reliability and data retention. This remains a key challenge in the practical use of SWOT data, particularly across large and diverse regions.

To address these limitations, we propose the SQRTL (SWOT Quality-Range Threshold for Lakes) filter, which combines SWOT quality flags, already widely used in the literature [21,35–42], with cross-track distance. By relying on variables provided in SWOT products themselves, the SQRTL approach can be readily integrated into existing frameworks, remains physically interpretable, and is intended as a replicable strategy that may be applied across regions and scales, rather than a filter tailored to a specific study or dataset.

Our study addresses these gaps by providing a nationwide evaluation of SWOT-derived WSE precision using in situ observations from 132 lakes across Brazil. We explore the factors controlling WSE precision and variability in the LakeSP, Raster\_250m, and Raster\_100m products, with particular emphasis on geometric and orbital influences, using a Random Forest (RF) approach. Based on these results, we propose an auxiliary filtering strategy that complements the current flagging system. Together, these analyses provide a large-scale comparison of three SWOT products and a practical framework for improving data screening.

## 2. Materials and Methods

### 2.1. Materials

The data sources used were: (i) in situ water level measurements provided by the Brazilian National Operator for the Electric System—ONS [43]; (ii) SWOT WSE observations from three distinct products: LakeSP, Raster\_250m, and Raster\_100m; and (iii) lake shape and size information for each hydroelectric power plant (HEPP) calculated using the ANA database [44]. The studied sites are illustrated in Figure 1. They are well distributed across Brazil's 12 hydrographic regions [45], with latitudes ranging from 29.45°S to 0.92°N and longitudes between 64.65°W and 37.80°W, covering approximately 3360 km north–south and 2950 km east–west.

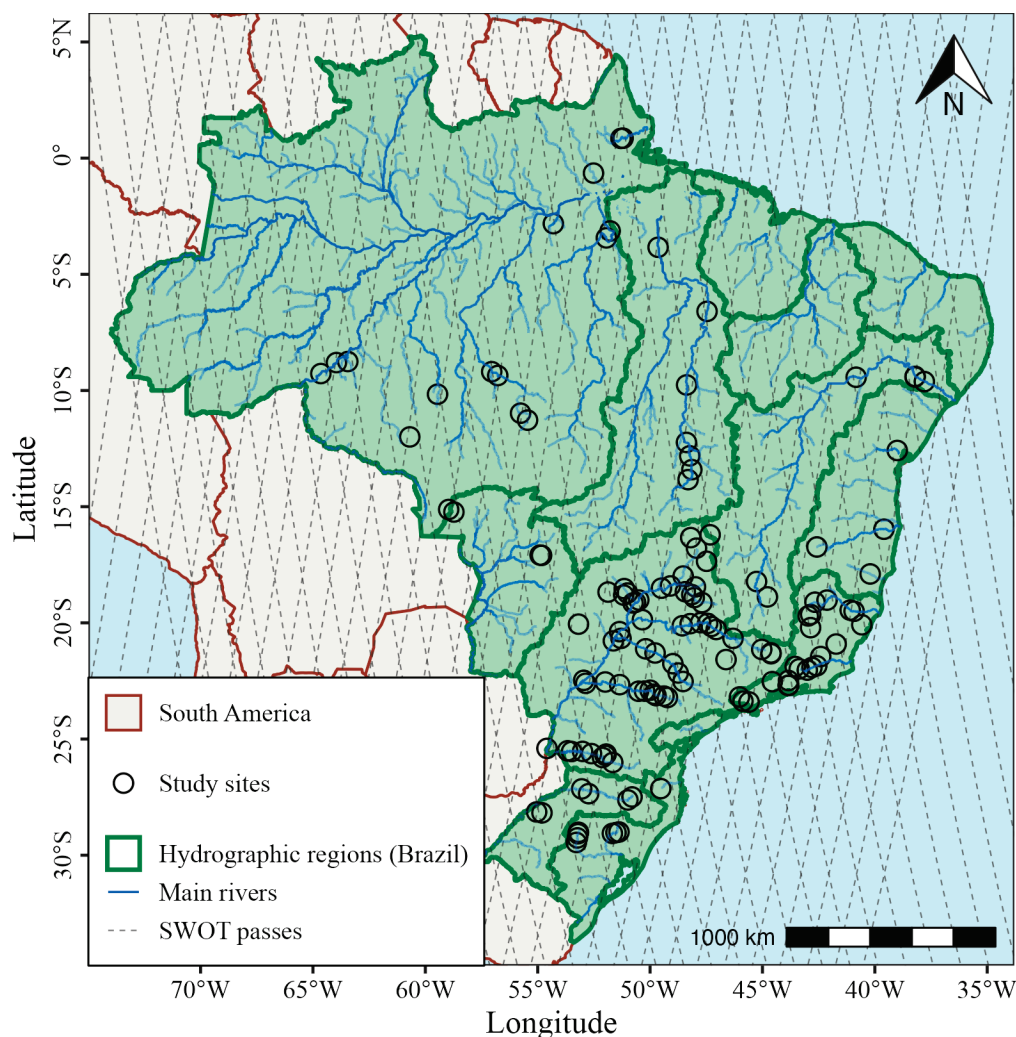
#### 2.1.1. ONS In Situ Observations

To evaluate SWOT-derived WSE, hourly water level observations from ONS were used as the in situ reference dataset. ONS operates approximately 180 hydroelectric power plants across Brazil, of which 132 intersected SWOT observations at least once during the study period (26 July 2023 to 31 March 2025) [43].

To address possible mismatches in vertical datum between SWOT and ONS observations, we paired the SWOT and the ONS WSE measurements, calculated the differences in the WSEs, and removed the median SWOT-ONS difference from each paired series, resulting in WSE anomalies. The adjusted series were used to quantify the precision of SWOT WSE measurements.

#### 2.1.2. SWOT Data for Lakes

The SWOT mission carries a Ka-band Radar Interferometer (KaRIn), designed to measure water surface extent and elevation for lakes with surface areas larger than  $250 \times 250$  m [31]. SWOT collects data in swaths, repeating its orbit every 20.86 days, since 21 July 2023 [46], with most water bodies being observed between 1 and 3 times per cycle [31]. Three products are used in this study, with data from version C: the Lake Single Pass (LakeSP), and the rasters with 250 m (Raster\_250m) and 100 m (Raster\_100m) resolution.



**Figure 1.** Location of 132 study sites across Brazilian hydrographic regions, with overlapping SWOT passes.

SWOT data referenced to EGM2008 [47,48] were retrieved from 26 July 2023 to 31 March 2025 (“science” or “nominal” orbit), resulting in 4443 observations for LakeSP, 7766 for Raster\_250m, and 7678 for Raster\_100m, all matched with a corresponding value from ONS. Matching was performed both temporally, by pairing each SWOT observation to the closest hourly ONS record, and spatially, using the intersecting lake feature for LakeSP\_Obs and a  $3 \times 3$  pixel window for the raster products. For the HEPPs analyzed, this sampling corresponds to roughly three observations per month, with the median revisit intervals of ~10 days for the raster products and 13 days for LakeSP.

The LakeSP product is a vector dataset derived from the PLD [6]. The observation-oriented product (LakeSP\_Obs) was used in this study, as it has one element per observed lake feature, allowing flexibility for merging multiple PLD lakes or breaking up a PLD lake into smaller ones [46,48].

Throughout this study, the term “lake” refers specifically to water-surface features represented within the SWOT PLD, reflecting the SWOT data model rather than a geomorphological classification.

The Raster\_250m and Raster\_100m products only differ in spatial resolution (250 m and 100 m), distributed in  $64 \text{ km} \times 64 \text{ km}$  tiles [46,47]. These products were selected due to

processing considerations for a multitemporal and national-scale analysis, which would make the use of the pixel cloud more complex and computationally demanding.

Each SWOT product provides WSE values with ancillary data that can be helpful in selecting better quality data or provide insight into data anomalies, explored with an RF algorithm in Sections 2.2.2 and 2.2.3. Notable examples are the summary quality flags for each observation [46], which can be: 0 (Good), 1 (Suspect), 2 (Degraded), or 3 (Bad), determined by multiple factors [46–48].

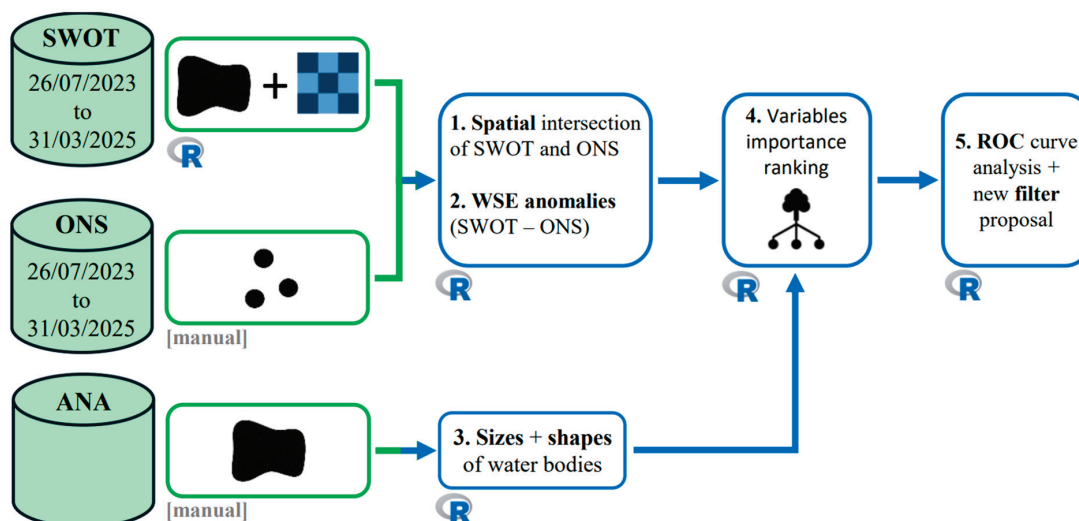
### 2.1.3. Brazilian Database for Water Bodies

To calculate geometry parameters of each HEPP, we used ANA's water body database, "Massas d'Água" [44]. Previous studies have indicated that lake size correlates with surface area variability, suggesting smaller lakes experience greater variability and more extreme values [8]. This relation could extend to WSE precision, as evaluated in this study.

## 2.2. Methods

This study compares SWOT-derived WSE with in situ ONS observations to quantify their anomalies. These differences should not be interpreted solely as measurement errors, as ONS data also contain uncertainty. Nevertheless, they provide a useful basis for assessing disagreement and its main drivers. An RF model was used to rank the explanatory variables associated with WSE differences, including both SWOT-derived metrics from the satellite products and non-SWOT information such as lake geometry and shape [49–51].

Based on these results, a Receiver Operating Characteristic (ROC) analysis was used to evaluate auxiliary filtering strategies using the key variables from the RF, resulting in the proposed SQRTL filter. Five differently sized lakes (four monitored and one unmonitored) were used as examples to illustrate the potential of this filter for gauged and ungauged systems. Figure 2 summarizes the workflow, from data collection to filter development.



**Figure 2.** Methodological workflow for WSE data processing and analysis.

### 2.2.1. Selection of ONS and SWOT Locations for WSE Anomalies Assessment

The geographic coordinates of the HEPPs typically correspond to dam structures, which are usually located near land. This can introduce biases when using SWOT data, as significant elevation gradients, transitions between the upstream reservoir and downstream river reach, and edge effects often occur near dams, while land contamination can further

degrade SWOT-derived WSE estimates. To reduce these effects, the geographic position was shifted approximately 500 m upstream.

This offset was defined based on the lowest spatial resolution pixel (250 m), for which the maximum distance from the center of a  $3 \times 3$  pixel window to its farthest point is less than 500 m. This displacement reduces the inclusion of pixels located downstream from dams and near-shore observations, limits land contamination effects, and helps ensure that the extracted pixels are representative of the lake environment. These shifted locations were also manually verified for all 132 HEPPs to confirm that the selected points remained within the water body and avoided any obvious contamination from the dam structure, nearby land, or downstream reach.

No additional filtering to remove land contamination pixels was applied, as excessive removal may introduce additional noise [26]. Next, we filtered SWOT data to retain only HEPP-intersecting observations. For LakeSP\_Obs, the intersection determined the WSE value used, whereas for raster products, the median of a  $3 \times 3$  pixel window centered on the 500 m upstream point was used, similar to [36], but we used the median instead of the mean to reduce sensitivity to extreme values.

## 2.2.2. Evaluating Potential Drivers of WSE Precision

To better understand SWOT's WSE precision, we selected known sources of measurement error [18,26,50], as shown in Table 1. We used the static lake geometries from ANA [44] as shapefiles to calculate six parameters that characterize the lake geometry: (i) *area*,  $A$  ( $\text{km}^2$ ); (ii) *perimeter*,  $P$  (km); (iii) *circularity index*,  $C_i$ , calculated as  $3.545\sqrt{A/P}$  [51]; (iv) *compactness coefficient*,  $C_c$ , calculated as  $0.2841P/\sqrt{A}$  [51]; (v) *minimum shore-to-shore distance* and (vi) *maximum shore-to-shore distance* (km) estimated from the lake area and perimeter assuming a characteristic rectangular lake shape [49].

**Table 1.** Explanatory variables used in the random forest algorithm.

| Variable                    | Description of Variable                             | Data Source   |
|-----------------------------|-----------------------------------------------------|---------------|
| <i>cross_track</i>          | Distance to the satellite's cross-track [m]         | LakeSP/Raster |
| <i>wse_u</i>                | Uncertainty in the WSE measurement [m]              | LakeSP/Raster |
| <i>wse_qual</i>             | Quality Flag assigned by SWOT                       | LakeSP/Raster |
| SWOT Product                | LakeSP, Raster_100m or Raster_250m                  | LakeSP/Raster |
| <i>layovr_val</i>           | Estimate of the WSE error due to layover            | LakeSP        |
| <i>wse_std</i>              | Standard deviation of WSE from lake's pixels        | LakeSP        |
| <i>dark_frac</i>            | Fraction of the lake covered by dark water          | LakeSP        |
| <i>partial_f</i>            | Flag that indicates partial lake coverage           | LakeSP        |
| <i>xovr_cal_c</i>           | Cross-over calibration applied to wse [m]           | LakeSP        |
| <i>xovr_cal_q</i>           | Quality flag for the cross-over calibration         | LakeSP        |
| <i>inc</i>                  | Incidence angle of radar measurement [deg.]         | Raster        |
| <i>darkwater</i>            | Fraction of <i>water_area</i> covered by dark water | Raster        |
| <i>water_frac</i>           | Estimated fraction of pixel covered by water        | Raster        |
| <i>sig0</i>                 | Normalized radar cross section, or $\sigma^0$ [dB]  | Raster        |
| <i>sig0_cor_atmos_model</i> | Atmos. correction model applied to $\sigma^0$ [dB]  | Raster        |
| <i>sig0_qual</i>            | Quality flag for the $\sigma^0$ at water pixels     | Raster        |
| <i>xover</i>                | Height correction to WSE [m]                        | Raster        |
| Number of pixels            | Number of pixels with data in $3 \times 3$ window   | Raster        |
| Area                        | Surface area of the lake [ $\text{km}^2$ ]          | ANA           |
| Perimeter                   | Perimeter of the lake outline [km]                  | ANA           |
| Min. shore-to-shore dist.   | Parameter for shape and size of lake [km]           | ANA           |
| Max. shore-to-shore dist.   | Parameter for shape and size of lake [km]           | ANA           |
| Circularity index           | Similarity to a circular shape                      | ANA           |
| Compactness coefficient     | Deviation from circularity                          | ANA           |

SWOT-specific metrics such as cross-track distance, quality flags, and additional ancillary variables (Table 1) were also considered. All variables in italic are named exactly as given in the SWOT product and can be found in the user handbook or in the product manuals [46–48].

### 2.2.3. Identifying Key Drivers of WSE Precision via Random Forests

An RF classifier implemented in R with the randomForest package [52,53] was used to investigate the relationship between the absolute WSE anomaly (target variable) and the explanatory variables listed in Table 1. We chose the RF classifier because it is robust to correlation among explanatory variables and is well suited for capturing complex, non-linear relationships between explanatory and target variables [52,54,55]. Moreover, RF is relatively robust to noise and outliers and does not rely on linearity or distributional assumptions, while providing established measures of variable relative importance that enable the identification of dominant drivers of WSE precision [52,54].

The absolute anomaly was treated as a categorical variable representing six intervals:  $\leq 0.1$  m, 0.1–0.3 m, 0.3–0.5 m, 0.5–0.8 m, 0.8–1.0 m,  $> 1.0$  m. Subsequently, we evaluated the Mean Decrease Accuracy (MDA) metric to quantify the relative importance of the explanatory variables in the composition of the WSE measurement error, with higher values indicating a higher sensitivity and thus a greater influence [53,56,57]. The evaluated lake-geometry variables represent similar measurements of lake size and shape; hence, their MDA values are interpreted comparatively and collectively.

The data were split 70/30 for training and testing, while maintaining representativeness, with the parameters being tuned via a grid search (number of trees = 1500, node size = 4, max. number of nodes = 500, and number of sampled variables per split = 4). These values also aid in the use of correlated variables in RF.

### 2.2.4. Auxiliary Filtering Strategy Based on ROC Analysis

As filtering strategies inherently involve trade-offs between data availability and reliability, we adopted a parsimonious approach. Based on the error distributions shown in Section 3.3, Flag = 0 measurements consistently exhibit low errors and do not require additional filtering, while Flag = 3 observations are dominated by large errors and were excluded. Flag = 2 data contain only a small fraction of high-quality measurements and a high proportion of large errors; hence, they were also excluded.

Consequently, we propose an auxiliary filter to allow the inclusion of observations flagged as suspect (Flag = 1) in addition to all observations flagged as good (Flag = 0). Rather than including all observations flagged as 1, we aim to only retain a subset of points with the highest likelihood of having good agreement with in situ WSE observations. Therefore, we only include suspect points if they lie within an optimal part of the swath. We made use of the ROC [58] curve to identify the optimal threshold for the filtering.

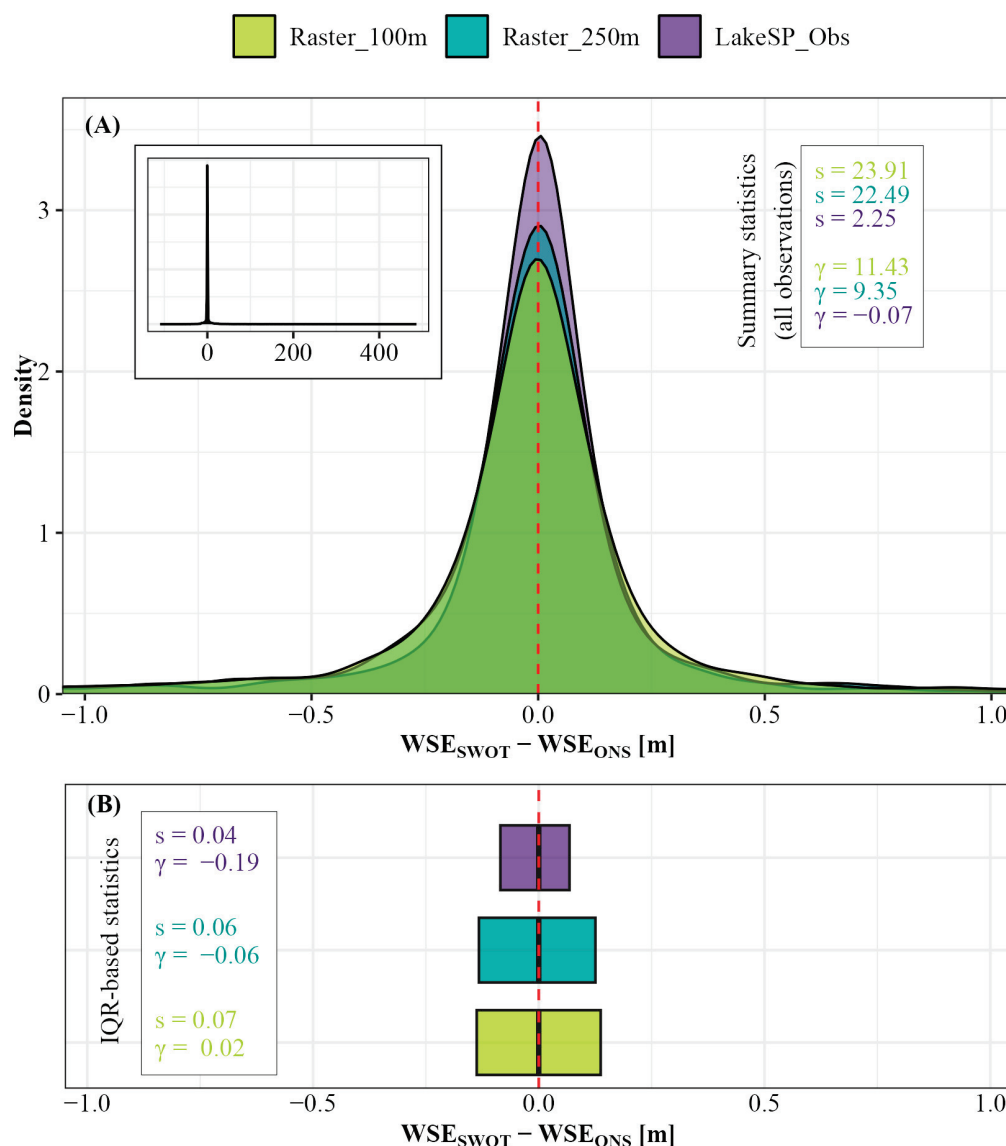
The ROC curve plots the hit rate (true positive rate) against the false alarm rate (false positive rate). We used the same threshold values used in the RF classifier (0.1 m, 0.3 m, 0.5 m, 0.8 m, 1.0 m) for the construction of the curves. These threshold values are defined in terms of the absolute difference between the SWOT and the in situ data with a bias removal (WSE anomaly). With a grid search calibration, the optimal cross-track lower and upper boundaries are identified. This optimization increases the number of retained observations, while also minimizing the addition of anomalous observations. We then illustrate the application of this filter for an unmonitored lake.



### 3. Results

#### 3.1. Water Surface Elevation (WSE) from SWOT

Initial results include density plots (Figure 3A) and interquartile range (IQR) boxplots (Figure 3B) of WSE differences between SWOT and ONS. The smaller panel shows the full distribution, highlighting the presence of a small number of anomalous values, likely representing erroneous measurements. The existence of anomalous data has been documented and merits further attention [20–22,34–37,39–42], motivating the use of a zoomed-in plot and IQR-based analyses aimed at characterizing the central tendencies and dispersion. The main panel focuses on differences ranging from  $-1.0$  m to  $+1.0$  m, encompassing most valid observations, while the boxplots only show values within the IQR (middle 50% of the dataset) for each product. The red dotted line denotes the optimal value, where SWOT and ONS measurements coincide.



**Figure 3.** WSE differences (SWOT-ONS) after bias correction. (A) Histogram, with inset plot showing full range; and (B) Boxplots. Symbols  $s$  and  $\gamma$  represent the standard deviation and skewness.

The two statistical metrics calculated for each product were the standard deviation ( $s$ ) and skewness coefficient ( $\gamma$ ), considering both the full dataset and only values within the

IQR. The standard deviation is calculated by the square root of the second central moment and the skewness by the Fisher–Pearson standardized moment coefficient. For the raster products, the WSE value used is the median over a  $3 \times 3$  pixel window.

When restricting the analysis to the IQR, thereby focusing on the most representative portion of the distributions and not on the outliers (Figure 3B), the standard deviation is significantly smaller and closer between products, with 4 cm for LakeSP\_Obs and 6 to 7 cm for both raster products. These results are in agreement with the current literature regarding the presence of spurious WSE data from SWOT.

The raster products exhibit positive skewness, indicating that the data are more prevalently overestimating the WSE from SWOT relative to ONS (Figure 3), also observed in Australian water bodies [37]. LakeSP\_Obs shows greater stability, possibly because its WSE values represent a spatially aggregated water body, reducing localized variability and noise. When constraining to the IQR, all products have a near-zero skewness, reinforcing greater stability in central values and highlighting the importance of targeted screening to improve hydrological interpretation.

The spatial distribution of the WSE anomalies observed reveals little geographic structure (Supplementary Material Figure S1). Higher differences tend to occur in coastal and southern regions, whereas lower median errors are more frequent in central Brazil, but are also observed elsewhere.

### 3.2. Explanatory Variables Associated with WSE Precision from SWOT

The RF analysis (Figure 4) ranks the explanatory variables (Table 1) according to their contribution to classify WSE error magnitude. The “All” product in the legend key corresponds to the pooled dataset including LakeSP\_Obs, Raster\_100m, and Raster\_250m. Here, we ranked the importance of the variables according to the MDA metric [52,53,56,57]. On the test datasets, the models achieved success rates of approximately 60%, outperforming a random classifier expected to correctly identify 16% of cases (1/6).

Higher MDA values imply a greater influence of a variable on model performance, thus being able to categorize (within the thresholds used) the magnitude of the difference between SWOT and ONS measurements. Even though the importance values vary depending on the model, i.e., the SWOT product used, due to differing explanatory variables available, some consistent patterns emerge.

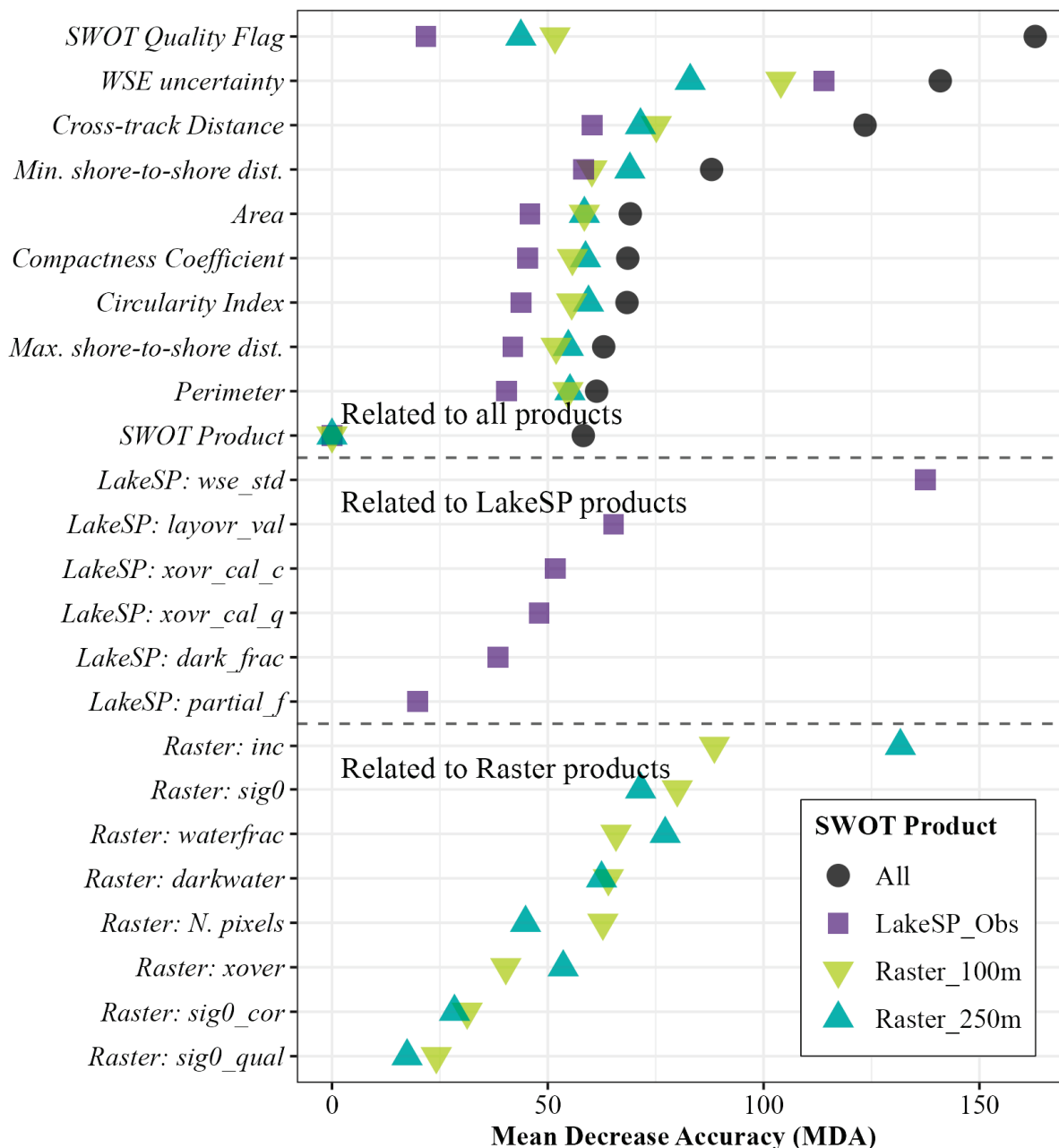
The higher MDA of the SWOT quality flag when using “All” three products may occur due to the lack of many ancillary variables that are product specific (only available in product-specific models), thus acting as a strong discriminator of the errors, when little additional information is given.

As anticipated, variables representing measurement uncertainty (WSE uncertainty and *wse\_std*) have considerable importance, confirming that SWOT’s uncertainty metrics effectively aid in characterizing the magnitude of measurement errors. Furthermore, SWOT quality flags also showed considerable importance, along with the cross-track distance, *wse\_std* and *layovr\_val* (for LakeSP\_Obs) and *inc* (for both raster products).

Notably, minimum shore-to-shore distance is the most influential non-SWOT variable across all products, with other shape and geometry parameters (area, perimeter, circularity index, compactness coefficient, and maximum shore-to-shore distance) clustered together in their importance. Their relative order should not be overinterpreted, as correlated predictors may redistribute importance within the RF model. This overall importance indicates that lake geometry (size and shape) is relevant to WSE precision.

It is worth noting that a lower MDA does not indicate a lack of importance, rather it suggests that within this RF model, these variables have a lesser impact than other,

more influential, variables. Therefore, filtering strategies based on the most influential variables may yield better results. Subsequent sections further explore the influence of the top-ranked variables on WSE precision, specifically the flags, geometry parameters, cross-track distance, layover error and incidence angle.

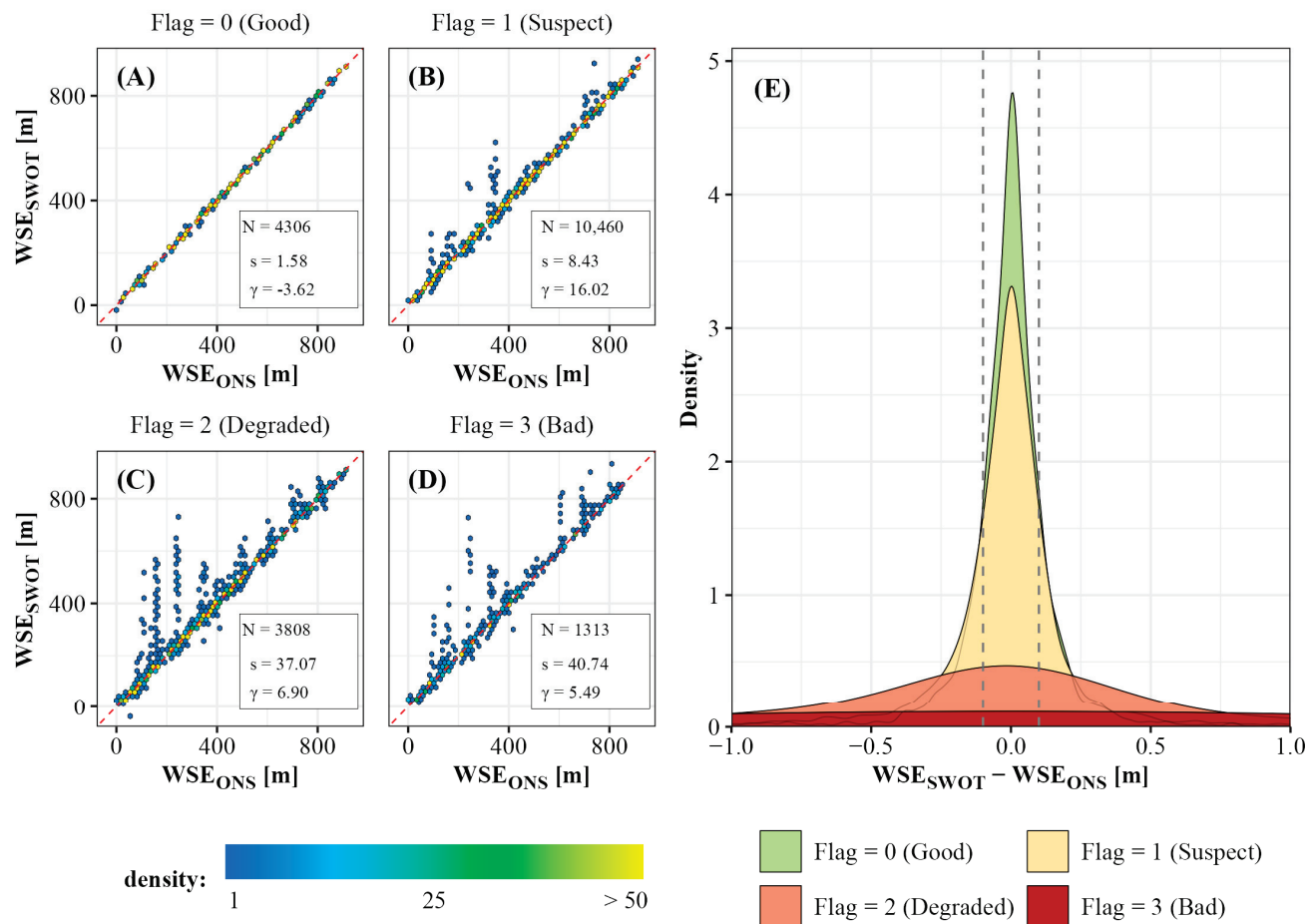


**Figure 4.** Mean Decrease Accuracy (MDA) in the random forest (higher values have higher influence), with each variable meaning described in Table 1. “All” refers to the RF model fitted to the dataset containing all three SWOT products.

### 3.3. Performance and Limitations of SWOT Quality Flags

With several studies using SWOT quality flags as a filter, we analyzed them, ranging from 0 (“Good”) to 3 (“Bad”). Figure 5A–D display heatmaps of the 19,887 WSE observations separated by their assigned quality flags. Flags 2 (“Degraded”) and 3 (“Bad”) clearly identify a substantial fraction of the poor-quality data, indicated by their high error standard deviations (s) of 37.07 m and 40.74 m, respectively. The lower performance of

measurements flagged as degraded and bad is also visible in Figure 5E, which shows the histogram of the measurement error categorized per quality flag value. While Flags 0 and 1 are mostly concentrated near 0, measurements classified as degraded and bad have a less clear peak and considerably higher density around high error values.



**Figure 5.** WSE comparisons (SWOT-ONS). (A–D) Heatmaps of the WSE observed for each Flag; and (E) Distribution of the errors in each flag, with the  $\pm 10$  cm error mark as gray dashed lines.

However, while Flag = 0 significantly filters poor-quality data, resulting in a closer alignment with in situ measurements, it still has a high standard deviation of 1.58 m. From Figure 5E, we can see that this value is probably due to the presence of outliers within this flag, as its histogram clearly shows a concentration of data inside these boundaries and near zero. Moreover, this flag only contains 22% (4306 out of 19,887 points) of available data; hence, most of the information gathered from SWOT does not fall within this criterion, suggesting that filtering only Flag = 0 might be overly restrictive (Table 2).

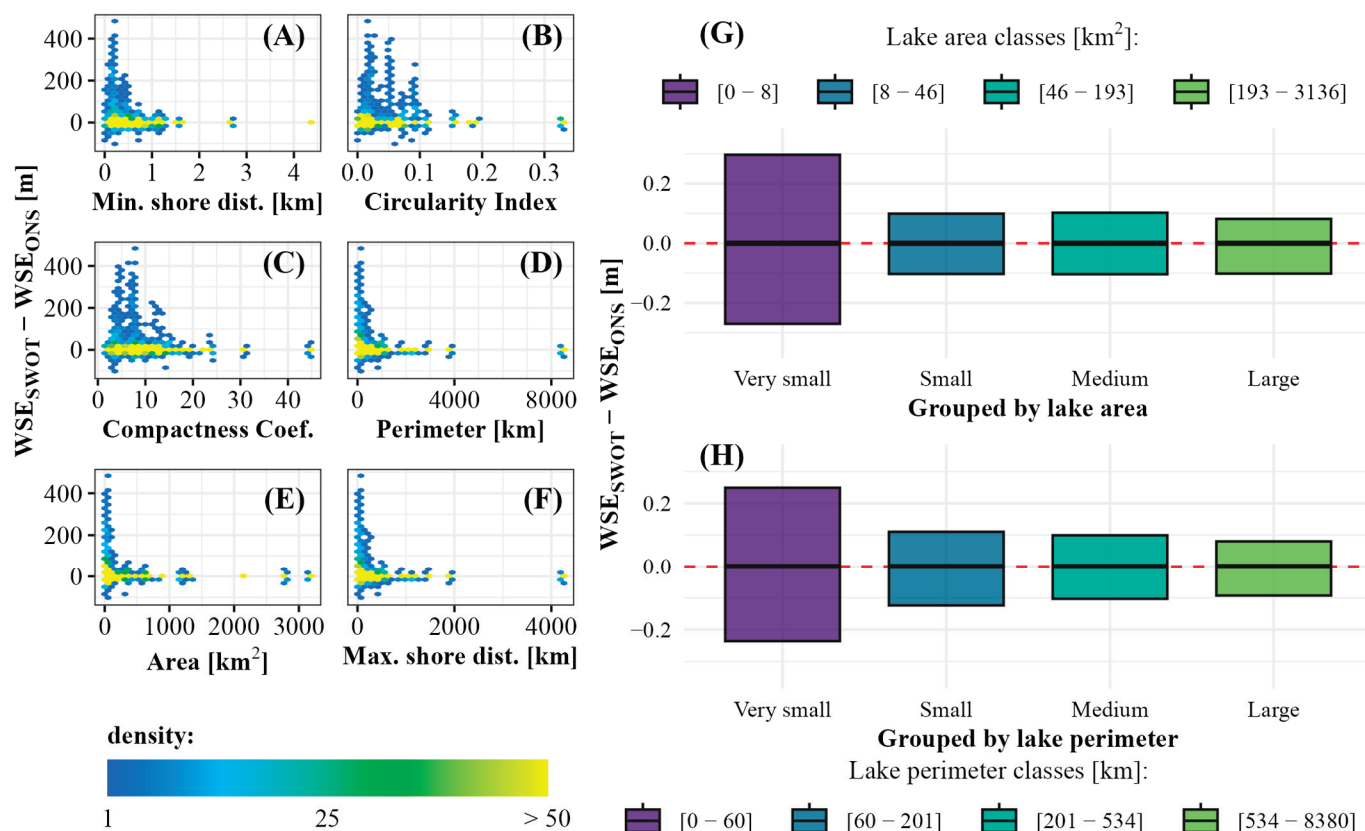
In addition, Flag = 1 contains most of the data points, including 53% of the total, or 10,460 observations (Table 2). Despite exhibiting higher variability than those points flagged as good, i.e., a standard deviation of 8.43 m, it still retains a significant number of reliable observations near optimal values. Table 2 presents the number and percentage of observations within absolute WSE difference thresholds (0.1 m, 0.3 m, 0.5 m, 0.8 m, 1.0 m) across flagging criteria, allowing performance evaluation under different tolerance levels. The SQRTL column reflects an alternative filtering strategy proposed in Section 3.6. This strategy retains 66% of the dataset, with 13,070 measurements (more than three times that of Flag = 0), while maintaining a similar proportion of accurate observations.

**Table 2.** Number of SWOT observations within different thresholds per filter (e.g., Flags).

| Threshold | Total    |    | Flag = 0 |    | Flag ≤ 1 |    | Flag ≤ 2 |    | Flag ≤ 3 |    | SQRTL    |    |
|-----------|----------|----|----------|----|----------|----|----------|----|----------|----|----------|----|
|           | <i>n</i> | %  | <i>n</i> | %  | <i>n</i> | %  | <i>n</i> | %  | <i>n</i> | %  | <i>n</i> | %  |
| ≤  0.1 m  | 9212     | 46 | 2796     | 65 | 8219     | 56 | 9011     | 49 | 9212     | 46 | 7564     | 58 |
| >  0.1 m  | 10,675   | 54 | 1510     | 35 | 6547     | 44 | 9563     | 51 | 10,675   | 54 | 5506     | 42 |
| ≤  0.3 m  | 13,681   | 69 | 3837     | 89 | 11,955   | 81 | 13,311   | 72 | 13,681   | 69 | 10,821   | 83 |
| >  0.3 m  | 6206     | 31 | 469      | 11 | 2811     | 19 | 5263     | 28 | 6206     | 31 | 2249     | 17 |
| ≤  0.5 m  | 14,738   | 74 | 3983     | 92 | 12,672   | 86 | 14,301   | 77 | 14,738   | 74 | 11,437   | 88 |
| >  0.5 m  | 5149     | 26 | 323      | 8  | 2094     | 14 | 4273     | 23 | 5149     | 26 | 1633     | 12 |
| ≤  0.8 m  | 15,429   | 78 | 4041     | 94 | 13,056   | 88 | 14,941   | 80 | 15,429   | 78 | 11,767   | 90 |
| >  0.8 m  | 4458     | 22 | 265      | 6  | 1710     | 12 | 3633     | 20 | 4458     | 22 | 1303     | 10 |
| ≤  1.0 m  | 15,710   | 79 | 4071     | 95 | 13,188   | 89 | 15,198   | 82 | 15,710   | 79 | 11,884   | 91 |
| >  1.0 m  | 4177     | 21 | 235      | 5  | 1578     | 11 | 3376     | 18 | 4177     | 21 | 1186     | 9  |
| Total     | 19,887   | -  | 4306     | -  | 14,766   | -  | 18,574   | -  | 19,887   | -  | 13,070   | -  |

### 3.4. Influence of Lake Geometry on WSE Anomalies

We examined the relationship between the geometry variables, identified by the RF analysis, with WSE precision for each parameter: area [km<sup>2</sup>], perimeter [km], circularity index [nondimensional], compactness coefficient [nondimensional], and minimum and maximum shore-to-shore distance [km] (Figure 6). Higher variance and difference values between SWOT and ONS were observed for smaller lakes, across all parameters in the heatmaps of Figure 6A–F and in the IQR boxplots (Figure 6G,H).



**Figure 6.** (A–F) WSE anomalies heatmap by lake's geometry parameters; and (G,H) IQR boxplots (no whiskers or outliers) of this difference grouped by area and perimeter (33 lakes per group).



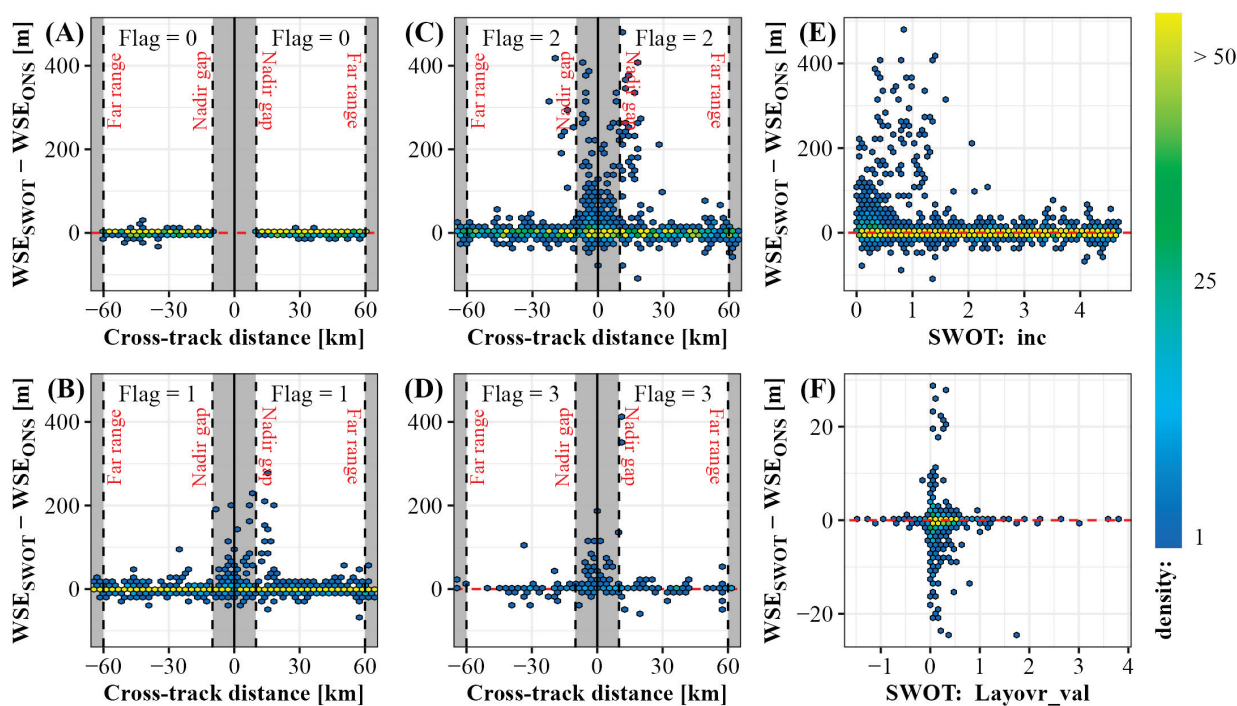
The lakes were grouped into four equally sized categories, ranging from very small to large, with 33 in each one. For each group, an IQR-based boxplot was drawn, only displaying the data within the 25th and 75th percentile. This excluded extreme values (Figure 6G,H) and showed a decrease in error variability with the increase in lake size (by both area and perimeter), with all IQR limits being under 25 cm. Designing groups with an equal number of lakes resulted in size classes with heterogeneous ranges.

Individual lake-geometry assessments were repeated separately for the three studied products (LakeSP\_Obs, Raster\_250m, and Raster\_100m), showing the same overall patterns across them, but with a more pronounced effect for the raster products, particularly for the very small lakes and short perimeters (Supplementary Material Figure S2).

### 3.5. SWOT-Specific Drivers of WSE Precision: Cross-Track, Angle, and Layover Effects

The precision of SWOT WSE measurements is influenced not only by the geometry of water bodies, but also by the observation geometry. Among these, three were found to be important according to the RF analysis (Figure 4), such as the relative position of an observation within the SWOT swath, the radar incidence angle, and the estimated error due to the layover effect. By examining these SWOT-specific drivers, we can better understand their influence on errors and broaden the understanding of the physical characteristics of radar-based observations and their interaction with Earth's water surface.

Figure 7 illustrates the relationship between WSE (SWOT–ONS) precision and cross-track distance. Our approach shows that the poorer-quality data are concentrated in the near range, although most observations have differences that are comparable to those observed in the mid-range of the KaRIn sensor (optimal range). The heatmap shows that most data points in flags 0, 1, and 2 (Figure 7A–C) are around the optimal line (difference = 0), although Flag = 2 has a much higher variance.



**Figure 7.** Heatmap of WSE differences (SWOT–ONS) against SWOT-derived factors. (A–D) Cross-track distance by quality flags; (E) Incidence angle of the measurements, in degrees; and (F) SWOT's estimated WSE error due to layover.

Two product-specific variables (*layovr\_val* and *inc*) are shown as influencing WSE precision from the RF analysis. *Layovr\_val* estimates WSE error due to layover and is only available in LakeSP, whereas *inc* is the incidence angle of the satellite for the measurement and is only available for the rasters. Figure 7E,F present heatmaps of the WSE differences between SWOT and ONS as a function of these variables. As expected, raster data (Figure 7E) exhibits greater variability (higher values for the differences) with more observations, while the vector data (Figure 7F) has greater stability at the cost of fewer data points.

As the incidence angle is related to the cross-track distance, the observed effects of the cross-track distance are extended to this variable. In contrast, the apparent relationship between *layovr\_val* and measurement performance does not necessarily imply a direct causal effect, as layover effects may be less pronounced under the relatively low topographic relief conditions of the large lakes analyzed in this study. Therefore, even though *layovr\_val* was identified by the RF as a potentially important variable, it was not considered a sufficiently robust standalone indicator of measurement quality in this context, and we chose not to include the layover error estimate in the proposed filtering strategy.

The cross-track distance, incidence angle, and the *layovr\_val* were also evaluated separately for each of the three studied products (LakeSP\_Obs, Raster\_250m, and Raster\_100m), showing the same overall patterns across them and confirming that near-nadir observations are consistently more problematic across the products, particularly for the raster products (Supplementary Material Figure S3).

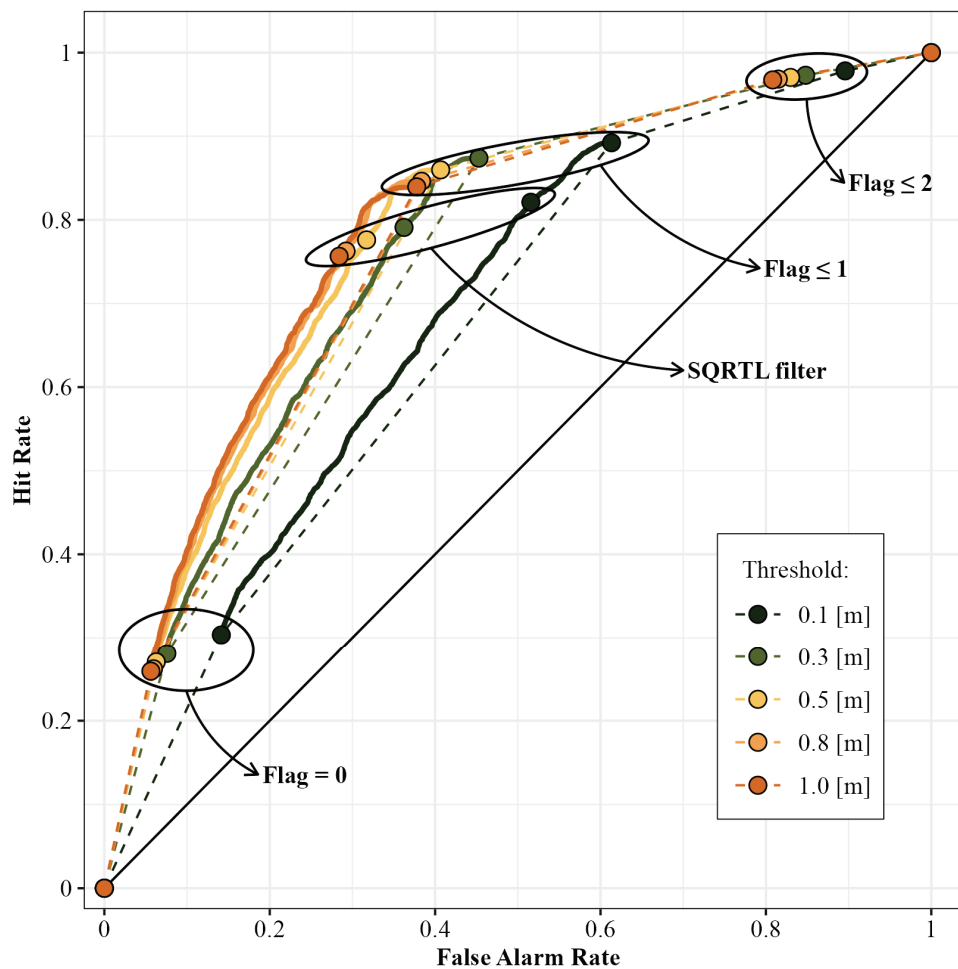
### 3.6. Proposal of Filtering Based on the Key Drivers

Based on the SWOT-derived variables assessed in the RF analysis, cross-track distance emerged as a key factor influencing WSE precision. Therefore, we propose its integration with the current flagging system to improve data availability while preserving the overall precision. The resulting SQRTL filter combines quality flags and observation range (cross-track distance), with ROC curves shown in Figure 8.

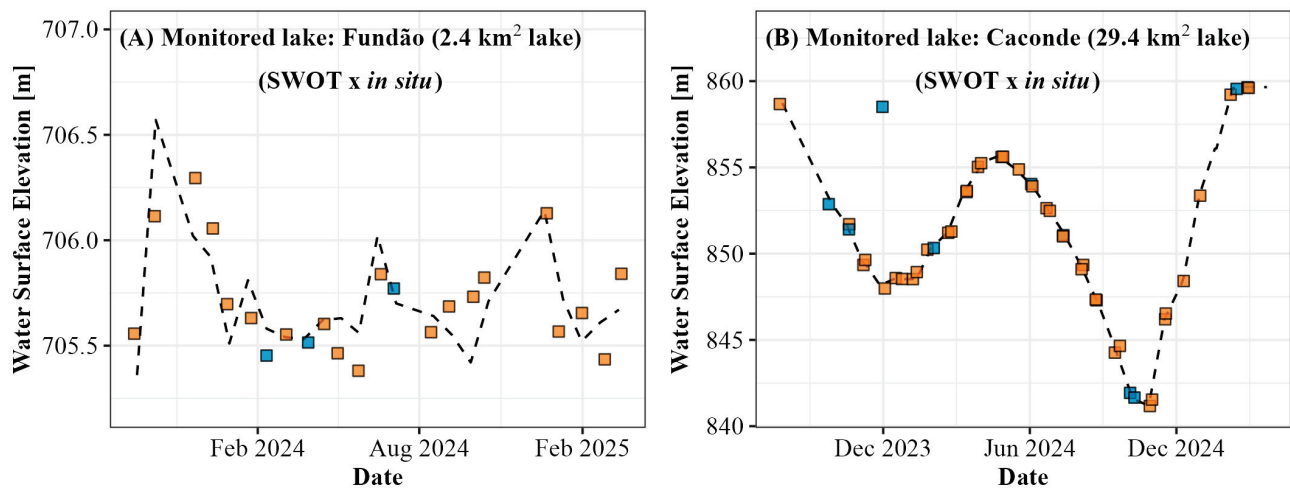
Figure 8 displays ROC curves (hit rate vs. false alarm rate) across selected thresholds (Table 2). Dashed lines represent increments in the performance of increasingly permissive flag combinations (e.g.,  $\text{Flag} \leq 0, 1, 2, 3$ ), with solid lines forming the Pareto front, i.e., the optimal result from a grid search calibration of cross-track distance boundaries. Highlighted points correspond to the use of specific flags or the proposed filtering method (i.e., SQRTL).

Across all thresholds,  $\text{Flag} = 0$  consistently exhibits low hit rates due to its restrictive nature, although with even lower false alarm rates, while  $\text{Flag} = 1$  achieves substantially higher hit rates, but at the cost of an also substantial increase in false alarm rates. To refine the selection within  $\text{Flag} = 1$ , we performed a grid search to identify the cross-track distance range that maximized the Euclidean distance from a ROC point (Hit Rate, False Alarm Rate) to the 45° line, optimizing the balance between sensitivity and specificity. The optimal range found was between 8100 m and 58,100 m, differing from the nominal 10 km to 60 km overall recommended window.

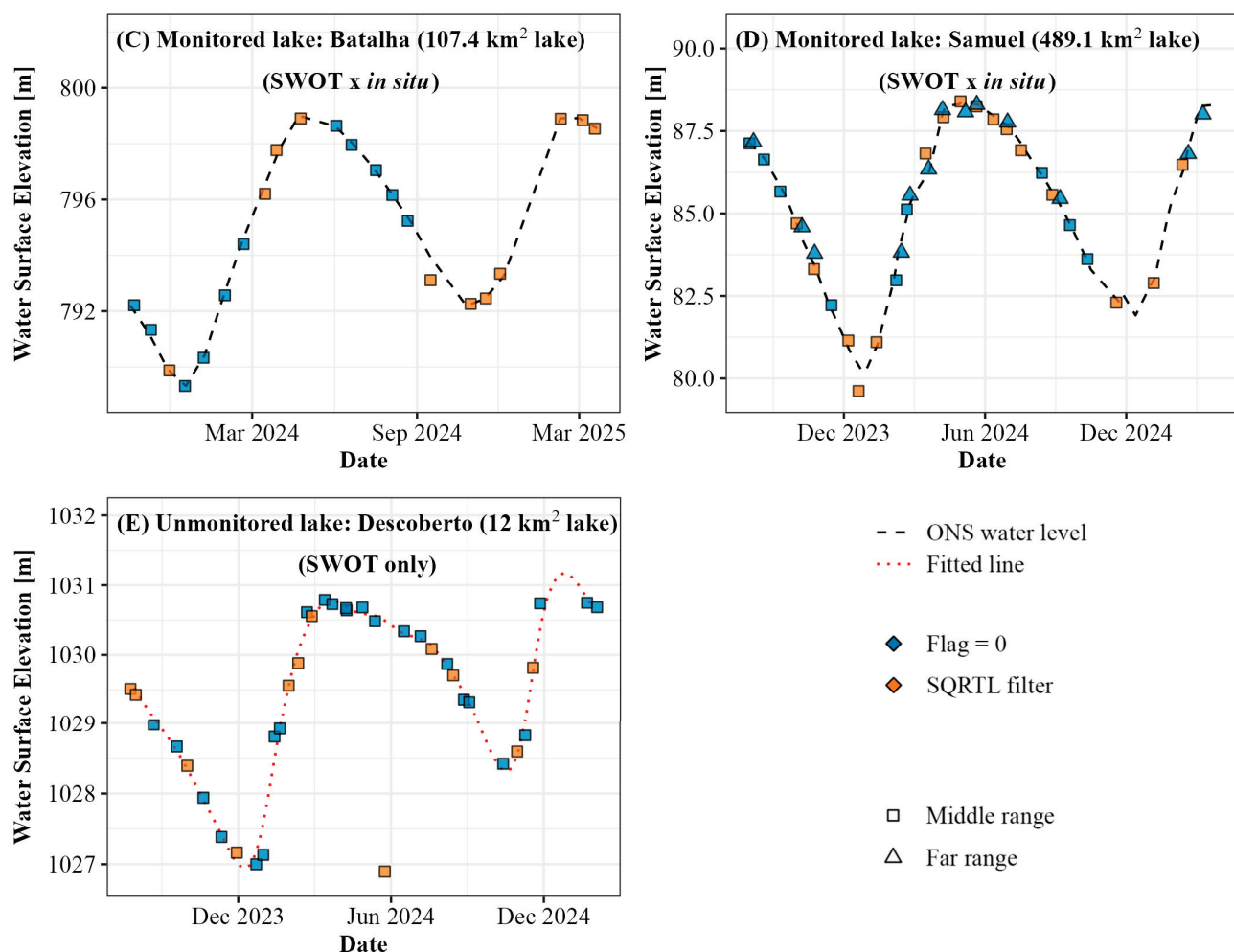
Figure 9 shows the potential of the proposed filter approach for SWOT lake water level monitoring, including in the absence of in situ data. The panels in Figure 9A–D show monitored lakes with different sizes (very small, small, medium, and large) and cross-track distances (near range below 10 km, middle range between 10 and 60 km, and far range above 60 km). This enables direct comparison between SWOT and in situ measurements.



**Figure 8.** ROC curves for different accepted thresholds, for the current flagging system and the proposed filter (SQRTL), which includes Flag = 0 and Flag = 1 constrained by cross-track distance.



**Figure 9.** Cont.



**Figure 9.** Illustrative example of SWOT temporal sampling and measurement quality for reservoir WSE monitoring using the LakeSP\_Obs product after applying the SQRTL filter with observations grouped into cross-track distance classes. (A–D) Four differently sized monitored lakes with *in situ* measurements from ONS (black dashed line); and (E) A 12 km<sup>2</sup> unmonitored lake, with only SWOT-derived WSE, and a dotted line representing a spline-based fit.

Figure 9A shows a “very small lake” (Fundão; 2.4 km<sup>2</sup>; 25.70°S, 52.00°W), Figure 9B a “small lake” (Caconde; 29.4 km<sup>2</sup>; 21.61°S, 46.58°W), Figure 9C a “medium-sized lake” (Batalha; 107.4 km<sup>2</sup>; 17.35°S, 47.49°W), and Figure 9D a “large lake” (Samuel; 489.1 km<sup>2</sup>; 8.75°S, 63.45°W). These sites were selected as representative of the overall dataset in terms of mean SWOT–ONS differences and revisit intervals. Figure 9E presents an unmonitored 12 km<sup>2</sup> lake (Descoberto; 15.75°S, 48.20°W), for which water levels are derived exclusively from SWOT observations.

#### 4. Discussion

Compared with conventional nadir altimeters (e.g., Jason-3, Sentinel-3, SARAL/AltiKa), SWOT represents a fundamental shift in inland water observation capability. Nadir missions are limited by narrow ground tracks and larger radar footprints, reducing their effectiveness for small and medium-sized lakes and increasing susceptibility to land contamination and topographic effects [16,59,60]. Although SAR processing and Ka-band altimetry have improved nadir performance [61,62], spatial coverage remains constrained by the nadir track sampling strategy. In contrast, SWOT provides two-dimensional spatial

coverage and enables the monitoring of lakes that are inaccessible to nadir altimeters, while maintaining strong consistency with independent satellite sources [31,40,63].

From our statistical analysis, it was shown that there is a key limitation of relying solely on global statistics for assessing measurement performance, as well as of the unadvised use of SWOT data without any screening. Nevertheless, the density distributions from Figure 3 show that most observations are concentrated near zero differences, indicating that most SWOT observations can capture accurate water level variations. For the LakeSP product, 59% of the observations fall within  $\pm 10$  cm, 82% within  $\pm 30$  cm, and 90% within  $\pm 1$  m without any filtering. Raster products show a similar behavior with lower precision, with approximately 43% of observations within  $\pm 10$  cm and 65% within  $\pm 30$  cm.

In the same analysis, the raster products contained approximately 73% more observations than LakeSP\_Obs during the 20-month study period, potentially introducing greater variability due to inclusion of lower-quality measurements, which are less frequent in LakeSP\_Obs since it employs different, often stricter, filtering [46,48]. Therefore, the higher variance observed in the raster products does not necessarily imply that the vector product is more reliable. When restricting analysis to dates when all three products were populated, raster standard deviations notably decreased from 4.12 m for Raster\_100m and 6.94 m for Raster\_250m in the full dataset to 5 cm and 4 cm, respectively, when restricted to those same dates and evaluated within the IQR. This suggests that significant variability is attributed to non-overlapping data.

For the RF evaluation, it is worth noting that the numerical values are relative rather than absolute and should be interpreted comparatively rather than quantitatively, e.g., a variable that is double the value of another does not imply that it is twice as important. From this analysis, the SWOT quality flag was shown to be a key variable in understanding WSE precision, emphasizing its role in SWOT-related data filtering. Other important variables were also identified, such as the cross-track distance and incidence angle (Figure 4). This aligns with known radar limitations at near and far ranges [50], especially within SWOT mission's Nadir gap (within 10 km of the satellite's center) and far range (beyond 60 km). Therefore, filtering strategies based on these variables (mainly quality flags and cross-track distances) may yield better results.

SWOT's KaRIn collects data across two swaths, each approximately 64 km wide. However, measurement performance is not uniform along the swath, with optimal precision expected within a 10 to 60 km range from the satellite's centerline [46,48]. The dependency of the WSE uncertainty on the position within the swath is included in some error models [26,29,49]. Consequently, more processed products (i.e., LakeSP) exclude data from the 10 km near range (Nadir gap), around the centerline and in far range areas beyond 60 km [48]. Figure 7 highlights the substantial amount of potentially useful data within the Nadir gap, which might be recovered with improved filtering criteria.

Far range observations do not appear to significantly impact WSE quality. This is expected since the main problem at that range is signal strength, not measurement precision. Additionally, the current 10 km Nadir gap might be insufficiently conservative for flags greater than 0 (suggested by Figure 7), as unreliable points appear outside this range, also noted by [35], who observed outliers up to 25 km from the nadir track. However, further investigation is necessary to optimize boundaries for both ranges. Even though users do not have control over the position of the lake in relation to the satellite ground track, they may have a choice of different passes, selecting those that offer more favorable viewing geometry in cases where a lake is observed by multiple passes. Thus, it is important to be aware that measurements in the near range are more likely to be affected by considerable measurement errors.



Because the incidence angle is intrinsically related to the cross-track distance, the observed effects of the cross-track distance can be extended to the incidence angle. Hence, as expected, values closer to  $0^\circ$  have more problematic measurements, with observations having  $inc < 2^\circ$  consistently showing a higher variance. Therefore, filtering methods based on this variable may improve data quality, similarly to the discussion of the cross-track distance. This can be particularly useful when analyzing an area that falls within multiple orbits, each having distinct incidence angles, with a possibility of relying more on those observations with higher angles.

The geometry of the lakes has been shown to be related to WSE precision as well. This aligns with [8], who noted that smaller lakes tend to have a greater surface area variability. Building on this, our results suggest a similar pattern but extended to WSE precision. A similar behavior, but less prominent, is found for the compactness coefficient and circularity index, implying a more accurate WSE from SWOT for more circular-shaped lakes (Figure 6). Therefore, the geometry and shape of the lake are important factors and should be noted before utilizing SWOT. However, unlike SWOT quality flags, one may not use them for screening strategies (a lake size and shape is mostly fixed).

Many studies using SWOT data rely on their quality flags to identify reliable measurements. Some studies use filters strictly based on them, using “good” or “suspect” flags, leading to limited data availability [35–37], while others further emphasized that flag-based filtering alone is often insufficient [21,22,34,39,40,42]. A potentially more robust strategy may be to use the factors influencing WSE alongside the quality flags.

Our results confirm that quality flags effectively categorize data by reliability, but the distribution of observations across flags is highly uneven. Flag = 0 retains only 22% of observations (a small fraction for many monitoring applications) with the 68th percentile errors below 12 cm, while Flag = 1 contains a large proportion of reliable measurements alongside higher variance data, increasing the 68th percentile errors up to 21 cm. For the full dataset, the 68th percentile errors are below 29 cm.

This suggests that targeted additional screening can recover meaningful observations without compromising overall precision. Compared to existing filtering approaches, SQRTL is intentionally minimalist in design. Studies that combine quality flags with statistical outlier removal [21,22,38,40–42] effectively reduce error variance but require paired in situ records or additional procedures, limiting their use over ungauged systems. In contrast, SQRTL relies only on two variables provided in every SWOT product (quality flag and cross-track distance).

By combining all Flag = 0 with cross-track constrained Flag = 1 observations, the SQRTL filter more than triples data availability while maintaining precision levels comparable to Flag = 0 alone (Table 2, Figure 8). The result is a subset of 66% of all observations with the 68th percentile errors below 16 cm, thus with the benefit of Flag = 1 data availability, but near Flag = 0 accuracy.

It becomes evident that a well-defined filtering strategy can improve the reliability of SWOT WSE estimates for hydrological applications. Studies report satisfactory performance only after discarding a significant fraction of available observations through quality flag screening and additional filtering procedures [21,22,34–37,39,40,42]. Collectively, these findings show that SWOT can provide accurate hydrological information when observations are carefully filtered and selected.

Studies using other satellite missions have also reported satisfactory results in WSE evaluation. Sentinel-6 RMSE values ranging from 5 to 50 cm have been reported [64], and RMSE values ranging from 4 to 36 cm have been reported for ICESat-1, ICESat-2, and GEDI [65]. Nevertheless, all these missions rely on much narrower sampling geometries

than SWOT's wide-swath capability. Therefore, SWOT can achieve comparable precision (with 68th percentile errors below 16 cm using the SQRTL filter) while having a broader spatial coverage and shorter revisit time.

The thresholds applied in the proposed framework here are not assumed to be universal, but instead reflect the hydrological, geomorphological, and orbital conditions of the study area. However, the framework itself demonstrates potential for transfer to wherever SWOT observations and reliable in situ water level data are available. Regional differences in geomorphology, shoreline complexity, hydrological regime, and SWOT pass geometry may lead to different optimal threshold values. Because the proposed filter is intentionally conservative, it may also serve as a baseline approach in regions where in situ data are sparse or unavailable.

The proposed SQRTL filter strategy inevitably reduces the number of retained data overall. However, it substantially increases data availability to using only Flag = 0 observations, with the average revisit time remaining adequate and proximate to the optimal values. With all available data, from which 54% of the measurements have errors in excess of 10 cm (Table 2), the median revisit times were 11, 10 and 13 days for the Raster\_100m, Raster\_250m and LakeSP products, respectively. However, using only Flag = 0, these revisit times increased to 88, 123 and 25 days, respectively. After applying the proposed filter, the median revisit times are 18, 16 and 14 days for Raster\_100m, Raster\_250m and LakeSP, respectively, showing a major improvement in contrast to Flag = 0, affording higher measurement quality with a modest decrease in temporal sampling.

As expected, the LakeSP product is more stable, as it showed only a small change relative to the unfiltered case, whereas the raster products exhibit a larger increase in revisit intervals. Even with filtering, the effective temporal sampling remains close to two observations per month on average. Given that the study period spans both major flood events and drought conditions in Brazil, the revisit frequency remains suitable for hydrological monitoring of reservoir water levels.

However, these gains in spatial detail come with a more complex observational geometry. SWOT's wide-swath interferometric design introduces error dependencies related to incidence angle, cross-track position, and surface characteristics [46,50,66–68]. Consequently, data quality assessment should not rely solely on conservative flag-based strategies when the objective is to balance precision and data availability for hydrological applications. Instead, filtering approaches must explicitly account for the physical and orbital factors governing when SWOT observations are most reliable.

The application of the proposed filter for monitored and unmonitored lakes (Figure 9) showed positive results for different sized lakes (2.4 km<sup>2</sup> to 489.1 km<sup>2</sup>) and different overpasses with varying distances from the Nadir gap (near to far range). The smooth temporal evolution captured demonstrates the ability of SWOT, when combined with the proposed filtering strategy, to characterize seasonal water level variability without any in situ measurements. Spurious data remain evident in panels 9B and 9E across quality flags 0 and 1, indicating that occasional anomalous observations persist and complementary outlier-removal or smoothing remains encouraged. However, the additional number of WSE observations is clear, showcasing the filter's potential.

By providing both a validated filtering strategy and a reproducible framework, this study contributes a practical foundation for expanding systematic lake monitoring, particularly in data-scarce regions where SWOT may represent the only available source of continuous water level information.

## 5. Conclusions

This study describes a scalable quality control and data filtering technique designed to facilitate the use of SWOT for monitoring (un)gauged lakes and evaluating the expected quality of WSE measurements. Here we evaluated the performance of these measurements across Brazil, demonstrating good precision with low standard deviations when focusing on the values within the IQR. As a novel satellite mission, SWOT significantly expands the spatial and temporal coverage of hydrological monitoring, especially for ungauged lakes and rivers. Importantly, it also delivers a rich set of supporting variables that can assist users in evaluating data quality.

An RF analysis identified key drivers of WSE precision based on the MDA metric. Among the top-ranked variables were some SWOT-derived variables (particularly the quality flags, the WSE uncertainty, and the cross-track distance), and geometric characteristics of lakes, with smaller and less circular lakes associated with higher WSE variability. These findings suggest that the PLD, which already characterizes each lake, could be enhanced to include shape-based indicators to support the current quality flagging process.

Most critically, this study highlights existing limitations in relying solely on SWOT's quality flags for filtering high-quality WSE data. Flag = 0 ("Good") may be overly restrictive for many hydrological applications, including only 22% of all observations, while still retaining anomalous data. Although the flags remain a useful starting point, they can be refined with improved filtering strategies. Here, we proposed the inclusion of observations with quality flags 0 and 1, retaining a subset of Flag = 1 observations constrained by a cross-track distance between 8.1 km and 58.1 km. Using these targeted filters as a complement to the existing flagging system provided better results, enhancing both data availability and reliability. While this threshold is regionally derived, the proposed framework could be transferable to wherever SWOT and in situ data are available or can be used as a baseline where only limited in situ data are available.

The proposed filter (SQRTL) strategy is particularly valuable for timeseries hydrology applications, where occasional extreme errors can dominate derived signals such as storage change estimates, anomaly detection, and model calibration, yet sufficient temporal sampling is required to resolve seasonal dynamics. With major improvements in revisit intervals (from 1 to 2 observations per month in LakeSP), effective sampling is adequate for many water resource applications. By enabling the systematic use of SWOT observations in poorly monitored systems, this framework has the potential to substantially expand lake and reservoir monitoring globally. For Brazil, it may increase the current coverage from 714 reservoirs to more than 100,000 lakes identified in the PLD, increasing the monitoring and availability of the data. An example over an ungauged lake showed the feasibility of this approach, with promising results.

Although the SQRTL framework can be applied elsewhere, the threshold values derived in this study are specific to the Brazilian dataset and should not be interpreted as universal. In addition, future work should evaluate the framework under different hydrological and geomorphological settings to assess its broader applicability.

Overall, this work demonstrates the value of advanced data screening techniques and strategic use of explanatory variables in strengthening the validity of satellite-derived hydrological data. While this study contributes to the current state of the art regarding radar use for hydrological applications, further research is needed on outlier detection and systematic bias assessment. More importantly, the results provide a practical foundation for incorporating SWOT WSE estimates into hydrological analysis and reservoir monitoring, particularly in poorly gauged regions where reliable water level information has previously been unavailable.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs18101609/s1>, Figure S1: Spatial distribution of the median WSE anomalies (SWOT–ONS) for the analyzed lakes. Figure S2: WSE precision related to the lake geometry for the three products (LakeSP\_Obs, Raster\_250m, and Raster\_100m). Figure S3: WSE precision related to the cross-track distance, incidence angle and layovr\_val for the three products (LakeSP\_Obs, Raster\_250m, and Raster\_100m).

**Author Contributions:** Conceptualization, T.L., D.B., L.O.S., R.P.d.M.F., R.E.C. and T.A.; methodology, T.L., G.G.P. and R.P.d.M.F.; software, T.L.; validation, T.L., G.G.P. and R.P.d.M.F.; resources, T.L., D.B., T.A., S.S. and R.E.C.; writing—original draft preparation, T.L.; writing—review and editing, T.L., D.B., L.O.S., T.A., G.G.P., R.P.d.M.F. and R.E.C.; visualization, T.L., T.A., R.P.d.M.F. and R.E.C.; supervision, T.A., R.P.d.M.F. and R.E.C.; project administration, T.A., S.S. and R.E.C.; funding acquisition, T.L., D.B., T.A., S.S. and R.E.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by ANA, grant number 04687620240003, and in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

**Data Availability Statement:** All data used in this study are available at <https://github.com/Lappicy/RS-manuscript>, accessed on 1 April 2026.

**Acknowledgments:** The authors thank the Brazilian National Water and Sanitation Agency (ANA) for its support and cooperation, as well as CAPES, CNPq, and FAPDF (grant numbers 23547.93.27555.30052018 and 00193.00000335/2026-19). R.P.d.M. Frasson would like to acknowledge the support received through the NASA Earth Action Water Resources program (Award number 21-WATER21-2-0004). A portion of this work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. The authors thank the reviewers for their feedback, which helped improve this paper.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

|       |                                                     |
|-------|-----------------------------------------------------|
| ANA   | Brazilian Water and Sanitation Agency               |
| HEPP  | Hydroelectric Power Plant                           |
| IQR   | Interquartile Range                                 |
| KaRIn | Ka-band Radar Interferometer                        |
| MDA   | Mean Decrease Accuracy                              |
| ONS   | Brazilian National Operator for the Electric System |
| PLD   | Prior Lake Database                                 |
| RF    | Random Forest                                       |
| ROC   | Receiver Operating Characteristic                   |
| SQRTL | SWOT Quality-Range Threshold for Lakes              |
| SWOT  | Surface Water and Ocean Topography                  |
| WSE   | Water Surface Elevation                             |

## References

1. Downing, J.A.; Prairie, Y.T.; Cole, J.J.; Duarte, C.M.; Tranvik, L.J.; Striegl, R.G.; McDowell, W.H.; Kortelainen, P.; Caraco, N.F.; Melack, J.M.; et al. The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol. Oceanogr.* **2006**, *51*, 2388–2397. [CrossRef]
2. Verpoorter, C.; Kutser, T.; Seekell, D.A.; Tranvik, L.J. A global inventory of lakes based on high-resolution satellite imagery. *Geophys. Res. Lett.* **2014**, *41*, 6396–6402. [CrossRef]

3. Messenger, M.L.; Lehner, B.; Grill, G.; Nedeya, I.; Schmitt, O. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nat. Commun.* **2016**, *7*, 13603. [CrossRef]
4. Pi, X.; Luo, Q.; Feng, L.; Xu, Y.; Tang, J.; Liang, X.; Ma, E.; Cheng, R.; Fensholt, R.; Brandt, M.; et al. Mapping global lake dynamics reveals the emerging roles of small lakes. *Nat. Commun.* **2022**, *13*, 5777. [CrossRef]
5. ANA, Agência Nacional de Águas e Saneamento Básico. Conjuntura dos Recursos Hídricos do Brasil 2025: Dados Análises, Tendências e Governança dos Recursos Hídricos. Relatório Pleno. Available online: <https://www.snirh.gov.br/portal/centrais-de-conteudos/conjuntura-dos-recursos-hidricos> (accessed on 1 April 2026).
6. Wang, J.; Pottier, C.; Cazals, C.; Battude, M.; Sheng, Y.; Song, C.; Sikder, M.S.; Yang, X.; Ke, L.; Gosset, M.; et al. The Surface Water and Ocean Topography Mission (SWOT) Prior Lake Database (PLD): Lake mask and operational auxiliaries. *Water Resour. Res.* **2025**, *61*, e2023WR036896. [CrossRef]
7. ANA, Agência Nacional de Águas e Saneamento Básico. System for Monitoring Reservoirs (SAR). Available online: <https://www.ana.gov.br/sar/> (accessed on 1 April 2026).
8. Bonnema, M.; David, C.H.; Frasson, R.P.d.M.; Oaida, C.; Yun, S.H. The global surface area variations of lakes and reservoirs as seen from satellite remote sensing. *Geophys. Res. Lett.* **2022**, *49*, e2022GL098987. [CrossRef]
9. Cheng, C.; Zhang, F.; Li, X.; Tan, M.L.; Kumar, P.; Johnson, B.A.; Shi, J.; Zhao, Q.; Liu, C. Variations in water storage of Bosten Lake, China, over the last two decades based on multi-source satellite data. *J. Hydrol. Reg. Stud.* **2023**, *49*, 101496. [CrossRef]
10. Farouk, M.I.H.Z.; Jamil, Z.; Latip, M.F.A. Towards online surface water quality monitoring technology: A review. *Environ. Res.* **2023**, *238*, 117147. [CrossRef] [PubMed]
11. Huang, Z.; Sun, R.; Wang, H.; Wu, X. Trends and Innovations in Surface Water Monitoring via Satellite Altimetry: A 34-Year Bibliometric Review. *Remote Sens.* **2024**, *16*, 2886. [CrossRef]
12. Beltrão, D.; Cicerelli, R.E.; Lappicy, T.; Sales, L.O.; Teixeira, A.A.; Cunha, L.S.; Andrade, A.C.F.; Almeida, T. SWOT: Reviewing challenges and opportunities in hydrological applications to inland waters. *IEEE Access* **2026**, *early access*. [CrossRef]
13. Vignudelli, S.; Scozzari, A.; Abileah, R.; Gómez-Enri, J.; Benveniste, J.; Cipollini, P. Water Surface Elevation in Coastal and Inland Waters Using Satellite Radar Altimetry. In *Extreme Hydroclimatic Events and Multivariate Hazards in a Changing Environment: A Remote Sensing Approach*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 87–127. [CrossRef]
14. Du, B.; Li, J.C.; Jin, T.Y.; Zhou, M.; Gao, X.W. Synthesis Analysis of SWOT KaRIn-Derived Water Surface Heights and Local Cross-Calibration of the Baseline Roll Knowledge Error Over Lake Baikal. *Earth Space Sci.* **2021**, *8*, e2021EA001990. [CrossRef]
15. Enguehard, P.; Frappart, F.; Zeiger, P.; Blarel, F.; Satgé, F.; Bonnet, M.P. Contribution of automatically generated radar altimetry water levels from unsupervised classification to study hydrological connectivity within Amazon floodplains. *J. Hydrol. Reg. Stud.* **2023**, *47*, 101397. [CrossRef]
16. Le Traon, P.Y.; Dibarboure, G.; Lellouche, J.M.; Pujol, M.I.; Benkiran, M.; Drevillon, M.; Drillet, Y.; Faugère, Y.; Remy, E. Satellite altimetry and operational oceanography: From Jason-1 to SWOT. *Ocean Sci.* **2025**, *21*, 1329–1347. [CrossRef]
17. Bonassies, Q.; Fatras, C.; Peña-Luque, S.; Dubois, P.; Piacentini, A.; Cassan, L.; Ricci, S.; Nguyen, T.H. A comprehensive study of Surface Water and Ocean Topography (SWOT) Pixel Cloud data for flood extent extraction. *Remote Sens. Environ.* **2026**, *333*, 115101. [CrossRef]
18. Solander, K.C.; Reager, J.T.; Famiglietti, J.S. How well will the Surface Water and Ocean Topography (SWOT) mission observe global reservoirs? *Water Resour. Res.* **2016**, *52*, 2123–2140. [CrossRef]
19. Xiong, J.; Jiang, L.; Qiu, Y.; Wongchuig, S.; Abhishek, G.S.; Chen, J. On the Capabilities of the SWOT Satellite to Monitor the Lake Level Change over the Third Pole. *Environ. Res. Lett.* **2023**, *18*, 044008. [CrossRef]
20. Riggs, R.M.; Dickinson, J.E.; Brinkerhoff, C.B.; Sikder, M.S.; Wang, J.; Gao, H.; Allen, G.H. Characterizing operational signatures of reservoirs with the SWOT satellite by comparing natural lake and reservoir dynamics. *Environ. Res. Lett.* **2026**, *21*, 044008. [CrossRef]
21. Jing, R.; Liao, J.; Ma, S.; Liu, X.; Wu, Y. Exploring the performance of SWOT satellite to monitor lake volumes: A case study of Chinese lakes. *J. Remote Sens.* **2026**, *6*, 1026. [CrossRef]
22. Zhan, P.; Wang, J.; Chen, T.; Luo, S.; Liu, K.; Ke, L.; Fan, C.; Lin, Y.; Song, C. Integrating SWOT with multi-source satellite observations for near-daily reservoir water level monitoring. *Water Resour. Res.* **2026**, *62*, e2024WR039711. [CrossRef]
23. Wongchuig-Correa, S.; Paiva, R.C.D.; Biancamaria, S.; Collischonn, W. Assimilation of future SWOT-based river elevations, surface extent observations and discharge estimations into uncertain global hydrological models. *J. Hydrol.* **2020**, *590*, 125473. [CrossRef]
24. Lubczynski, M.W.; Leblanc, M.; Batelaan, O. Remote sensing and hydrogeophysics give a new impetus to integrated hydrological models: A review. *J. Hydrol.* **2024**, *633*, 130901. [CrossRef]
25. Domeneghetti, A.; Schumann, G.P.; Frasson, R.P.M.; Wei, R.; Pavelsky, T.M.; Castellarin, A.; Brath, A.; Durand, M.T. Characterizing water surface elevation under different flow conditions for the upcoming SWOT mission. *J. Hydrol.* **2018**, *561*, 848–861. [CrossRef]



26. Durand, M.; Chen, C.; Frasson, R.P.M.; Pavelsky, T.M.; Williams, B.; Yang, X.; Fore, A. How will radar layover impact SWOT measurements of water surface elevation and slope, and estimates of river discharge? *Remote Sens. Environ.* **2020**, *247*, 111883. [CrossRef]
27. Durand, M.; Dai, C.; Moortgat, J.; Yadav, B.; Frasson, R.P.M.; Li, Z.; Wadkwoski, K.; Howat, I.; Pavelsky, T.M. Using river hypsometry to improve remote sensing of river discharge. *Remote Sens. Environ.* **2024**, *315*, 114455. [CrossRef]
28. Meem, T.H.; Rhoads, B.L.; Wang, J.; Frasson, R.P.M. Detecting water-surface superelevation in meandering rivers using Surface Water and Ocean Topography (SWOT) satellite data. *Geophys. Res. Lett.* **2026**, *53*, e2025GL119167. [CrossRef]
29. Frasson, R.P.M.; Schumann, G.J.P.; Kettner, A.J.; Brakenridge, G.R.; Krajewski, W.F. Will the Surface Water and Ocean Topography (SWOT) Satellite Mission Observe Floods? *Geophys. Res. Lett.* **2019**, *46*, 10435–10445. [CrossRef]
30. Altenau, E.H.; Pavelsky, T.M.; Durand, M.T.; Yang, X.; Frasson, R.P.M.; Bendezu, L. The Surface Water and Ocean Topography (SWOT) Mission River Database (SWORD): A Global River Network for Satellite Data Products. *Water Resour. Res.* **2021**, *57*, e2021WR030054. [CrossRef]
31. Biancamaria, S.; Lettenmaier, D.P.; Pavelsky, T.M. The SWOT mission and its capabilities for land hydrology. In *Survey in Geophysics*; Springer International Publishing: Cham, Switzerland, 2016; pp. 117–147. [CrossRef]
32. Moreira, D.; Papa, F.; Fassoni-Andrade, A.; Fleischmann, A.; Wongchuig, S.; Paiva, R.; Paris, A.; Frappart, F.; Melo, J.; Crétaux, J.F.; et al. Widespread and exceptional reduction in river water levels across the Amazon Basin during the 2023 extreme drought revealed by satellite altimetry and SWOT. *Geophys. Res. Lett.* **2025**, *52*, e2025GL116180. [CrossRef]
33. Yao, J.; Xu, N.; Wang, M.; Liu, T.; Lu, H.; Cao, Y.; Tang, X.; Mo, F.; Chang, H.; Gong, H.; et al. SWOT satellite for global hydrological applications: Accuracy assessment and insights into surface water dynamics. *Int. J. Digit. Earth* **2025**, *18*, 2472924. [CrossRef]
34. Xu, Y.; Frappart, F.; Tang, G.; Zhang, G.; Lin, P.; Jiang, L.; Papalexiou, S.; Yao, F.; Han, X.; Xia, J. A global intercomparison of SWOT and traditional nadir radar altimetry for monitoring river water surface elevation. *Remote Sens. Environ.* **2026**, *334*, 115219. [CrossRef]
35. Kica, S.; Pavelsky, T.M.; Fayne, J.V.; Williams, B.A. SWOT Water Surface Elevation in Herbaceous Wetlands of Florida's Everglades. *Geophys. Res. Lett.* **2025**, *52*, e2025GL114956. [CrossRef]
36. Laipelt, L.; de Paiva, R.C.; Fan, F.M.; Collischonn, W.; Papa, F.; Ruhoff, A. SWOT reveals how the 2024 disastrous flood in South Brazil was intensified by increased water slope and wind forcing. *Geophys. Res. Lett.* **2025**, *52*, e2024GL111287. [CrossRef]
37. Maubant, L.; Dodd, L.; Tregoning, P. Assessing the accuracy of SWOT measurements of water bodies in Australia. *Geophys. Res. Lett.* **2025**, *52*, e2024GL114084. [CrossRef]
38. Das, P.; Hossain, F. Multi-satellite tracking of surface water storage change in the era of Surface Water and Ocean Topography (SWOT) satellite mission. *Earth Space Sci.* **2025**, *12*, e2024EA004178. [CrossRef]
39. Patidar, G.; Indu, J.; Karmakar, S. Performance assessment of Surface Water and Ocean Topography (SWOT) mission for WSE measurement across India. *Geophys. Res. Lett.* **2025**, *52*, e2025GL115804. [CrossRef]
40. Wu, S.; Cai, Y.; Ke, C.Q.; Xiao, Y.; Li, H.; He, Z.; Duan, Z. SWOT mission enables high-precision and wide-coverage lake water levels monitoring on the Tibetan Plateau. *J. Hydrol. Reg. Stud.* **2025**, *59*, 102357. [CrossRef]
41. Hamoudzadeh, A.; Ravanelli, R.; Crespi, M. SWOT Level 2 Lake Single-Pass Product: The L2\_HR\_LakeSP Data Preliminary Analysis for Water Level Monitoring. *Remote Sens.* **2024**, *16*, 1244. [CrossRef]
42. Sales, L.O.; Lappicy, T.; Beltrão, D.; Teixeira, A.A.; Cicerelli, R.; Almeida, T. SWOT Satellite Nodes as Virtual Stations During the 2024 Extreme Flood in Southern Brazil. *Hydrology* **2025**, *12*, 248. [CrossRef]
43. ONS, Operador Nacional do Sistema Elétrico. Brazilian National Operator for the Electric System. Hydraulic Data by Reservoir-Hourly Basis. Available online: [https://dados.ons.org.br/dataset/dados\\_hidrologicos\\_ho](https://dados.ons.org.br/dataset/dados_hidrologicos_ho) (accessed on 1 April 2026).
44. ANA, Agência Nacional de Águas e Saneamento Básico. Massas d'água. Brazilian National Water and Sanitation Agency's (ANA) Metadata Catalog. Available online: <https://metadados.snirh.gov.br/geonetwork/srv/eng/catalog.search#/metadata/7d054e5a-8cc9-403c-9f1a-085fd933610c> (accessed on 1 April 2026).
45. Brazil. Conselho Nacional de Recursos Hídricos. Resolução nº 32, de 15 de Outubro de 2003. Dispõe Sobre as Regiões Hidrográficas do Brasil. Available online: <https://metadados.snirh.gov.br/geonetwork/srv/eng/catalog.search#/metadata/fb87343a-cc52-4a36-b6c5-1fe05f4fe98c> (accessed on 1 April 2026).
46. JPL. *SWOT Science Data Products User Handbook*; Jet Propulsion Laboratory Internal Document: Pasadena, CA, USA, 2026.
47. JPL. *SWOT Product Description. Level 2 KaRIn High Rate Raster Product Short Name: L2\_HR\_Raster. Revision C*; Internal Document; Jet Propulsion Laboratory: Pasadena, CA, USA, 2026.
48. SWOT. *SWOT Product Description. Level 2 KaRIn High Rate Lake Single Pass Vector Product. Short Name: L2\_HR\_LakeSP Revision C*; Internal Document SWOT-TN-CDM-0673-CNES; Jet Propulsion Laboratory: Pasadena, CA, USA, 2024.
49. Frasson, R.P.M.; Ardila, D.R.; Pease, J.; Hestir, E.; Bright, C.; Carter, N.; Dekker, A.G.; Thompson, D.R.; Green, R.O.; Held, A. The impact of spatial resolution on inland water quality monitoring from space. *Environ. Res. Commun.* **2024**, *6*, 101003. [CrossRef]

50. JPL. *SWOT Project Mission Performance and Error Budget, Revision A*; Jet Propulsion Laboratory: Pasadena, CA, USA, 2017; 117p.
51. Das, B.C.; Islam, A.; Sarkar, B. Drainage basin shape indices to understanding channel hydraulics. *Water Resour. Manag.* **2022**, *36*, 2523–2547. [CrossRef]
52. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
53. Breiman, L.; Cutler, A.; Liaw, A.; Wiener, M. *Package 'Randomforest'*; University of California: Berkeley, CA, USA, 2018.
54. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*, 2nd ed.; Springer: New York, NY, USA, 2021. [CrossRef]
55. Mushagalusa, C.A.; Fandohan, A.B.; Kakai, R.G. Random forest and spatial cross-validation performance in predicting species abundance distributions. *Environ. Syst. Res.* **2024**, *13*, 23. [CrossRef]
56. Liu, H.; Xia, R.; Chen, Y.; Jia, R.; Wei, Y.; Yan, C.; Li, L.; Zhang, K.; Wang, Y.; Li, X. Spatial patterns of hydroecological health in the semi-arid yellow river basin: Revelations from machine learning models. *Ecol. Indic.* **2024**, *168*, 112799. [CrossRef]
57. Zhou, Y.; Wei, G.; Wang, Y.; Wang, B.; Quan, Y.; Wu, Z.; Liu, J.; Bian, S.; Li, M.; Fan, W.; et al. Estimating Regional Forest Carbon Density Using Remote Sensing and Geographically Weighted Random Forest Models: A Case Study of Mid-to High-Latitude Forests in China. *Forests* **2025**, *16*, 96. [CrossRef]
58. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]
59. Michailovsky, C.I.; McEnnis, S.; Berry, P.A.M.; Smith, R.; Bauer-Gottwein, P. River monitoring from satellite radar altimetry in the Zambezi River basin. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 2181–2192. [CrossRef]
60. Schwatke, C.; Dettmering, D.; Bosch, W.; Seitz, F. DAHITI—an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 4345–4364. [CrossRef]
61. Nair, A.S.; Kumar, N.; Indu, J.; Vivek, B. Monitoring lake levels from space: Preliminary analysis with SWOT. *Front. Water* **2021**, *3*, 717852. [CrossRef]
62. Nair, A.S.; Soman, M.K.; Girish, P.; Karmakar, S.; Indu, J. Evaluating SWOT water level information using a large scale hydrology simulator: A case study over India. *Adv. Space Res.* **2022**, *70*, 1362–1374. [CrossRef]
63. Morrow, R.; Fu, L.L.; Ardhuin, F.; Benkiran, M.; Chapron, B.; Cosme, E.; d'Ovidio, F.; Farrar, J.T.; Gille, S.T.; Lapeyre, G.; et al. Global Observations of Fine-Scale Ocean Surface Topography with the Surface Water and Ocean Topography (SWOT) Mission. *Front. Mar. Sci.* **2019**, *6*, 232. [CrossRef]
64. Domingo, X.; Gilbert, F.; Molina, R.; Escorihuela, M.J. Improved Inland Water Level Estimates with Sentinel-6 Fully Focused SAR Processing: A Case Study in the Ebre River Basin. *Remote Sens.* **2025**, *17*, 531. [CrossRef]
65. Xiang, J.; Li, H.; Zhao, J.; Cai, X.; Li, P. Inland water level measurement from spaceborne laser altimetry: Validation and comparison of three missions over the Great Lakes and lower Mississippi River. *J. Hydrol.* **2021**, *597*, 126312. [CrossRef]
66. Fjørtoft, R.; Gaudin, J.M.; Pourthie, N.; Lalaurie, J.C.; Mallet, A.; Nouvel, J.F.; Martinot-Lagarde, J.; Oriot, H.; Borderies, P.; Ruiz, C.; et al. KaRIn on SWOT: Characteristics of Near-Nadir Ka-Band Interferometric SAR Imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2172–2185. [CrossRef]
67. Desrochers, N.M.; Peters, D.L.; Siles, G.; Charest, E.C.; Trudel, M.; Leconte, R. A Remote Sensing View of the 2020 Extreme Lake-Expansion Flood Event into the Peace-Athabasca Delta Floodplain-Implications for the Future SWOT Mission. *Remote Sens.* **2023**, *15*, 1278. [CrossRef]
68. Salameh, E.; Desroches, D.; Deloffre, J.; Fjørtoft, R.; Mendoza, E.T.; Turki, I.; Froideval, L.; Levailant, R.; Déchamps, S.; Picot, N.; et al. Evaluating SWOT's interferometric capabilities for mapping intertidal topography. *Remote Sens. Environ.* **2024**, *314*, 114401. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## Article

# An Intelligent Gated Fusion Network for Waterbody Recognition in Multispectral Remote Sensing Imagery

Tong Zhao, Chuanxun Hou \*, Zhili Zhang and Zhaofa Zhou

Intelligent Control Laboratory, Force University of Engineering, Xi'an 710025, China;  
2021226056@chd.edu.cn (T.Z.)

\* Correspondence: houcxun@sina.com

## Highlights

### What are the main findings?

- This study proposes a novel Intelligent Gated Fusion Network (IGF-Net). The dual-branch feature encoder is designed to alleviate the input channel mismatch between pre-trained RGB models and multi-band data. Its core Intelligent Gated Fusion Module (IGFM) facilitates adaptive fusion of spectral and visual features.
- Extensive experiments indicate that IGF-Net achieves highly competitive performance (IoU: 0.8742, Dice: 0.9239) on the newly constructed dataset and shows favorable generalization on an independent Sentinel-2 dataset, performing competitively compared with mainstream segmentation models.

### What are the implications of the main findings?

- The work provides an effective and robust deep learning solution for accurate water body recognition from multispectral imagery, which can directly benefit practical applications such as hydrological monitoring, environmental management, and disaster assessment.
- We construct and publicly release the “Tiangong-2 Remote Sensing Image Water Body Semantic Segmentation Dataset” along with the complete model implementation code, offering a valuable benchmark and reproducible research foundation for the community.

## Abstract

Accurate water body segmentation from multispectral remote sensing imagery is critical for hydrological monitoring and environmental management. However, leveraging transfer learning with pre-trained models remains challenging due to the dimensional mismatch between three-channel RGB-based architectures and multi-band spectral data. To address this, this study proposes a novel segmentation network, termed Intelligent Gated Fusion Network (IGF-Net), built upon a dual-branch feature encoder module and a core Intelligent Gated Fusion Module (IGFM). The IGFM achieves adaptive fusion of visual and spectral features through a cascaded mechanism integrating differences-and-commonalities parallel modeling, channel-context priors, and adaptive temperature control. We evaluate IGF-Net on the newly constructed Tiangong-2 remote sensing image water body semantic segmentation dataset, which comprises 3776 meticulously annotated multispectral image patches. Comprehensive experiments demonstrate that IGF-Net achieves strong and consistent performance on this dataset, with an Intersection over Union of 0.8742 and a Dice coefficient of 0.9239, consistently outperforming the evaluated baseline methods, such as FCN, U-Net, and DeepLabv3+. It also exhibits strong cross-dataset generalization

capabilities on an independent Sentinel-2 water segmentation dataset. Ablation studies and visualization analyses confirm that the proposed fusion strategy significantly enhances segmentation accuracy and stability, particularly in complex scenarios. Placeholder.

**Keywords:** deep learning; semantic segmentation; intelligent gated fusion; water-body recognition

## 1. Introduction

Water bodies constitute the general designation for water aggregates including rivers, lakes, and oceans. Rapid and accurate identification of water bodies from remote sensing imagery holds significant practical value for applications such as flood prediction, coastline monitoring, water resource evaluation, and agricultural planning. The extraction of water body information from remote sensing data through digital image processing techniques represents a crucial research focus within the field of remote sensing applications. Relevant scholars in this domain have developed multiple water body recognition methodologies, which can be systematically classified into three primary categories: traditional methods, machine-learning-based methods, and deep-learning-based methods.

Traditional methods primarily include visual interpretation and water index approaches. Among these, visual interpretation refers to the process where experts identify water body areas by observing differences in color and shape of features within remote sensing imagery. While offering high accuracy, this method heavily depends on expert experience and is labor-intensive, leading to low efficiency. Water index methods leverage the spectral characteristics of water bodies to construct mathematical models through band selection in remote sensing imagery. Threshold segmentation is then applied to achieve land-water separation. Commonly used water indices include the Normalized Difference Water Index (NDWI) [1], Revised Normalized Difference Water Index (RNDWI) [2], Shadow Water Index (SWI) [3], and Enhanced Water Index (EWI) [4]. Different indices are suitable for distinct application scenarios; for instance, EWI performs better in semi-arid regions, while SWI is more applicable to mountainous water bodies.

Machine-learning-based methods for water body identification first require manual feature selection. These features can be regarded as abstract representations of water body information within remote sensing imagery. Subsequently, machine learning classifiers use these features during training. Once trained, the classifiers can rapidly identify water body pixels in remote sensing imagery based on these features. Commonly used classifiers include decision trees [5], support vector machines [6], and other algorithms. Compared to water index methods, machine learning approaches typically achieve higher identification accuracy. However, their performance heavily depends on feature selection, and such methods often exhibit limited generalization capability.

Following the breakthrough of AlexNet [7] in the 2012 ImageNet [8] competition, deep learning methods have rapidly advanced and have been extensively applied to water body identification tasks in remote sensing imagery. These methods typically rely on large volumes of annotated samples to train neural networks for the accurate identification of water pixels. Generally speaking, existing research has proceeded along two primary trajectories: one focusing on water body extraction in challenging scenarios, enhancing adaptability to boundary ambiguity, shadow interference, complex backgrounds, and cross-regional variations through network architecture improvements; the other emphasizing

multispectral information utilization by integrating non-visible bands (e.g., NIR and SWIR) to enhance the model's discriminative capabilities between water and non-water features.

Regarding water extraction in complex scenarios, researchers have undertaken extensive structural optimization efforts targeting specific application contexts. For instance, to tackle the challenges of subtle spectral/spatial feature differences between water bodies and background, blurred boundaries, and difficulties in extracting small water bodies within tidal flat environments, Zhang et al. developed FYOLOv3. This approach accomplishes fine-grained extraction of tidal flat water bodies through enhancements to YOLOv3, incorporating both a pooling-free fully convolutional network architecture and pixel-wise similarity discriminative mechanisms [9]. To address the inadequacies of baseline semantic segmentation models in complex remote sensing imagery and the challenge of blurred water body boundaries, Weng et al. developed a U-Net-based framework incorporating the Object Context module, while integrating Atrous Spatial Pyramid Pooling (ASPP) and self-attention mechanisms. This architectural design enhances multi-scale contextual feature extraction capabilities and improves water boundary segmentation accuracy [10]. In ultra-high spatial resolution urban scenarios, Li et al. conducted a systematic evaluation of the water body extraction efficacy of a fully convolutional network (FCN) under constrained training sample conditions. The experimental results demonstrate that this FCN-based approach effectively leverages spatial contextual information to differentiate water bodies from shadow regions, achieving overall performance superiority over benchmark methods including NDWI and SVM [11]. In the context of global-scale water body identification, Isikdogan et al. introduced DeepWater Map, which formulates water extraction as a multi-scale fully convolutional semantic segmentation framework. This methodology significantly reduces the dependence of conventional spectral-index-based approaches on empirically determined thresholds while simultaneously minimizing misclassification errors associated with snow, ice, cloud cover and shadowed regions [12]. To tackle challenging interference conditions in agricultural watersheds including shadow effects, water quality variability, and vegetation occlusion, Liao et al. developed a lightweight architecture termed LKF-DCANet. This framework utilizes deformable convolutions for adaptive geometric feature extraction of water bodies, while incorporating learnable Kalman filtering during feature decoding for denoising. While maintaining a compact parameter footprint, the approach demonstrates competitive segmentation performance, with additional gains achieved through subsequent knowledge distillation [13]. The EU-Net developed by Cao et al. strengthens multi-scale contextual modeling through dilated convolutions at multiple scales, attention mechanisms, and enhanced residual connections. This architecture integrates multi-scale feature fusion to deliver superior boundary segmentation accuracy, particularly for extracting complex-textured water bodies, small-scale aquatic features, and narrow channel regions [14]. To address the paucity of labeled samples for small water body delineation in sub-meter resolution satellite imagery, Li et al. established a cross-sensor transfer learning framework bridging Sentinel-2 and PlanetScope datasets. Their systematic evaluation of the state-space model VMamba demonstrates its promising potential for small water body mapping, thereby providing an innovative computational approach for precise extraction of hydrological features from ultra-high-resolution imagery [15].

Conversely, non-visible-band spectral information from multispectral remote sensing imagery, particularly NIR and SWIR, substantially enhances spectral discrimination capabilities for water-body identification. A comparative investigation by Ngo et al. utilizing unmanned aerial vehicle-acquired multispectral data demonstrated that fusing ancillary features including NIR reflectance and NDWI yields statistically significant improvements in water detection and segmentation accuracy within complex humid tropical ecosystems,



outperforming RGB-only input configurations [16]. Building upon these insights, researchers have developed multiple fusion-based segmentation networks specifically designed for multi-spectral information utilization. Weng et al. introduced SCR-Net, which employs a four-band RGB + NIR input configuration within a dual-branch encoder-decoder framework. This architecture utilizes a ConvFormer branch for global context modeling, while integrating a ResNet-50-based residual pathway with a GAM channel-spatial attention mechanism to enhance feature representation and detail delineation. This design enables effective multispectral information fusion and fine-scale water body segmentation [17]. To tackle the challenge of heterogeneous spatial resolutions across Sentinel-2 multispectral bands and feature degradation induced by conventional interpolation-based upsampling, Yuan et al. developed an end-to-end multi-channel water body detection network named MC-WBDN. This architecture performs front-end fusion of RGB, NIR, and SWIR spectral bands while integrating enhanced atrous spatial pyramid pooling modules. The proposed approach significantly enhances multispectral water segmentation accuracy and robustness against illumination variations and atmospheric disturbances, demonstrating particular efficacy in detecting micro-scale hydrological features [18]. To tackle the challenges of spectral reflectance instability in complex aquatic environments and interference from reflections, ripples, and specular highlights, Hu et al. developed a shape-spectrum fusion-based water detection approach leveraging multi-frame multispectral imagery. The methodology establishes a dual-path feature extraction architecture with separate spectral-spatial processing streams. In the spectral pathway, temporal-spectral fusion significantly reduces water-surface dynamic disturbances, while channel-wise and spatial attention mechanisms are synergistically embedded in both the spatial branch and fusion stages, achieving high-precision and robust pixel-level water segmentation [19]. Beyond multispectral information, researchers have incorporated active-mode microwave remote sensing data (e.g., SAR) to develop multi-source/multi-modal fusion approaches, aiming to enhance water segmentation and recognition capabilities in challenging environments. To overcome the inherent limitations of standalone optical or microwave remote sensing data for small water body identification in rugged terrain regions, Yang et al. introduced a multispectral-SAR image fusion algorithm named MASF. This method successfully accomplishes synergistic fusion of both data modalities through a multi-resolution analytical framework. By subsequently integrating spectral, textural, and geometric features with multi-scale segmentation and RF-based classification, the approach achieves precise delineation of small water bodies in complex topographic settings [20]. Wang et al. developed MFGF-UNet and established the WIPI multimodal dataset, which synthesizes Sentinel-2 water-index-derived features with Sentinel-1 polarimetric information. The architecture incorporates a gated multi-scale filter Inception module and channel-wise attention-based GCT skip connections, significantly improving multi-scale water body feature representation. This approach exhibits superior segmentation accuracy and generalization capability across diverse benchmarks (WIPI, Chengdu, GF2020) while maintaining high parameter efficiency [21]. In addition, related studies in other vision tasks have also explored attention mechanisms and feature enhancement strategies to improve boundary representation and robustness under complex backgrounds [22].

Collectively, contemporary deep learning approaches have demonstrated substantial advancements in both scenario adaptability and multispectral information utilization. Notably, multi-spectral fusion segmentation networks have validated that non-visible-band spectral data significantly enhances water-body identification efficacy under complex environmental conditions. For systematic comparison of the aforementioned representative methodologies, Table 1 synthesizes their principal input modalities, core architectural designs, and distinctive methodological features. Comprehensive analysis of Table 1 reveals persistent

common limitations across current approaches. In terms of fusion strategies, contemporary approaches predominantly employ conventional paradigms involving unified multi-band inputs or channel-level concatenation. While these methods enable partial integration of multi-source information, they critically lack explicit differentiation between the distinct learning objectives associated with visible and non-visible-band spectral modalities. This fundamental limitation severely constrains the potential synergy that could be derived from the complementary characteristics of these two data domains. For example, while SCR-Net employs a dual-branch architecture, both branches process identical four-channel RGB + NIR inputs. The distinction between branches predominantly manifests in their architectural configurations and targeted feature modeling strategies, rather than through explicit input-modality-based processing objective separation [17]. While MFGF-UNet implements enhancements in attention mechanisms and multi-scale feature modeling via GCT skip connections and the GMF-Inception module, its cross-modality fusion strategy fundamentally operates by organizing pre-extracted water indices and SAR polarimetric features into input-stage multi-channel representations. These are subsequently processed through a shared backbone network, thereby remaining fundamentally within the unified fusion paradigm characterized by channel-wise concatenation [21]. From the viewpoint of pre-trained knowledge utilization, multispectral/multimodal inputs and their purpose-built fusion architectures exhibit fundamental incompatibility with the conventional RGB-based pre-training paradigm established in natural image processing. This inherent discrepancy frequently prevents direct reuse of mature pre-trained representations in adapted model architectures. Prominent approaches like LKF-DCANet [13] and EU-Net [14] remain fundamentally grounded in optical image segmentation frameworks, directing primary research efforts toward spatial feature augmentation, contextual modeling, and network architecture lightweighting. Consequently, these methods fail to establish explicit spectral decoupling and fusion mechanisms specifically designed for multispectral bands. Specifically, while LKF-DCANet exclusively employs RGB optical imagery for experimental validation, EU-Net—despite being tested on high-resolution optical data containing NIR bands—does not methodologically emphasize multispectral information modeling in its architectural innovations. While the cross-sensor transfer learning framework developed by Li et al. [15] successfully addresses sample scarcity challenges in high-resolution small water body mapping, its input organization maintains a consolidated four-band RGB-NIR configuration lacking explicit spectral decoupling and spectral-interaction modeling mechanisms. This highlights a critical research gap: simultaneously preserving RGB pre-training benefits through explicit separation of visible/non-visible-band learning pathways, while enabling bi-directional adaptive feature interaction across modalities—both remaining fundamental challenges in multispectral remote sensing water segmentation.

To address this issue, this study proposes an Intelligent Gated Fusion Network (IGF-Net). Our dual-branch encoder processes synthetic RGB images (from selected RGB bands) in one branch for generic visual representation, while the complementary branch mines unique spectral patterns from non-visible bands. An adaptive gating mechanism dynamically modulates cross-branch feature interactions, prioritizing channel-wise discriminative cues through learnable weights. This study follows a complete workflow covering data processing, model construction, and systematic experimental validation. Preprocessing is first performed on the multispectral remote sensing image dataset. Next, a segmentation network centered on a Dual-Branch Feature Encoder Module and an Intelligent Gated Fusion Module (IGFM) is constructed. Finally, model performance is evaluated through comparative experiments against mainstream methods and ablation studies with replaced fusion modules. The remainder of this paper is organized as follows: Section 2 introduces the dataset characteristics and preprocessing protocol. Section 3 systematically elaborates

on the architecture and implementation details of the proposed intelligent gated fusion network. Section 4 presents the complete experimental setup and result analysis. Section 5 concludes the paper and provides an outlook for future work.

**Table 1.** Comparison of representative remote sensing water body identification methods.

| Representative Method                                                      | Input Type                      | Key Architecture                                                                                                                                      | Main Strengths                                                                                                                             |
|----------------------------------------------------------------------------|---------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|
| Weng et al.: SCR-Net [17]                                                  | RGB + NIR                       | ConvFormer branch + ResNet-50 branch + GAM attention module                                                                                           | Balances global context and local details for effective multispectral fusion and accurate water body segmentation.                         |
| Yang et al.: Multispectral and SAR Fusion algorithm [20]                   | Multispectral + SAR             | MASF + multi-scale segmentation + random forest                                                                                                       | Fuses complementary multispectral and SAR information to identify fragmented small water bodies in complex terrain.                        |
| Wang et al.: MFGF-UNet [21]                                                | SAR and the seven water indexes | U-Net + gated multi-filter inception module + GCT skip connection                                                                                     | Leverages multimodal and multiscale features for strong, robust, and low-complexity performance on the WIPI, Chengdu, and GF2020 datasets. |
| Liao et al.: LKF-DCANet [13]                                               | RGB                             | Channel attention-enhanced deformable convolution module + convolutional additive token mixer + learnable Kalman filter                               | Achieves precise boundary delineation and strong robustness to noise and appearance ambiguity with only 0.22 M parameters.                 |
| Cao et al.: EU-Net [14]                                                    | RGB + NIR                       | Improved residual connections + multi-scale dilated convolution module + multi-scale feature fusion module + channel and spatial attention mechanisms | Maintains water-body geometry and clear boundaries, especially in small water bodies, narrow channels, and complex scenes.                 |
| Li et al.: transfer learning framework from Sentinel-2 to PlanetScope [15] | RGB + NIR                       | Transfer learning framework from Sentinel-2 to PlanetScope + assessment of VMamba for small water-body mapping                                        | Reduces manual annotation effort and improves small water body mapping in cross-sensor transfer learning.                                  |

## 2. Dataset

To comprehensively evaluate the performance and generalization capability of the proposed model, this study employs two multispectral water body semantic segmentation datasets: (1) a custom-built Tiangong-2 remote sensing dataset with water body annotations, and (2) the publicly released Sentinel-2 Water Segmentation Dataset introduced by Yuan et al. [18]. Subsequent sections present detailed descriptions of both datasets.

### 2.1. Tiangong-2 Remote Sensing Image Water Body Semantic Segmentation Dataset

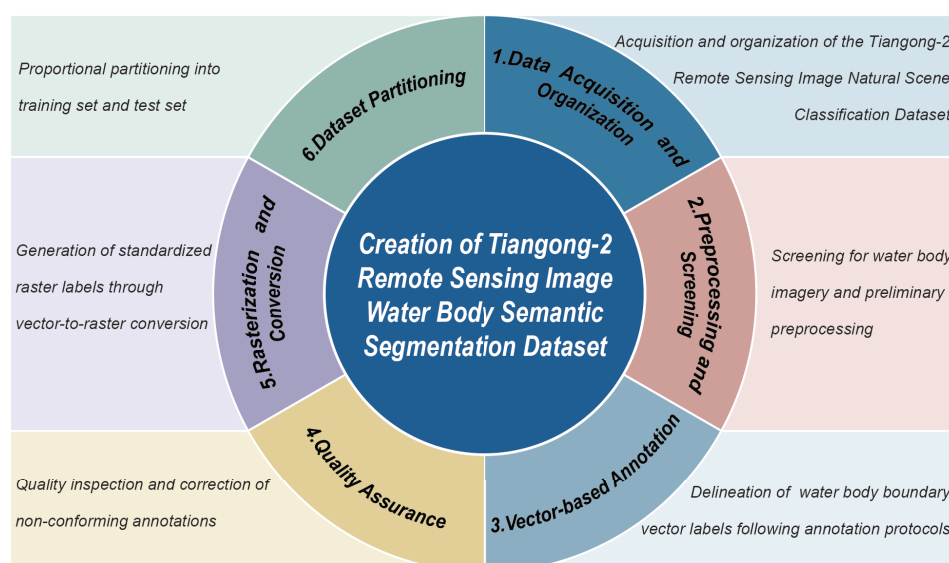
The multispectral data used in the Tiangong-2 Remote Sensing Image Water Body Semantic Segmentation Dataset originate from the “Tiangong-2 Remote Sensing Image Natural Scene Classification Dataset” [23]. This dataset was developed through systematic consideration of critical factors including geographic diversity and temporal coverage. It comprises two primary components: (1) multispectral imagery with 14 discrete spectral bands spanning 0.403–0.990  $\mu\text{m}$ , encompassing both visible and near-infrared regions, and (2) corresponding RGB true-color composites generated from specific bands (7, 11, and 12). The non-contiguous spectral sampling across this range is designed to enhance the differentiation of surface features, including water bodies. Complete specifications of individual band wavelengths are provided in Table 2.

**Table 2.** Spectral band specifications of the Tiangong-2 Remote Sensing Image Natural Scene Classification Dataset.

| Channel Number | Spectral Range ( $\mu\text{m}$ ) |
|----------------|----------------------------------|
| V1             | 0.970–0.990                      |
| V2             | 0.930–0.950                      |
| V3             | 0.895–0.915                      |
| V4             | 0.845–0.885                      |
| V5             | 0.810–0.830                      |
| V6             | 0.740–0.760                      |
| V7             | 0.6775–0.6875                    |
| V8             | 0.655–0.675                      |
| V9             | 0.610–0.630                      |
| V10            | 0.555–0.575                      |
| V11            | 0.510–0.530                      |
| V12            | 0.480–0.500                      |
| V13            | 0.433–0.453                      |
| V14            | 0.403–0.423                      |

Note: Band specifications were compiled from the dataset metadata of the Tiangong-2 Remote Sensing Image Natural Scene Classification Dataset (V1).

A total of 3776 images containing water bodies were selected from the aforementioned dataset. Based on this subset, we established the “Tiangong-2 Remote Sensing Image Water Body Semantic Segmentation Dataset” through manual visual interpretation. The annotation procedure employed the LabelMe tool to precisely delineate surface water bodies [24]. Annotated results were converted into binary raster labels (TIFF format) using a custom-developed Python (v3.10.18) script. In these label images, water body pixels are assigned a grayscale value of 255 (maximum intensity), while background pixels are assigned 0 (minimum intensity). To optimize data access efficiency, all original multispectral images in TIFF format were systematically transformed into NPY format—a specialized extension for NumPy that enables efficient loading of multi-dimensional arrays. After completing the aforementioned preprocessing, the dataset was randomly partitioned into a training set (3402 images) and a test set (374 images) at a 9:1 ratio. The complete data preparation pipeline is depicted in Figure 1.

**Figure 1.** Workflow of the Tiangong-2 Remote Sensing Image Water Body Semantic Segmentation Dataset construction.

## 2.2. Sentinel-2 Water Segmentation Dataset

To ensure fair comparison with existing representative methods and validate the model's generalizability on a public benchmark, this study adopts the publicly released Sentinel-2 Water Segmentation Dataset by Yuan et al. [18]. The dataset encompasses Chengdu City, Sichuan Province, China, along with its adjacent peri-urban transitional zones, utilizing Sentinel-2 satellite imagery. Notably, all training, validation, and quantitative testing in this study strictly employ the first temporal phase data (April 2018). While the original study primarily utilized the second (December 2018) and third (February 2019) phases as cross-temporal images to assess model robustness under varying illumination and weather conditions, we repurpose these phases exclusively for visual analysis in multi-temporal experiments.

To ensure spatial alignment among the multi-band data and compliance with network input specifications, the following preprocessing steps were applied:

- (1) **Spatial Resolution Unification:** Given the lower spatial resolution of SWIR bands (typically 20 m for Sentinel-2) compared to RGB/NIR bands (10 m), bilinear interpolation was applied to upsample SWIR data to 10 m, ensuring consistent spatial scale across all bands.
- (2) **Image Cropping and Tiling:** A grid-based cropping method was employed to divide the resolution-unified multi-band images into non-overlapping square patches of size  $256 \times 256$  pixels.
- (3) **Dataset Splitting:** In accordance with the 9:1 ratio established for the custom-built dataset, all valid image patches were randomly partitioned into training and test sets.

These preprocessing steps resulted in a curated collection of standardized image patches, ready for model input and subsequent analysis.

## 2.3. Comparison of the Two Datasets

Significant differences exist between the Tiangong-2 Remote Sensing Image Water Body Semantic Segmentation Dataset and the Sentinel-2 Water Segmentation Dataset regarding data sources, sensor platforms, spectral characteristics, and experimental purposes. Specifically, the former is a custom-built dataset constructed from Tiangong-2 multispectral imagery through manual annotation, containing 14 spectral bands (detailed in Table 2), primarily used to evaluate the proposed method's effectiveness on the custom-built multispectral data. The latter is a publicly available benchmark dataset derived from Sentinel-2 imagery, encompassing visible, near-infrared, and shortwave infrared bands, mainly employed to validate the proposed model's performance on a standardized public benchmark as well as its cross-temporal generalization capability. A detailed comparison of the two datasets is presented in Table 3.

**Table 3.** Comparison between the Tiangong-2 Remote Sensing Image Water Body Semantic Segmentation Dataset and the Sentinel-2 Water Segmentation Dataset.

| Aspect                   | Tiangong-2 Remote Sensing Image Water Body Semantic Segmentation Dataset                    | Sentinel-2 Water Segmentation Dataset                     |
|--------------------------|---------------------------------------------------------------------------------------------|-----------------------------------------------------------|
| Data source              | Custom-built from the Tiangong-2 Remote Sensing Image Natural Scene Classification Dataset. | Public dataset released by Yuan et al.                    |
| Spectral characteristics | 14 discrete spectral bands spanning 0.403–0.990 $\mu\text{m}$ .                             | Multispectral imagery including RGB, NIR, and SWIR bands. |
| Annotation strategy      | Manually annotated in this study using LabelMe.                                             | Labels provided in the original public benchmark dataset. |



Table 3. Cont.

| Aspect                         | Tiangong-2 Remote Sensing Image Water Body Semantic Segmentation Dataset         | Sentinel-2 Water Segmentation Dataset                                                                                                               |
|--------------------------------|----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| Data preparation               | TIFF images converted to NPY format; binary labels generated by a custom script. | SWIR bands upsampled from 20 m to 10 m and cropped into $256 \times 256$ patches.                                                                   |
| Temporal setting in this study | Used for supervised training and quantitative evaluation.                        | April 2018 data used for training, validation, and quantitative testing; December 2018 and February 2019 data used for qualitative visual analysis. |
| Primary role in this study     | Task-specific evaluation on a custom-built multispectral dataset.                | Benchmark evaluation and cross-temporal generalization analysis on a public dataset.                                                                |

For clarity, the Tiangong-2 Remote Sensing Image Water Body Semantic Segmentation Dataset is abbreviated as TG2-WaterSeg, and the Sentinel-2 Water Segmentation Dataset is abbreviated as S2-WaterSeg in the remainder of this paper.

### 3. Methodology

#### 3.1. Intelligent Gated Fusion Network Architecture

To tackle the challenges of water body segmentation in multispectral remote sensing imagery, this study proposes an end-to-end IGF-Net. The network adopts an encoder-decoder architecture [25], taking multispectral images as input and generating corresponding binary segmentation masks for water bodies as output. The overall structure of the network is illustrated in Figure 2.

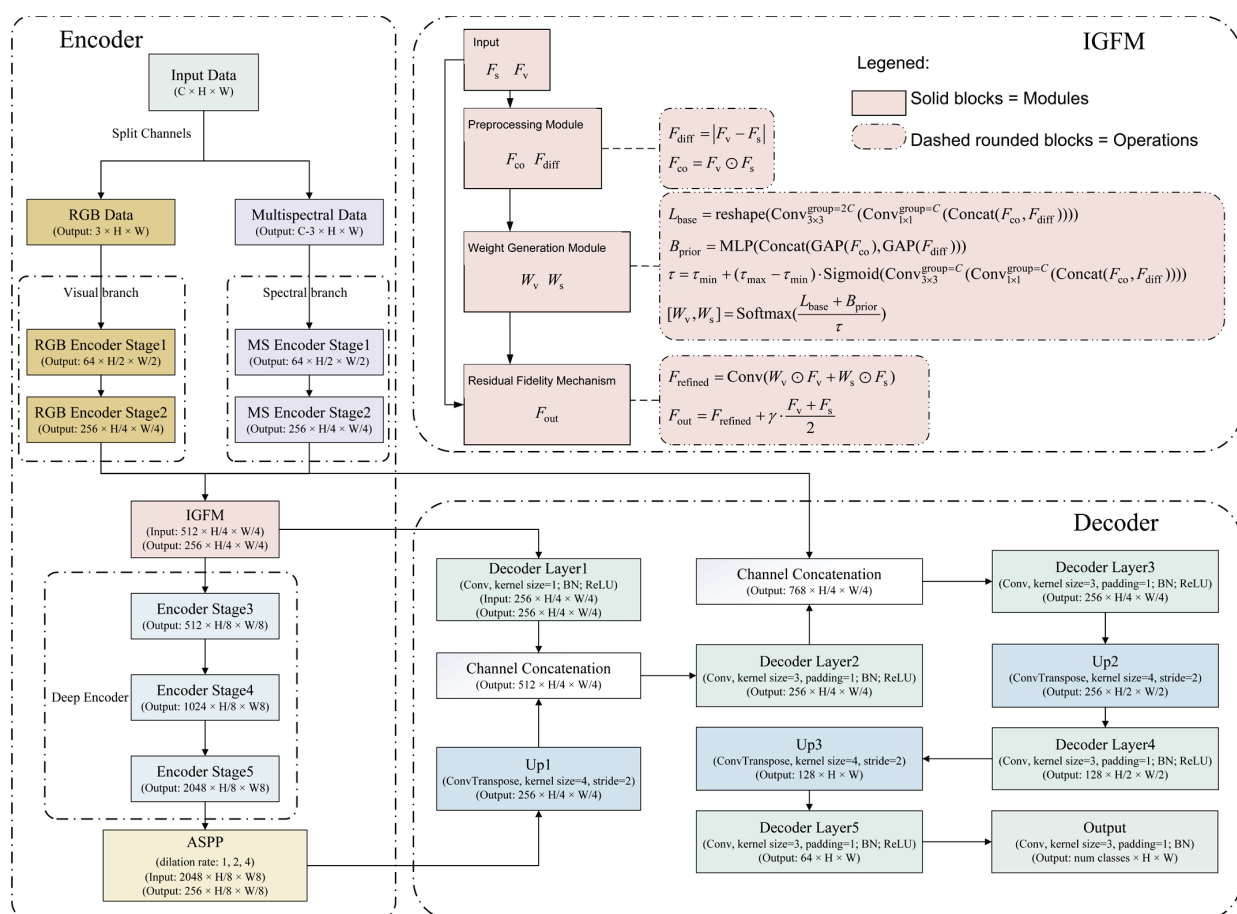


Figure 2. Structure diagram of Intelligent Gated Fusion Network.

As the feature extraction core of the model, the encoder is comprised mainly of the following three modules:

- (1) **Dual-Branch Feature Encoder Module:** This module employs a parallel dual-branch structure to process distinct band subsets of the multispectral data. The visual branch takes the red, green, and blue bands as input and loads weights pre-trained on ImageNet to extract general visual features. The spectral branch processes the remaining bands with randomly initialized weights, specializing in learning spectral-specific features from the multispectral data. Both branches adopt the ResNet-50 backbone [26], producing feature maps at  $1/4$  of the input spatial resolution with 256 channels.
- (2) **Intelligent Gated Fusion Module:** As the core innovation of this study, this module receives outputs from the dual-branch feature encoder. Through a gating-based adaptive mechanism, it dynamically learns branch-wise attention weights via trainable parameters, enabling selective fusion of visual and spectral features to enhance the model's focus on key discriminative characteristics.
- (3) **Atrous Spatial Pyramid Pooling (ASPP) Module [27]:** Following further processing by the deep encoder, the fused features are passed to the ASPP module. This module aggregates multi-scale contextual information through parallel atrous convolutions with varying dilation rates, thereby strengthening the model's capacity to capture water bodies across diverse spatial scales.

The decoder employs a progressive upsampling scheme with multi-level feature fusion. It first processes the high-level semantic features from the encoder, reducing their dimensionality via a  $1 \times 1$  convolution and then upsampling them to a quarter of the input resolution. These upsampled features are then fused—via skip connections—with shallow features from the Intelligent Gated Fusion Module, the visual branch, and the spectral branch, integrating complementary spatial details. Finally, through two consecutive upsampling stages (each doubling the resolution), each refined by  $3 \times 3$  convolutions, the feature maps are restored to the original input resolution to yield high-precision segmentation predictions.

To clearly present the technical details of the model and ensure the reproducibility of the research, the detailed parameter settings and inter-layer connections of the network are provided in Table 4.

**Table 4.** Description of the Intelligent Gated Fusion Network architecture.

| Stage/Module |                         | Input Size                   | Output Size                                         | Details                                                                                              |
|--------------|-------------------------|------------------------------|-----------------------------------------------------|------------------------------------------------------------------------------------------------------|
| Input Data   |                         | $C \times H \times W$        | $(C-3) \times H \times W$ and $3 \times H \times W$ | Split the input into RGB and multispectral data by channel indices.                                  |
| Encoder      | Visual branch Encoder   | $3 \times H \times W$        | $256 \times H/4 \times W/4$                         | Extract RGB features using a shallow ResNet50 encoder with pretrained initialization.                |
|              | Spectral branch Encoder | $(C-3) \times H \times W$    | $256 \times H/4 \times W/4$                         | Extract multispectral features using a shallow ResNet50 encoder with random initialization.          |
|              | IGFM                    | $512 \times H/4 \times W/4$  | $256 \times H/4 \times W/4$                         | Adaptively fuse dual-branch features with dynamic weights while preserving residual information.     |
|              | Deep Encoder            | $256 \times H/4 \times W/4$  | $2048 \times H/8 \times W/8$                        | Further extract high-level semantic features using the deep ResNet50 encoder.                        |
|              | ASPP                    | $2048 \times H/8 \times W/8$ | $256 \times H/8 \times W/8$                         | Extract multi-scale contextual features through parallel convolutions with different dilation rates. |

Table 4. Cont.

| Stage/Module | Input Size                                                                                                                                                                        | Output Size                     | Details                                                                                                                          |
|--------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| Decoder      | ASPP: $256 \times H/8 \times W/8$ ;<br>IGFM: $256 \times H/4 \times W/4$ ;<br>RGB Encoder Stage2: $256 \times H/4 \times W/4$ ;<br>MS Encoder Stage2: $256 \times H/4 \times W/4$ | Num classes $\times H \times W$ | Generate the final prediction result via transposed convolution upsampling, channel concatenation, and convolutional refinement. |

### 3.2. Dual-Branch Feature Encoder Module

Transfer learning, by leveraging parameters from models pre-trained on large-scale datasets (e.g., ImageNet), significantly enhances both model performance and training efficiency for target tasks. However, applying this methodology to multispectral remote sensing image segmentation encounters a critical issue: prevalent pre-trained architectures (such as ResNet-50) are inherently designed for three-channel color images, with their input dimensions strictly confined to three channels. When processing multispectral data containing more than three spectral bands, this channel dimension discrepancy precludes direct initialization with pre-trained weights, thus diminishing the efficacy of transfer learning.

To address the aforementioned challenge, our model employs a dual-branch feature encoder module. This architecture comprises two distinct branches built upon the ResNet-50 backbone: a visual branch and a spectral branch. The visual branch processes the three specific bands (red, green, blue) from the input multispectral data, initializing with weights pre-trained on ImageNet. This design enables direct transfer of generic visual features (such as edges and textures) acquired from large-scale datasets, endowing the model with robust spatial semantic priors. Conversely, the spectral branch handles all residual spectral bands (excluding RGB channels), maintaining identical architectural configuration to the visual branch but utilizing randomly initialized weights. Specialized for multispectral analysis, this branch captures minute reflectance variations among different land cover categories across extensive spectral ranges.

The design of this dual-branch feature encoder module is founded on a principle of explicit complementarity. The visual branch inherits the robust generic feature representation from pre-trained models, while the spectral branch confers specialized learning capacity for multispectral signatures. This separate encoding strategy simultaneously harnesses the benefits of transfer learning and avoids the structural conflicts or information loss associated with forced input dimensionality alterations. Consequently, the module provides dual feature representations for downstream fusion, incorporating both general visual knowledge and domain-specific spectral information.

### 3.3. Intelligent Gated Fusion Module

The fusion quality of features extracted by the dual-branch feature encoder module critically determines the final segmentation performance. Since the visual branch and the spectral branch are derived from heterogeneous inputs, their encoded features usually contain both modality-shared semantic structures and modality-specific responses. Therefore, effective fusion should not simply aggregate the two feature maps, but should adaptively determine how much each branch contributes according to their consistency and complementarity.

Existing fusion strategies, including direct concatenation, element-wise summation, and conventional attention-based reweighting, are insufficient for this purpose. Direct fusion operations ignore the interaction patterns between heterogeneous branches, while standard attention mechanisms mainly estimate saliency over channels or spatial locations within a feature representation. Although such methods can enhance informative responses,

they do not explicitly characterize whether the two branches are mutually consistent, complementary, or conflicting. This limitation is particularly critical in RGB-spectral fusion, where the spectral diversity and dimensional heterogeneity may lead to substantial inter-branch discrepancy.

To address this issue, this study introduces an Intelligent Gated Fusion Module (IGFM) for adaptive heterogeneous feature integration. Different from generic attention or conventional cross-modal fusion methods, IGFM formulates fusion from the perspective of feature complementarity modeling rather than simple response enhancement. Specifically, it explicitly decomposes the interaction between the two branch features into a disparity cue and a coherence cue, which respectively describe modality-specific differences and modality-consistent activations. Based on these two complementary descriptors, the module generates branch-wise adaptive fusion weights to regulate the contribution of each feature source. In this way, IGFM performs complementarity-aware gated fusion, enabling the network to preserve discriminative complementary information while suppressing redundant or inconsistent responses. Moreover, a residual fidelity mechanism is introduced to preserve original information completeness and improve training stability. As illustrated in Figure 2, IGFM consists of three components: a preprocessing module, a weight generation module, and a residual fidelity mechanism.

### 3.3.1. Preprocessing Module

The preprocessing module is designed to explicitly model the interaction relationship between the visual branch feature ( $F_v$ ) and the spectral branch feature ( $F_s$ ). Instead of directly using the raw concatenated features for fusion, we decompose their interaction into two complementary aspects: disparity and coherence. The former reflects branch-specific responses that may provide complementary cues, while the latter captures co-activated structures consistently supported by both branches.

Accordingly, the preprocessing module contains two parallel pathways: a disparity extraction stream and a coherence extraction stream. The disparity features are obtained by computing the element-wise absolute difference between the two input features:

$$F_{\text{diff}} = |F_v - F_s| \quad (1)$$

This operation highlights spatial locations and channel responses where the two branches exhibit evident divergence, thereby emphasizing modality-specific information that may be useful for downstream segmentation.

The coherence features are defined as:

$$F_{\text{co}} = F_v \odot F_s \quad (2)$$

where  $\odot$  denotes element-wise multiplication. This operation enhances regions where both branches are simultaneously activated, thereby preserving shared semantic structures and mutually corroborated evidence.

By jointly extracting  $F_{\text{diff}}$  and  $F_{\text{co}}$ , the preprocessing module provides an explicit and interpretable representation of feature complementarity. Compared with directly using concatenated features alone, this design offers a more informative basis for subsequent weight generation, since it simultaneously encodes branch discrepancy and branch agreement.

### 3.3.2. Weight Generation Module

The weight generation module learns spatially and channel-wise adaptive fusion weights from the interaction cues produced by the preprocessing module. Unlike conven-

tional attention mechanisms that directly estimate saliency scores from a single feature tensor, the proposed module predicts branch-wise contribution coefficients from the jointly encoded disparity and coherence features. Therefore, the learned weights are explicitly associated with inter-branch complementarity rather than merely local activation strength.

Specifically,  $F_{\text{diff}}$  and  $F_{\text{co}}$  are first concatenated along the channel dimension. The concatenated features are then processed sequentially by a  $1 \times 1$  group convolution (with group = C) and a  $3 \times 3$  group convolution (with group = 2C), denoted as  $\text{Conv}_{3 \times 3}^{\text{group}=2C}$  and  $\text{Conv}_{1 \times 1}^{\text{group}=C}$  respectively. Finally, a reshape operation is applied to obtain the base weight tensor  $L_{\text{base}}$ . This process is formulated as:

$$L_{\text{base}} = \text{reshape}(\text{Conv}_{3 \times 3}^{\text{group}=2C}(\text{Conv}_{1 \times 1}^{\text{group}=C}(\text{Concat}(F_{\text{co}}, F_{\text{diff}})))) \quad (3)$$

Here, the  $1 \times 1$  group convolution performs local channel interaction and lightweight feature projection, while the subsequent  $3 \times 3$  group convolution introduces spatial contextual modeling. Together, they allow the module to infer fusion preferences from both channel-level and spatial-level complementarity patterns.

In parallel, a dedicated pathway predicts a channel-wise and spatially adaptive temperature parameter  $\tau$ . This branch first transforms the concatenated features  $\text{Concat}(F_{\text{co}}, F_{\text{diff}})$  via a  $1 \times 1$  group convolution (with group = C), producing an initial value  $\tau_{\text{init}}$ . This is followed by a  $3 \times 3$  group convolution (with group = C) and a Sigmoid activation, yielding a normalized intermediate value  $\tau_{\text{process}} \in [0, 1]$ . Finally,  $\tau_{\text{process}}$  is linearly mapped to the predefined interval  $[\tau_{\text{min}}, \tau_{\text{max}}]$  to dynamically adjust the distribution sharpness of the Softmax operation. The complete formulation is:

$$\begin{aligned} \tau_{\text{init}} &= \text{Conv}_{1 \times 1}^{\text{group}=C}(\text{Concat}(F_{\text{co}}, F_{\text{diff}})) \\ \tau_{\text{process}} &= \text{Sigmoid}(\text{Conv}_{3 \times 3}^{\text{group}=C}(\tau_{\text{init}})) \\ \tau &= \tau_{\text{min}} + (\tau_{\text{max}} - \tau_{\text{min}}) \cdot \tau_{\text{process}} \end{aligned} \quad (4)$$

The adaptive temperature plays an important role in controlling the competition-cooperation relationship between the two branches. When one branch exhibits significantly stronger evidence than the other, a relatively lower effective temperature leads to a sharper distribution and more decisive branch selection. In contrast, when both branches provide useful complementary cues, a smoother distribution encourages cooperative fusion rather than excessive suppression.

To further introduce global semantic guidance, the module additionally incorporates a channel-wise contextual prior mechanism. Specifically, global average pooling is separately applied to the disparity feature  $F_{\text{diff}}$  and the coherence feature  $F_{\text{co}}$ , and the resulting channel descriptors are concatenated and fed into a lightweight multi-layer perceptron (MLP) to generate a channel prior bias:

$$B_{\text{prior}} = \text{MLP}(\text{Concat}(\text{GAP}(F_{\text{co}}), \text{GAP}(F_{\text{diff}}))) \quad (5)$$

This prior branch captures global complementarity statistics beyond local convolutions, thereby providing a coarse but stable semantic bias for weight estimation.

Finally, the branch-wise fusion weights are derived by combining the base weights and the channel prior bias, followed by temperature-scaled Softmax normalization:

$$[W_v, W_s] = \text{Softmax}\left(\frac{L_{\text{base}} + B_{\text{prior}}}{\tau}\right) \quad (6)$$



where  $W_v$  and  $W_s$  denote the adaptive fusion weights for the visual branch and the spectral branch, respectively. Through this process, the module converts disparity-aware and coherence-aware interaction patterns into branch contribution coefficients, thereby achieving fine-grained and interpretable heterogeneous feature fusion.

### 3.3.3. Residual Fidelity Mechanism

Although adaptive gating can effectively regulate the contribution of the two branches, aggressive weighting may also lead to the suppression of potentially useful information. To alleviate this problem, we introduce a residual fidelity mechanism to preserve information completeness and stabilize optimization.

After obtaining the fusion weights, a weighted summation is first performed to produce the preliminary fused feature:

$$F_{\text{fused}} = W_v \odot F_v + W_s \odot F_s \quad (7)$$

The fused feature is then refined by a lightweight convolutional layer:

$$F_{\text{refined}} = \text{Conv}(F_{\text{fused}}) \quad (8)$$

To further avoid information loss and facilitate gradient propagation, a learnable residual connection is introduced. Specifically, the mean of the original input features is added back to the refined output, scaled by a learnable coefficient  $\gamma$ :

$$F_{\text{out}} = F_{\text{refined}} + \gamma \cdot \frac{F_v + F_s}{2} \quad (9)$$

where  $\gamma$  is a learnable scalar parameter initialized to 0.

This design ensures that the proposed fusion process remains selective rather than destructive. Even when the learned gating weights favor one branch, the residual pathway still preserves shared low-frequency structural information from the original inputs, thus preventing excessive information loss caused by overconfident branch selection. Meanwhile, the learnable residual coefficient allows the network to adaptively determine the necessity of residual compensation during training, which further improves optimization stability and final segmentation performance.

## 4. Experiments

### 4.1. Experimental Setup

#### 4.1.1. Implementation Details

The experimental framework comprised two stages: model training conducted on a cloud server and inference testing performed locally, to optimize computational efficiency and deployment practicality.

**Cloud-Based Training Configuration:** All training procedures were executed on the VirtAI 1.0 cloud platform utilizing a B1. Large instance. This instance featured 32 GB RAM and an GPU with 24 GB dedicated VRAM. The software environment incorporated Ubuntu 22.04 LTS, Python 3.10, PyTorch 2.0.1 framework, and CUDA 11.8 acceleration toolkit.

**Local Inference Setup:** Inference validation was conducted on a local workstation equipped with a 12th Generation Intel Core i7-12650H processor and an NVIDIA GeForce RTX 4060 mobile GPU, sourced from HASEE Computer Co., Ltd., Shenzhen, China. The system configuration maintained Windows 11 OS, Python 3.10, PyTorch 2.7.1, and CUDA 12.8. The hardware-software ecosystem remained unmodified to emulate authentic deployment conditions.

Model training employed the AdamW optimizer [28] with an initial learning rate of 0.0005. A cosine annealing scheduler [29] was used for adaptive learning rate adjustment, and a weight decay coefficient of 0.0001 was applied for regularization. Training proceeded for 100 epochs, optimizing the binary cross-entropy loss.

To enhance the model's generalization capability and robustness, data augmentation techniques were applied during preprocessing. These included random rotation, vertical/horizontal flipping, and scaling. In the evaluation phase, model-generated probability maps were binarized at a fixed threshold of 0.5 to produce segmentation masks, thereby facilitating quantitative performance assessment through direct comparison with ground truth annotations.

#### 4.1.2. Evaluation Metrics

To systematically assess the model's performance, four quantitative evaluation metrics were adopted: Recall, Precision, Dice Coefficient (Dice), and Intersection over Union (IoU). The corresponding mathematical formulations for these indicators are defined in Equation (10):

$$\begin{aligned}\text{Recall} &= \frac{TP}{TP+FN} \\ \text{Precision} &= \frac{TP}{TP+FP} \\ \text{Dice} &= 2 \times \frac{|A \cap B|}{|A| + |B|} \\ \text{IoU} &= \frac{|A \cap B|}{|A \cup B|}\end{aligned}\quad (10)$$

In the equations, TP (True Positive) represents the number of positive samples correctly identified by the model; FP (False Positive) denotes the number of samples incorrectly predicted as positive; FN (False Negative) indicates the number of samples incorrectly predicted as negative. Here,  $A$  denotes the ground truth segmentation label,  $B$  is the predicted segmentation result;  $A \cap B$  represents the overlapping area between  $A$  and  $B$ , while  $A \cup B$  indicates the area of their union. Recall focuses on the completeness of the model's coverage of target regions, emphasizing whether true targets are missed. Precision, on the other hand, highlights the accuracy of the predicted regions, with an emphasis on controlling false positives. Both the Dice and IoU provide a holistic assessment from the perspective of regional overlap. Among them, the Dice offers better robustness to class imbalance, whereas IoU provides an intuitive measure of geometric overlap and is widely adopted as a standard metric in segmentation tasks. All metrics range from 0 to 1, with values closer to 1 indicating more accurate segmentation results.

#### 4.1.3. Model Complexity and Inference Efficiency

Beyond segmentation accuracy, the practical applicability of remote sensing models hinges critically on computational complexity and inference efficiency. To conduct a thorough evaluation of the proposed IGF-Net, this study performs comparative analyses across multiple models using four key metrics: parameter count (Params), floating-point operations (FLOPs), average inference time per image (Avg-Time), and frames per second (FPS). These metrics collectively characterize distinct performance dimensions—Params quantifies model size and memory requirements, FLOPs measures computational complexity during forward propagation, while Avg-Time and FPS jointly assess inference latency and processing throughput.

All experiments were conducted under strictly controlled conditions to ensure fair benchmarking. Prior to speed measurements, each model underwent multiple warm-up iterations to stabilize system performance and mitigate initialization artifacts. Subsequently, repeated forward propagations were executed, with statistical averaging applied to derive

reliable estimates of per-image inference time. This dual-perspective assessment framework, integrating both accuracy metrics and computational costs, enables comprehensive methodological comparisons. Detailed quantitative results will be systematically presented in the subsequent experimental section.

## 4.2. Comparative Experiments

### 4.2.1. Comparative Experimental Setup

To systematically validate the effectiveness and generalization capability of IGF-Net in water body identification tasks, this paper designs comprehensive experiments encompassing three aspects: model comparison, input modality analysis, and cross-temporal generalization verification.

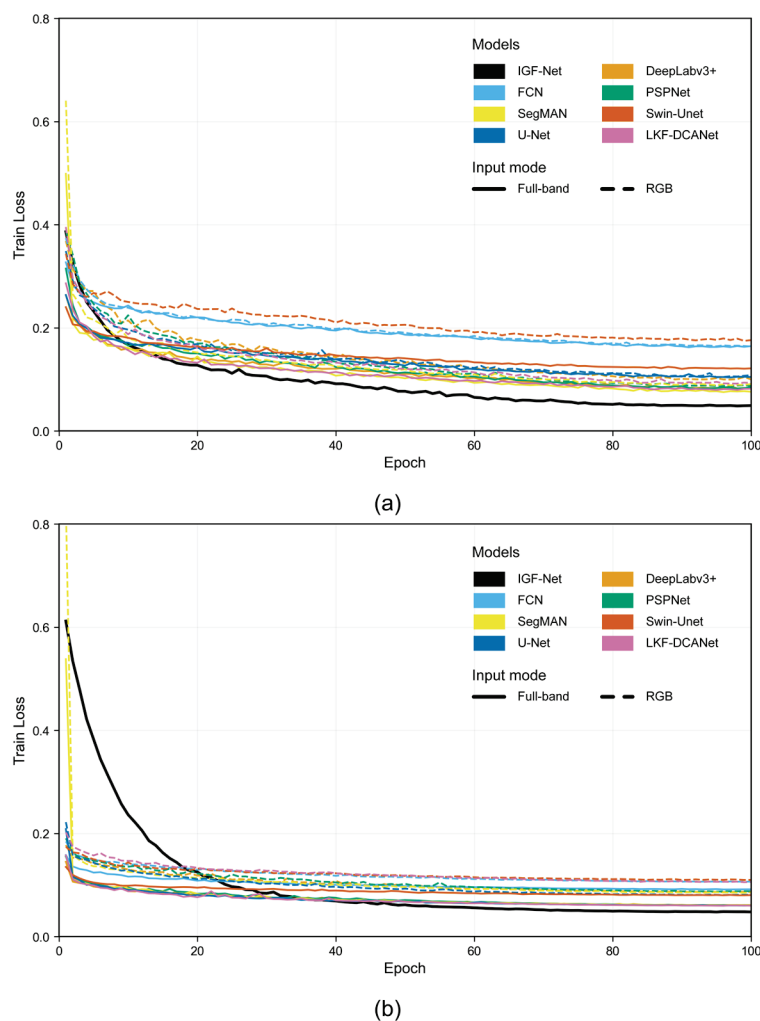
In terms of comparative model selection, seven representative semantic segmentation models—FCN [30], PSPNet [31], U-Net [32], DeepLabv3+ [27], LKF-DCANet [13], SegMAN [33], and Swin-Unet [34]—were chosen as baseline models. It should be noted that to ensure fairness in experimental comparison, this study did not directly adopt the official training configurations or published results from the original papers of these models. Instead, their core structures were uniformly adapted to the same training and testing framework. All models were retrained and evaluated under consistent conditions, including identical data partitioning, input modalities, optimization strategies, and hyperparameter settings. Therefore, the comparative models described here (e.g., FCN, DeepLabv3+, and LKF-DCANet) are unified framework implementations based on the core ideas of the original methods. Their results reflect the relative performance of different architectures under a unified experimental protocol, rather than strict reproductions of official versions or original literature results.

In terms of experimental data, this study employs two multispectral remote sensing water body segmentation datasets: the self-constructed Tiangong-2 Remote Sensing Image Water Body Semantic Segmentation Dataset (TG2-WaterSeg) and the publicly available Sentinel-2 Water Segmentation Dataset (S2-WaterSeg). The TG2-WaterSeg contains 3776 images covering 14 spectral bands, while the S2-WaterSeg includes 5 spectral bands. These two datasets differ in imaging conditions, spectral composition, and land cover scenarios, effectively enabling validation of model robustness and adaptability. To systematically evaluate the impact of spectral information on segmentation performance, two input modalities were designed for each dataset: RGB mode and Full-band mode. Specifically, RGB mode uses only the visible red, green, and blue bands as input, whereas Full-band mode incorporates all available spectral bands from each dataset. This configuration allows for a unified analysis of how limited visible light information versus multispectral information affects water extraction outcomes.

For the validation of generalization capability, this study conducted multi-dataset independent validation and cross-temporal transfer analysis, considering data availability. For TG2-WaterSeg, strictly multi-temporal experiments could not be performed because its source dataset (Tiangong-2 Remote Sensing Image Natural Scene Classification Dataset) does not provide observational images of the same area at other times. For S2-WaterSeg, in addition to the original temporal data, the source dataset includes images from two additional time phases for the same region. However, as these new phases lack accurate and reliable pixel-level annotations, this study did not conduct strict quantitative evaluations on them. Instead, the model trained on the original temporal data was directly transferred to the two unseen temporal phases (not involved in training) for inference. Representative samples from all three phases were selected for visual comparison to qualitatively assess the model's transferability, robustness, and prediction stability under temporal variations.

#### 4.2.2. Quantitative Comparison Results

Figure 3 illustrates the training loss curves of the IGF-Net model and its comparative models. From an overall trend perspective, the training loss of all models decreased significantly as the number of training epochs increased, gradually converging to a stable state after approximately 70 epochs, indicating an effective and stable training process. Furthermore, by comparing the loss curves under the two input modes, it can be observed that models trained with the Full-band mode generally exhibited lower loss values in most cases compared to their counterparts using only the RGB mode, and the converged curves showed smaller fluctuations, suggesting a more stable optimization process.



**Figure 3.** Training loss convergence curves of different models on (a) TG2-WaterSeg and (b) S2-WaterSeg.

Tables 5 and 6 present the experimental results of IGF-Net and various comparison methods on the TG2-WaterSeg and S2-WaterSeg datasets, respectively, evaluating both model complexity and segmentation performance. It should be noted that due to differences in input sizes between the two datasets and variations in the number of input channels between RGB mode and Full-band mode, the Params and FLOPs of the same model may vary across datasets or input modes. Furthermore, to objectively reflect the model's performance on the entire test dataset, all reported metrics are averaged over individual test results on every sample in the corresponding test sets.

As shown in Table 5, all models achieve high segmentation accuracy on the TG2-WaterSeg dataset. On this basis, the proposed IGF-Net achieves the best performance, with Dice, Precision, Recall, and IoU reaching 0.9239, 0.9331, 0.9235, and 0.8742, respectively. Compared to the strong baseline SegMAN, IGF-Net improves Dice and IoU by approximately 1.01% and 1.37%, respectively; compared to LKF-DCANet, the improvements are 1.17% and 1.57%. This indicates that even under conditions of high overall accuracy, the proposed method further enhances the overlap between water regions and ground truth, demonstrating superior region modeling and boundary delineation capabilities.

**Table 5.** Quantitative comparison of different models on TG2-WaterSeg.

|                                              | Model                 | Params (M) | FLOPs (G) | FPS    | Avg-Time (ms) | Dice          | Precision     | Recall        | IoU           |
|----------------------------------------------|-----------------------|------------|-----------|--------|---------------|---------------|---------------|---------------|---------------|
| RGB mode<br>$3 \times 128 \times 128$        | DeepLabv3+            | 40.35      | 8.68      | 207.95 | 4.81          | 0.8936        | 0.9103        | 0.8934        | 0.8335        |
|                                              | FCN                   | 9.41       | 7.34      | 807.11 | 1.24          | 0.6668        | 0.7370        | 0.6493        | 0.6086        |
|                                              | PSPNet                | 48.94      | 23.51     | 202.25 | 4.94          | 0.8746        | 0.8928        | 0.8727        | 0.8094        |
|                                              | U-Net                 | 31.04      | 27.37     | 258.05 | 3.88          | 0.8927        | 0.9204        | 0.8865        | 0.8324        |
|                                              | Swin-Unet             | 27.15      | 3.85      | 107.05 | 9.34          | 0.8085        | 0.8765        | 0.7903        | 0.7415        |
|                                              | LKF-DCANet            | 4.30       | 10.15     | 334.81 | 2.99          | 0.8987        | 0.9056        | 0.9056        | 0.8414        |
|                                              | SegMAN                | 26.25      | 15.75     | 15.38  | 65.00         | 0.9021        | 0.9212        | 0.8982        | 0.8445        |
| Full-band mode<br>$14 \times 128 \times 128$ | DeepLabv3+            | 40.38      | 8.96      | 202.52 | 4.94          | 0.9096        | 0.9208        | 0.9109        | 0.8542        |
|                                              | FCN                   | 9.41       | 7.54      | 806.30 | 1.24          | 0.6801        | 0.7304        | 0.6691        | 0.6213        |
|                                              | PSPNet                | 48.98      | 23.79     | 203.62 | 4.91          | 0.8891        | 0.8991        | 0.8928        | 0.8281        |
|                                              | U-Net                 | 31.04      | 27.58     | 256.77 | 3.89          | 0.8941        | 0.9180        | 0.8907        | 0.8388        |
|                                              | Swin-Unet             | 27.17      | 3.89      | 102.23 | 9.78          | 0.8834        | 0.9123        | 0.8752        | 0.8237        |
|                                              | LKF-DCANet            | 4.31       | 10.25     | 324.26 | 3.08          | 0.9122        | 0.9223        | 0.9125        | 0.8585        |
|                                              | SegMAN                | 26.25      | 15.78     | 15.10  | 66.22         | 0.9138        | 0.9264        | 0.9123        | 0.8605        |
|                                              | <b>IGF-Net (ours)</b> | 45.15      | 51.48     | 97.92  | 10.21         | <b>0.9239</b> | <b>0.9331</b> | <b>0.9235</b> | <b>0.8742</b> |

**Table 6.** Quantitative comparison of different models on S2-WaterSeg.

|                                             | Model                 | Params (M) | FLOPs (G) | FPS    | Avg-Time (ms) | Dice          | Precision     | Recall        | IoU           |
|---------------------------------------------|-----------------------|------------|-----------|--------|---------------|---------------|---------------|---------------|---------------|
| RGB mode<br>$3 \times 256 \times 256$       | DeepLabv3+            | 40.35      | 34.73     | 160.03 | 6.25          | 0.4464        | 0.6188        | 0.4031        | 0.3403        |
|                                             | FCN                   | 9.41       | 29.35     | 296.13 | 3.38          | 0.1930        | 0.3492        | 0.1641        | 0.1368        |
|                                             | PSPNet                | 48.94      | 93.71     | 90.99  | 10.99         | 0.4230        | 0.5849        | 0.3782        | 0.3188        |
|                                             | U-Net                 | 31.04      | 109.48    | 89.53  | 11.17         | 0.4800        | 0.6567        | 0.4320        | 0.3660        |
|                                             | Swin-Unet             | 27.15      | 15.41     | 73.99  | 13.51         | 0.2966        | 0.5391        | 0.2456        | 0.2172        |
|                                             | LKF-DCANet            | 4.30       | 40.60     | 115.37 | 8.67          | 0.3079        | 0.5509        | 0.2578        | 0.2270        |
|                                             | SegMAN                | 26.25      | 61.78     | 9.15   | 109.24        | 0.4411        | 0.6316        | 0.3903        | 0.3327        |
| Full-band mode<br>$5 \times 256 \times 256$ | DeepLabv3+            | 40.35      | 34.93     | 160.05 | 6.25          | 0.6152        | 0.7225        | 0.5798        | 0.4785        |
|                                             | FCN                   | 9.41       | 29.50     | 300.05 | 3.33          | 0.2484        | 0.4249        | 0.2141        | 0.1735        |
|                                             | PSPNet                | 48.95      | 93.92     | 90.17  | 11.09         | 0.5937        | 0.7063        | 0.5616        | 0.4577        |
|                                             | U-Net                 | 31.04      | 109.63    | 90.69  | 11.03         | 0.6272        | 0.7395        | 0.5877        | 0.4909        |
|                                             | Swin-Unet             | 27.15      | 15.44     | 78.22  | 12.78         | 0.5192        | 0.6977        | 0.4683        | 0.3916        |
|                                             | LKF-DCANet            | 4.30       | 40.68     | 115.78 | 8.64          | 0.6346        | <b>0.7343</b> | 0.5949        | 0.4969        |
|                                             | SegMAN                | 26.25      | 61.80     | 9.11   | 109.81        | 0.6236        | 0.7123        | 0.5949        | 0.4855        |
|                                             | <b>IGF-Net (ours)</b> | 45.12      | 205.00    | 55.87  | 17.90         | <b>0.6370</b> | 0.7090        | <b>0.6198</b> | <b>0.5009</b> |

A comparison of the two input modes reveals that most methods perform better in Full-band mode than in RGB mode. For instance, the IoU of DeepLabv3+ increases from 0.8335 (RGB) to 0.8542 (Full-band), U-Net from 0.8324 to 0.8388, and SegMAN from 0.8445 to 0.8605. This suggests that incorporating multispectral bands provides richer spectral discriminative information for the models. In terms of model categories, traditional



convolutional networks (e.g., FCN, PSPNet, and U-Net) achieve reasonable results but generally lag behind advanced architectures like LKF-DCANet and SegMAN. This implies that relying solely on local convolutional features or conventional multi-scale aggregation strategies is insufficient to capture fine-grained differences in complex water regions. IGF-Net further outperforms these strong baselines, demonstrating its superior capability in spectral information fusion and water structure representation.

Table 6 presents the experimental results of various models on the S2-WaterSeg. Compared to the results on TG2-WaterSeg (Table 5), all methods exhibit varying degrees of performance decline on S2-WaterSeg. This is attributed to the higher segmentation difficulty of the S2-WaterSeg, characterized by more complex water boundaries, stronger background interference, and more pronounced inter-class confusion. In this more challenging scenario, IGF-Net achieves the highest Dice, Recall, and IoU scores of 0.6370, 0.6198, and 0.5009, respectively, demonstrating its robust water recognition capability in complex remote sensing environments. While LKF-DCANet attains the highest Precision (0.7343), slightly exceeding IGF-Net's 0.7090, its Recall and IoU remain lower. This indicates that LKF-DCANet's predictions are more conservative, reducing false positives but increasing false negatives. By contrast, IGF-Net achieves a superior balance between Precision and Recall, resulting in better overall performance on metrics like Dice and IoU.

Regarding input modes, the performance gains of Full-band mode over RGB mode are more pronounced on S2-WaterSeg. For example, DeepLabv3+, PSPNet, U-Net, and LKF-DCANet all show significant improvements in Dice. This highlights that RGB-only information is insufficient for distinguishing water from complex backgrounds, whereas multispectral data provides richer discriminative features. IGF-Net further capitalizes on these advantages, achieving optimal results and underscoring its effectiveness in leveraging multispectral inputs.

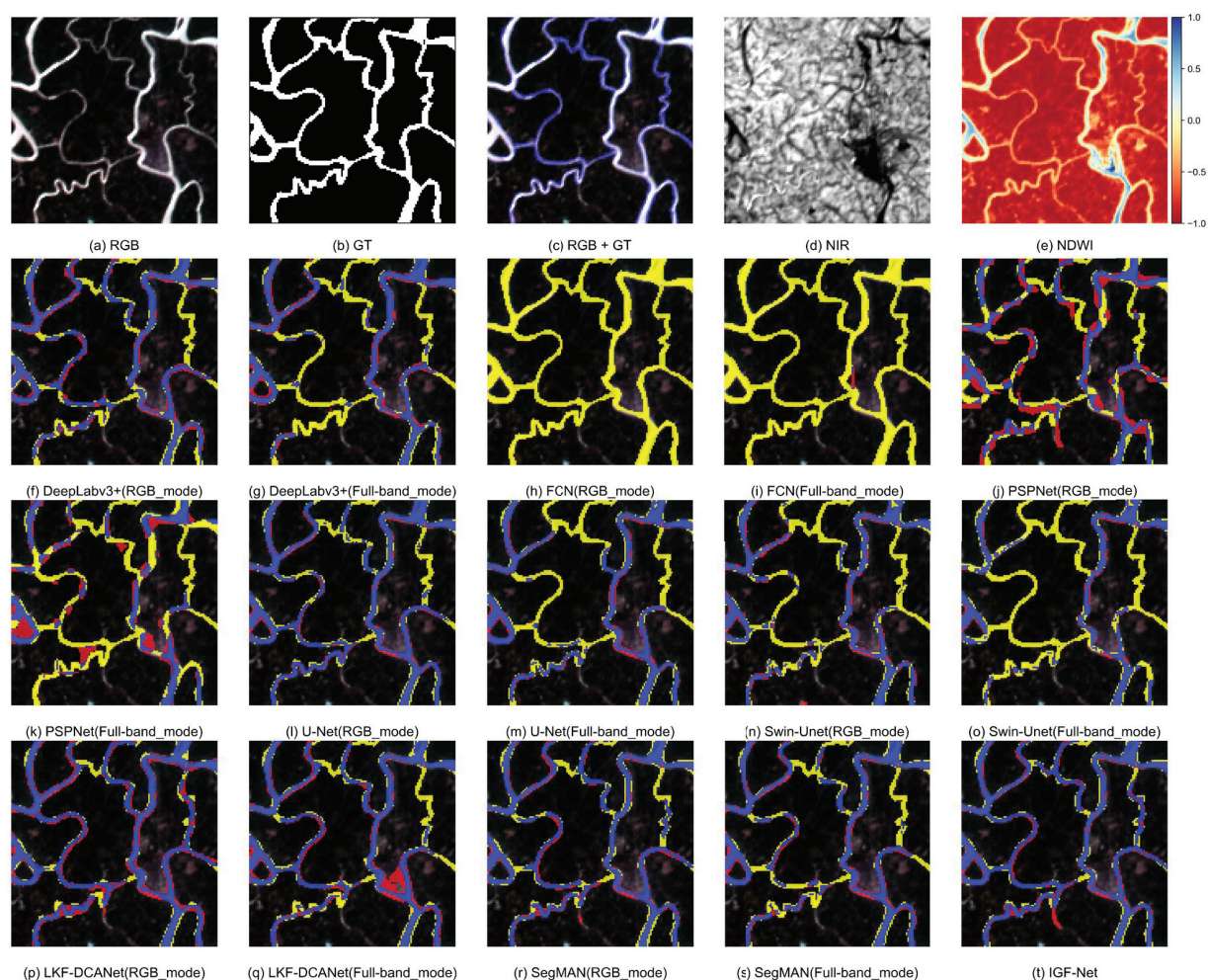
From the perspective of model complexity, IGF-Net exhibits higher Params and FLOPs compared to most comparison methods due to its Dual-Branch Feature Encoder Module. Specifically, on TG2-WaterSeg and S2-WaterSeg, IGF-Net's Params are 45.15 M and 45.12 M, respectively, with FLOPs of 51.48 G and 205.00 G. In terms of inference efficiency, it achieves FPS values of 97.92 and 55.87 on the two datasets, corresponding to Avg-Time of 10.21 ms and 17.90 ms. The reduced inference speed on S2-WaterSeg is primarily attributed to the increased computational overhead caused by higher spatial resolution in input images. Overall, while IGF-Net is not the fastest model, it maintains acceptable practical operational efficiency.

A comprehensive analysis of complexity and performance demonstrates that IGF-Net achieves significant accuracy improvements despite its increased model scale and computational cost. Although its inference efficiency lags behind lightweight models, it outperforms all others in key metrics such as Dice and IoU. This indicates that IGF-Net effectively trades moderate computational overhead for superior segmentation performance. For remote sensing water body segmentation tasks prioritizing high-precision extraction, this accuracy-efficiency trade-off is justified and practically valuable. Compared to pursuing extreme inference speed, achieving higher regional overlap accuracy and more complete water body identification is typically more critical in such applications.

#### 4.2.3. Qualitative Comparison Results and Discussion

To enable direct visual comparison of model performance, this subsection presents visualizations of prediction results and associated errors generated by IGF-Net and comparative models across four typical scenarios in the TG2-WaterSeg and S2-WaterSeg (Figures 4–7). Each figure incorporates multi-modal reference images: (a) input RGB image providing spatial context; (b) binary ground truth water mask where white denotes water

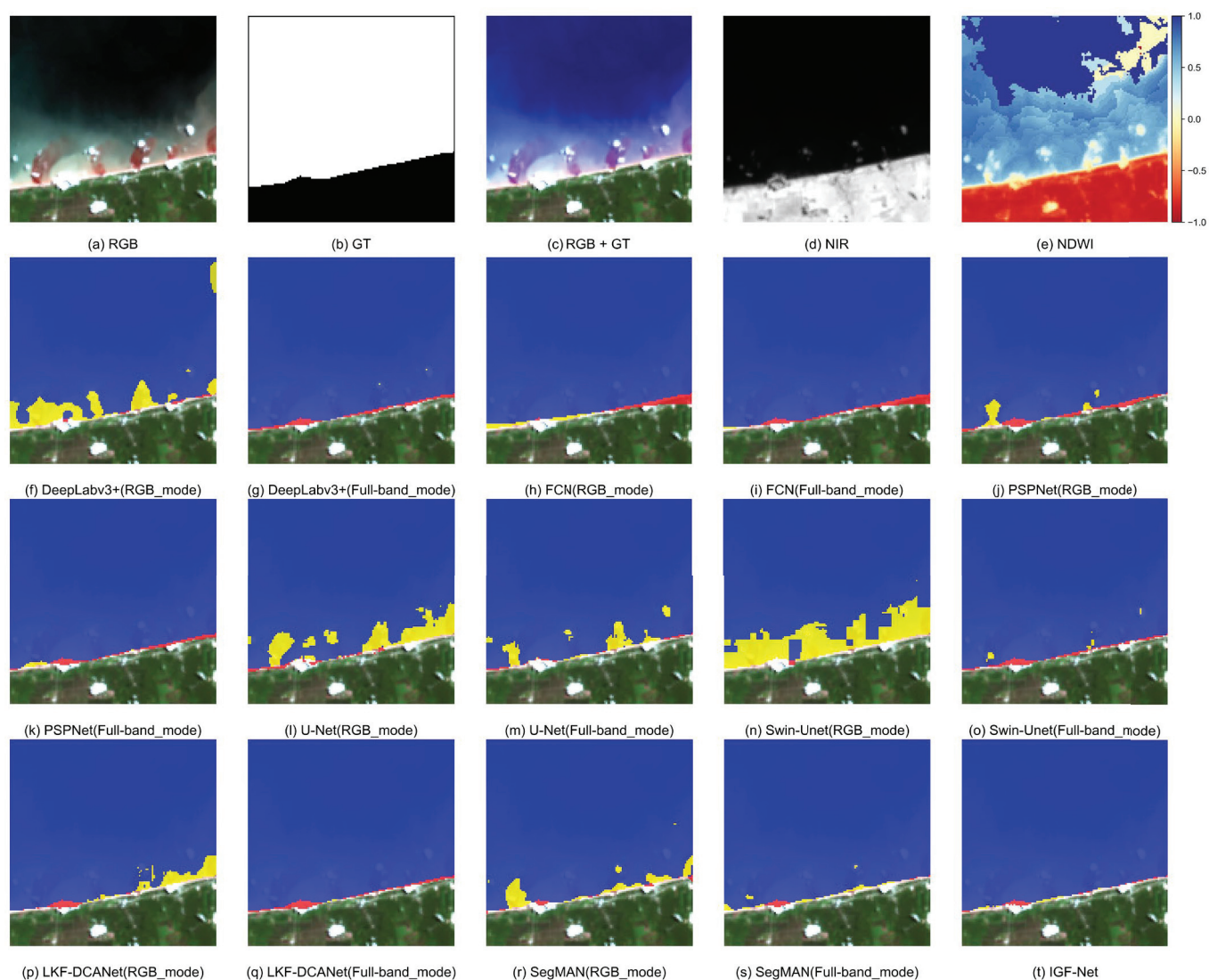
regions and black indicates background; (c) RGB image overlaid with semi-transparent blue water masks to clarify segmentation objectives; (d) water-sensitive near-infrared band grayscale image highlighting reflective characteristics; (e) NDWI map computed through conventional spectral index calculations serving as baseline reference. Standardized error maps in subfigures (f–t) of each figure (Figures 4–7) facilitate comparison of the segmentation results across all models, using blue, red, and yellow hues to denote true positives, false positives, and false negatives, respectively.



**Figure 4.** Visual comparison of water segmentation results and error maps in a complex river scene from the TG2-WaterSeg.

As shown in Figure 4, this sample is selected from the TG2-WaterSeg test set. It exhibits distribution characteristics of a narrow, continuous, and multi-branch water system, presenting a densely meandering spatial morphology. This configuration imposes high demands on the model's capabilities for fine-grained structure perception, boundary localization, and topological connectivity preservation. The comparison results in Figure 4f–s reveal consistent failure modes across different methods in this complex scenario. In RGB mode, most methods demonstrate varying degrees of omission and connectivity disruption in two critical areas: the upper multi-branch confluence region and the mid-lower zigzag narrow tributaries. These errors manifest as fragmented river segments, incomplete branch connections, and disappearance of terminal tributaries. When switching to Full-band mode, the incorporation of additional multispectral band information effectively reduces omissions in main channels. Most methods show improved response integrity for the right

main channel and upper branches compared to RGB mode, validating the discriminative value of supplementary spectral bands for water body identification. However, even with Full-band mode, comparative methods still exhibit notable limitations: (1) persistent local discontinuities or connection errors in mid-lower narrow tributaries and bifurcation nodes; (2) trade-offs between reduced omissions and new artifacts, including boundary misalignment, local morphological distortion, or river width estimation errors. These issues are especially pronounced in highly sinuous reaches and terminal narrow tributaries. In contrast, the IGF-Net result in Figure 4t demonstrates superior structural restoration capability. Its predictions not only achieve more complete tracking of the right longitudinal main channel but also better preserve: (1) the connectivity of upper multi-branch systems, (2) the continuous extension of mid-lower zigzag tributaries, and (3) tighter boundary alignment with ground truth contours at sharp bends and confluence nodes. Overall, IGF-Net surpasses other methods in three key aspects: narrow channel preservation, branch topology restoration, and complex boundary fitting. This performance underscores its enhanced fine-grained structural preservation capability in complex water network environments.



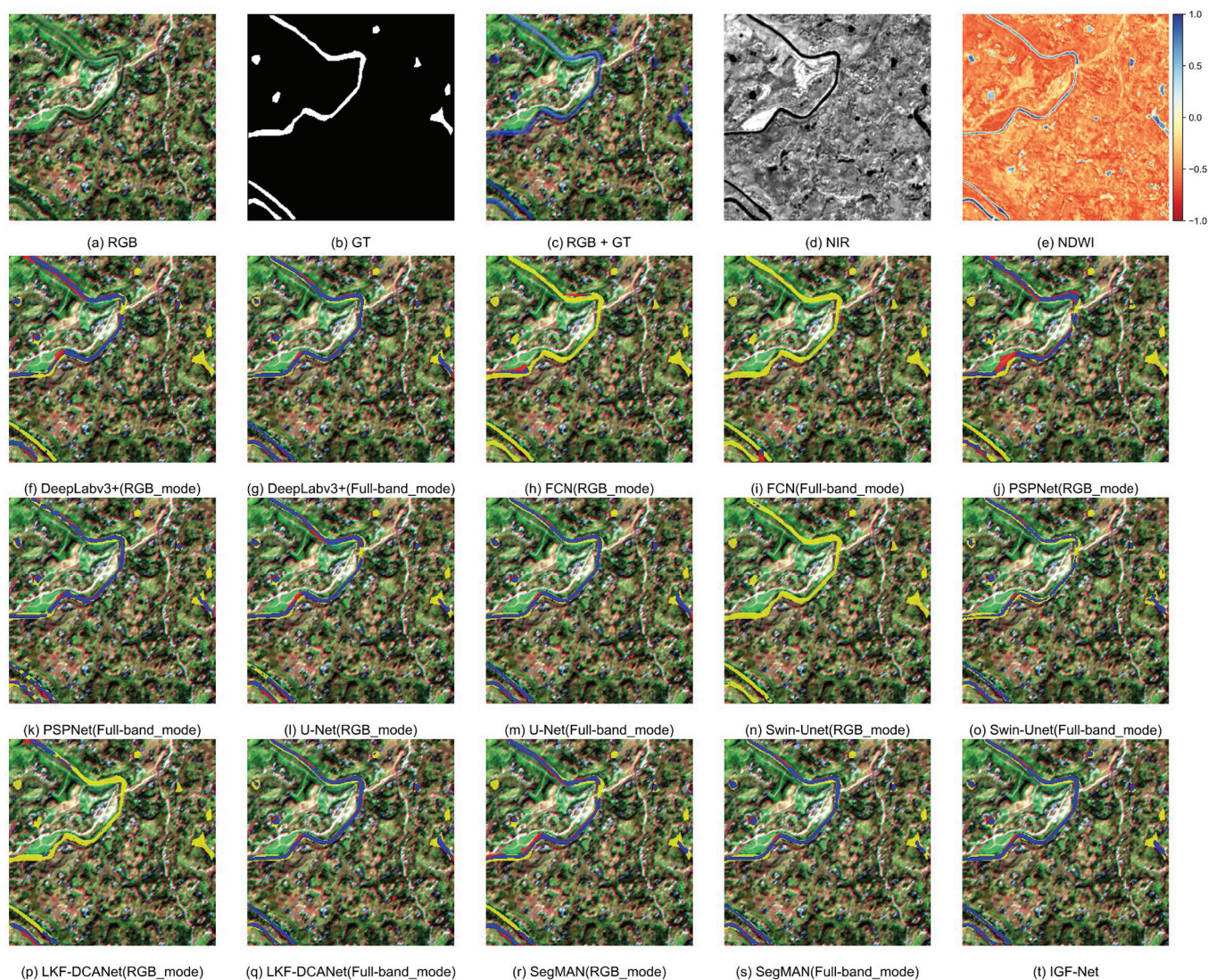
**Figure 5.** Visual comparison of water segmentation results and error maps in a typical coastal scene from the TG2-WaterSeg.



Figure 5 presents a comparative visualization of segmentation results from multiple methods applied to a typical coastal scene in the TG2-WaterSeg. This sample exhibits characteristic features of the land-water interface, with the coastline demonstrating a diagonal trajectory from the lower-left to upper-right. The upper region contains continuous open water, while the lower area comprises a narrow land strip. The land-water boundary is characterized by both considerable length and subtle local undulations. Analysis of Figure 5a,d reveals that nearshore shallow water and intertidal zones lack ideal uniform spectral responses, instead displaying pronounced grayscale gradients and textural disturbances. This scenario necessitates that models not only demonstrate robust water body identification but also effectively address challenges including boundary localization accuracy, transitional zone discrimination, and local noise suppression. Comparative analysis of prediction results in Figure 5f–s identifies two predominant error patterns among baseline methods. The first involves systematic deviations along the true shoreline, manifesting as boundary false positives/negatives relative to ground truth annotations—a consistent observation across most methods indicating generalized coastline localization errors. The second error type appears as contiguous misclassified patches within the nearshore transition zone, where certain models generate extensive blocked errors particularly adjacent to high-reflectance interference areas. Notably, these phenomena exhibit mode-dependent variations: RGB mode shows exacerbated coastline misalignment and speckle artifacts, whereas Full-band mode leverages additional multispectral information to improve main-body segmentation. While full-band inputs reduce overall error magnitude compared to RGB mode, persistent irregularities remain in shoreline details. This suggests that although multi-spectral data enhances general separability, it fails to resolve fundamental challenges in precise coastal boundary alignment. Conversely, IGF-Net demonstrates superior shoreline reconstruction capability. As evidenced in Figure 5t, its errors are constrained to an immediately adjacent narrow band along the true shoreline. The model significantly suppresses large-area patch errors in shallow waters while minimizing discrete noise responses. Particularly noteworthy is the high boundary consistency achieved in both mid-section and terminal right segments of the coastline. These results verify IGF-Net’s dual capability of accurately distinguishing main water bodies from land masses while maintaining precise boundary localization within complex transition zones.

Figure 6 presents a typical mountainous sample from the S2-WaterSeg, containing a meandering narrow main channel, a partially truncated river segment in the lower-left, and several discrete small water bodies distributed across the upper-middle, upper-right, and left-middle regions. The scene exhibits strong terrain-induced texture interference, pronounced curvature variations in the central bend of the main channel, and diminutive spatial extent of isolated water bodies, requiring models to preserve channel continuity while maintaining sensitivity to minuscule targets. Comparative analysis of Figure 6f–s reveals systematic error patterns among baseline methods, predominantly in two domains: high-curvature narrow sections of the main channel and discretely positioned small water bodies. Under RGB modality, several methods exhibit width shrinkage and local discontinuities within the central bend and narrow reaches. For isolated water bodies, multiple models generate fragmented responses confined to central regions while failing to reconstruct complete boundaries, occasionally resulting in total omission. Full-band implementations improve main-channel continuity, particularly in the central bend and lower-left segment, confirming that supplementary spectral information enhances water-background separability. However, for small water bodies covering minimal pixels, numerous comparative methods still demonstrate boundary contraction artifacts, capturing core regions while failing to recover full spatial extent. Conversely, IGF-Net achieves complete recovery

of the tortuous main channel while maintaining geometric fidelity at the highly contorted central bend. It consistently detects all small water bodies without global omission failures. Although minor under-segmentation persists at boundaries of extremely small targets, such errors remain strictly peripheral without compromising core region recovery. This demonstrates IGF-Net's superior robustness in handling size-variant water feature extraction under complex mountainous conditions.

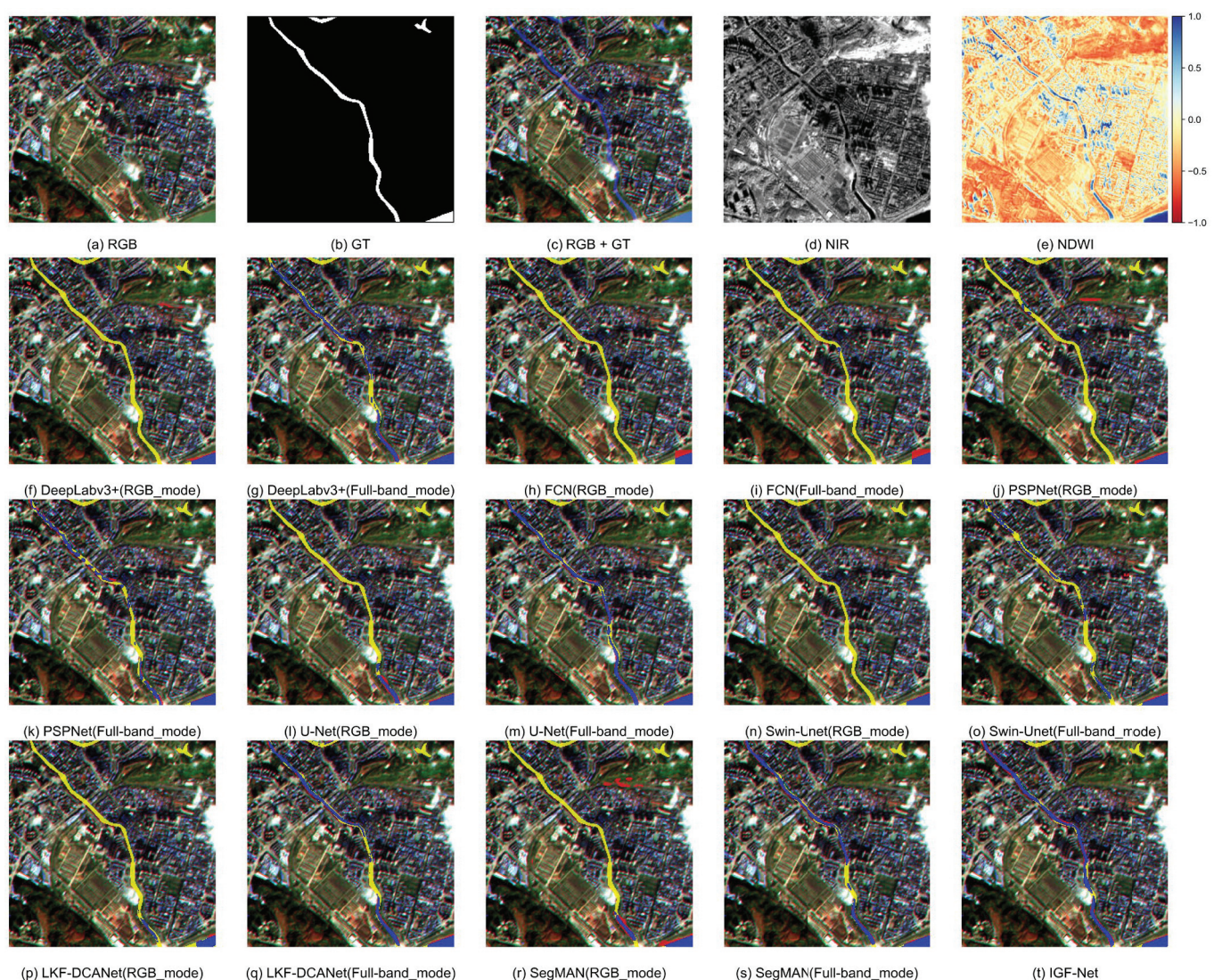


**Figure 6.** Visual comparison of water segmentation results and error maps in a mountainous scene from the S2-WaterSeg.

Figure 7 shows a representative urban scene from the S2-WaterSeg dataset. The target water body is dominated by a narrow main channel extending diagonally from the upper to the lower part of the image, with slight width variations and bends in the densely built central region. Several tiny isolated water bodies are also present near the upper-right and upper boundary. The scene is highly challenging because abundant buildings, roads, and other artificial structures introduce dense background textures, while linear roads, building edges, bright objects, and shadows produce strong spatial and spectral confusion with the narrow channel. As shown in Figure 7f–s, most baseline methods under RGB mode only respond to the upper and lower segments of the main channel, while the central urban section is frequently missing or severely fragmented, accompanied by width shrinkage and



boundary misalignment. Full-band inputs improve channel recognition and allow longer continuous segments to be recovered, but many methods still suffer from discontinuities, boundary drift, and inaccurate width estimation in the narrow central section and other complex urban areas. These results indicate that enhanced spectral information improves separability but remains insufficient for preserving elongated water structures in urban environments. In contrast, IGF-Net achieves complete and continuous extraction of the main channel across the image center, with substantially better alignment to the ground truth in densely built and highly disturbed regions. Although slight omissions remain for tiny isolated water bodies near the upper boundary, IGF-Net shows clear advantages in main-channel integrity, boundary stability, and overall structural consistency.



**Figure 7.** Visual comparison of water segmentation results and error maps in a typical urban scene from the S2-WaterSeg.

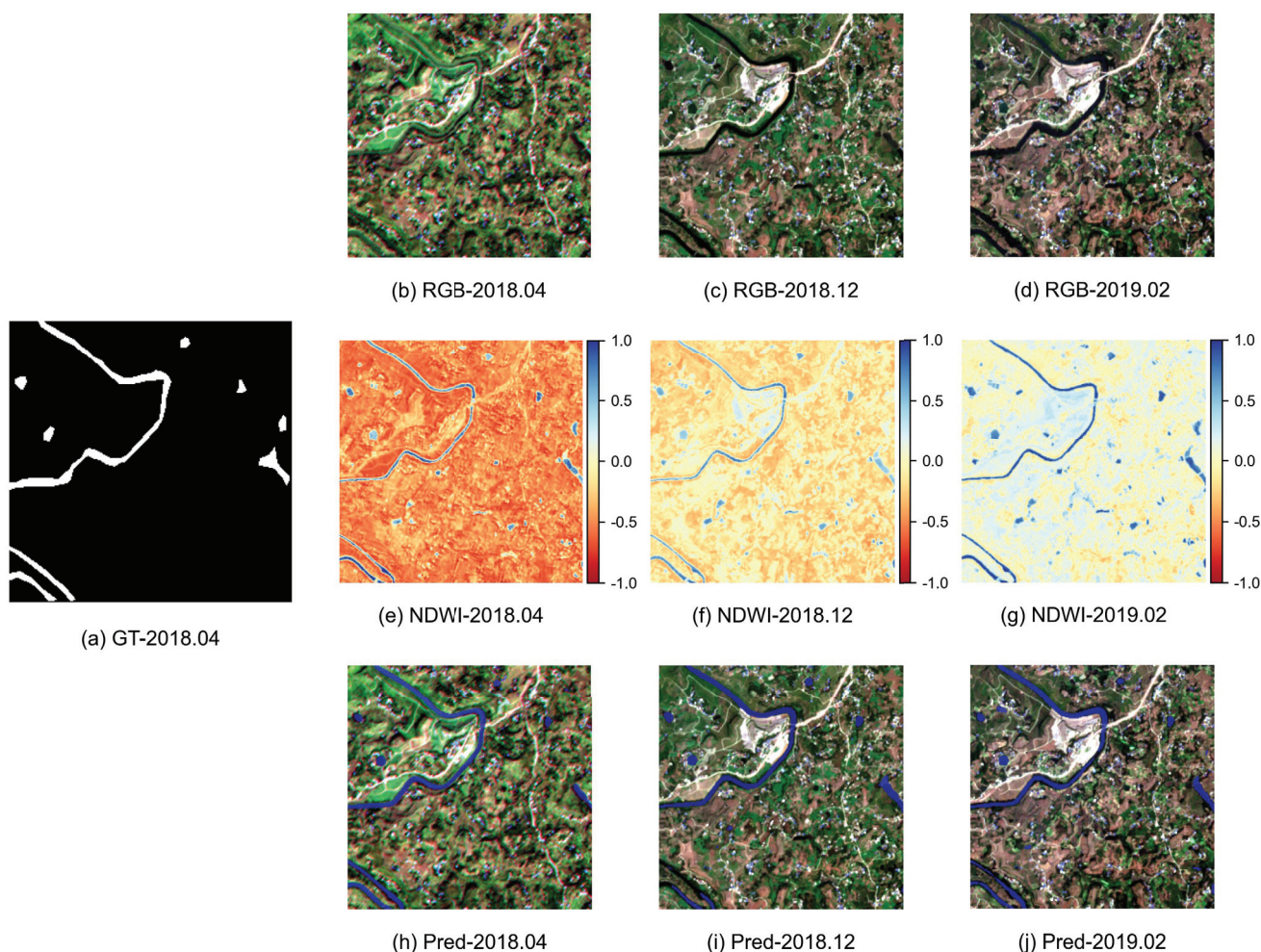
#### 4.2.4. Cross-Temporal Migration Experiment and Result Analysis

To systematically evaluate the model's generalization capability and prediction stability under temporal variations, this study conducts cross-temporal migration experiments using the temporally extended S2-WaterSeg. Unlike the “training-testing” paradigm within the same temporal phase, cross-temporal validation better aligns with real-world scenarios where models must process multi-temporal remote sensing observations. In practice,



models often encounter images acquired at different times, where factors such as solar elevation angle, atmospheric conditions, sensor status, and seasonal surface changes introduce significant variations in spectral response, brightness distribution, textural features, and background context. Such temporal shifts induce distribution discrepancies that increase the difficulty of object identification and segmentation. Due to the lack of pixel-level annotations for the newly added temporal data, unified quantitative metrics cannot rigorously evaluate model performance on unseen phases. Therefore, this study adopts a qualitative analysis: the IGF-Net model trained on the original phase (April 2018) is directly applied to two unseen phases (December 2018 and February 2019) for inference. Prediction results are then compared and analyzed alongside NDWI data for two representative samples.

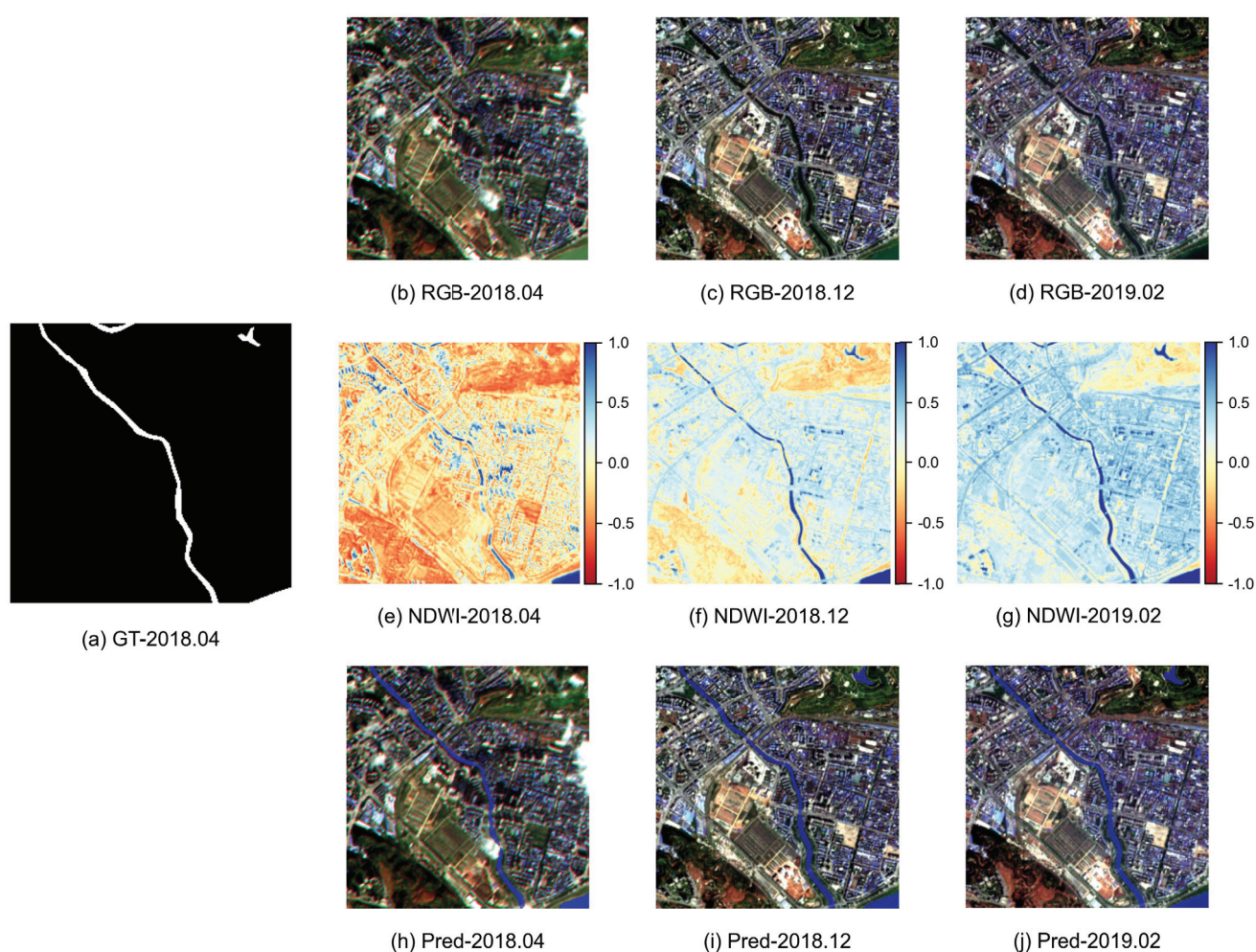
Figures 8 and 9 illustrate representative examples from the cross-temporal migration experiment. Panel (a) displays the ground truth annotation for the training phase (April 2018). Panels (b)–(d) show RGB images acquired in April 2018, December 2018, and February 2019, respectively. Panels (e)–(g) present the corresponding NDWI results for these phases. Finally, panels (h)–(j) depict prediction overlays generated by the model trained on the April 2018 phase and applied to all three temporal phases.



**Figure 8.** Visualization results of cross-temporal migration in a typical mountainous scene from the S2-WaterSeg.

As shown in Figure 8, this sample exhibits pronounced observational differences across the three temporal phases. Compared to April 2018, the RGB images from December 2018 and February 2019 display notable changes in overall color tone, brightness gradients,

and background textural patterns, indicating that temporal variations have induced a shift in the input feature distribution. For water regions, the main river channel retains its spatial positioning and overall trajectory across all three phases, with visual discrepancies primarily observed in localized shorelines, near-shore transitional areas, and the morphology of scattered small water bodies. The corresponding NDWI results further highlight that while the main channel sustains consistently high spectral responses across different acquisition dates, the index values over non-water backgrounds exhibit distinct alterations. Despite these input variations, the model trained on April 2018 data, when applied to the two unseen phases, successfully maintains continuous extraction of the main river channel. It preserves connectivity and curvilinear morphology effectively, with only minor adaptations at local boundaries and small-scale water features. This underscores the model's robust ability to retain dominant water structures and deliver stable predictions despite temporal shifts, demonstrating strong cross-temporal generalization.



**Figure 9.** Visualization results of cross-temporal migration in a typical urban scene from the S2-WaterSeg.

As illustrated in Figure 9, this sample demonstrates significant visual discrepancies across the three temporal phases. Notably, the April 2018 image appears overall darker with distinct bright spots, contrasting sharply with December 2018 and February 2019 in both imaging aesthetics and textural sharpness. This highlights that cross-temporal model migration must address not only seasonal shifts but also perturbations arising from variable imaging conditions. Furthermore, NDWI analysis confirms that while the main river



channel retains strong spectral responses across all dates, the index values over non-water backgrounds show pronounced alterations. Specifically, background regions in December 2018 and February 2019 exhibit systematically higher NDWI values compared to April 2018, exacerbating land-water ambiguity risks in urban settings. Despite these challenges, the primary river skeleton remains spatially consistent across phases, maintaining identical positional alignment and directional trends, with minor deviations confined to narrow channels, bends, and vicinities of small tributary water bodies. The model robustly sustains the main channel's connectivity throughout all phases. However, in April 2018, cloud cover and shadowing in the upper-right quadrant of the input image caused partial misdetection of small water features, whereas these areas were accurately resolved in later phases. This resilience underscores the model's capacity to generalize across temporal variability.

#### 4.3. Ablation Study: Replacing the Fusion Module

##### 4.3.1. Ablation Experiment Setup

To systematically evaluate the contribution of the proposed IGFM to water-body semantic segmentation in remote sensing images, this study designs a corresponding ablation experiment: On the IGF-Net architecture, the impact mechanism of sequentially replacing or removing the fusion module on multi-source feature fusion (the visual branch and the spectral branch) is analyzed from both quantitative metrics and interpretability visualization perspectives. To ensure comparability among different model variants and conclusion reliability, all experiments follow consistent parameter settings, with differences limited solely to the model fusion method.

Specific experimental models are as follows:

**Model 1:** The baseline control model. This model completely removes the IGFM and uses only simple channel concatenation followed by a  $1 \times 1$  convolution for feature fusion and dimensionality reduction. It is designed to examine whether direct feature concatenation and convolutional operations are sufficient for multi-source fusion tasks without explicit interaction/attention mechanisms.

**Model 2:** Replaces the IGFM with an iterative attention feature fusion module [35]. This module progressively refines fused features through cascaded/iterative attention while enhancing local salient regions. This configuration allows for a direct comparison between the IGFM and an iterative-optimization-based attention fusion approach.

**Model 3:** Replaces the IGFM with a bilinear attention fusion module [36]. This module captures high-order correlations between features through bilinear interaction modeling to obtain finer-grained interactive representations. It is used to evaluate the applicability of high-order feature interaction in multispectral water-body segmentation and to compare the module's interaction modeling approach with that of the IGFM.

**Model 4:** Replaces the IGFM with a cross-modal attention feature fusion module [37]. This module has demonstrated strong mutual-information extraction capability in multimodal data fusion. In this experiment, it is introduced into the multispectral remote-sensing scenario to examine the module's effectiveness in handling two types of feature sources: spectral and spatial.

**Model 5:** Replaces the IGFM with a dual cross-attention mechanism module [38]. This module enables mutual enhancement of the two branch features through two-stage, bidirectional attention interaction, providing strong semantic-level representation capacity. It is thus employed to benchmark the performance of the IGFM against an advanced cross-attention architecture.

As discussed in Section 3.2, the Dual-Branch Feature Encoder Module exhibits fundamental differences in feature representation: the visual branch primarily carries general spatial-semantic knowledge acquired through transfer learning, while the spectral branch

focuses on subtle reflectance variations corresponding to different land classes in non-RGB bands of multispectral data. To systematically investigate the internal working mechanisms of different fusion modules during the “visual-spectral” feature integration process, this study computes and visualizes feature conflict maps at the input end and post-fusion feature activation maps at the output end of the fusion modules, thereby revealing their intrinsic fusion mechanisms.

Specifically, given the two pre-fusion branch features,  $F_v$  and  $F_s$ , this paper first calculates the  $L_2$  norm of the feature vector at each spatial position as the response intensity for that position, thereby obtaining the response intensity maps for both branches:

$$\begin{aligned} M_v(x, y) &= \|F_v(:, x, y)\|_2 \\ M_s(x, y) &= \|F_s(:, x, y)\|_2 \end{aligned} \quad (11)$$

On this basis, the absolute difference in response intensity between the two branches is defined as the raw conflict intensity  $D_{\text{raw}}$ .

$$D_{\text{raw}}(x, y) = |M_v(x, y) - M_s(x, y)| \quad (12)$$

To visualize the spatial distribution of feature conflicts, the raw conflict intensity  $D_{\text{raw}}(x, y)$  is further normalized into the range  $[0, 1]$  using a min-max normalization strategy. Specifically, for each sample, the minimum and maximum values of the conflict map are computed, and the normalized conflict intensity is defined as:

$$D_{\text{norm}}(x, y) = \frac{D_{\text{raw}}(x, y) - D_{\min}}{D_{\max} - D_{\min} + \varepsilon} \quad (13)$$

where  $D_{\min}$  and  $D_{\max}$  denote the minimum and maximum values of the raw conflict map for the current sample, respectively, and  $\varepsilon$  is a small positive constant to avoid division by zero. This normalization rescales the conflict intensity of each sample to the interval  $[0, 1]$ , facilitating clearer visualization of spatial conflict patterns.

The post-fusion feature activation map is defined as the spatial response obtained by applying the  $L_2$  norm along the channel dimension to the fused feature tensor  $T$  output by the fusion module:

$$A(x, y) = \|T(:, x, y)\|_2 \quad (14)$$

Then, min-max normalization is applied to  $A(x, y)$  to map it into the range  $[0, 1]$ :

$$A_{\text{norm}}(x, y) = \frac{A(x, y) - A_{\min}}{A_{\max} - A_{\min} + \varepsilon} \quad (15)$$

Due to the use of sample-specific min-max normalization, the pixel values in the conflict map and post-fusion feature activation map represent relative differences between spatial locations rather than absolute quantities comparable across samples. The performance of the fusion module can be evaluated from two dimensions: first, the spatial concentration of boundary activations—whether high activations are localized in water-land transition zones; second, the quality of activation suppression in non-boundary regions—whether feature energy within water bodies and backgrounds is effectively suppressed. Furthermore, since both maps are generated by upsampling low-resolution features, the interpolation process tends to blur fine-grained structures and introduce visible diffusion artifacts in corresponding areas. Therefore, these maps are more suitable for observing overall distribution patterns and their correlation with error-prone regions, rather than serving as precise pixel-level boundary references.



#### 4.3.2. Quantitative Ablation Results

Module-replacement experiments were conducted on TG2-WaterSeg and S2-WaterSeg. Table 7 summarizes the quantitative results of each substitution scheme on the respective test sets, with the main network structure and training configuration held identical. In these experiments, IGF-Net serves as the baseline for the replacement experiments, and its results are consistent with those reported in the earlier comparative experiments.

**Table 7.** Quantitative Results of Module-Replacement Experiments on TG2-WaterSeg and S2-WaterSeg.

| Models       |                       | IoU           | Dice          | Precision     | Recall        |
|--------------|-----------------------|---------------|---------------|---------------|---------------|
| TG2-WaterSeg | Model 1               | 0.8712        | 0.9224        | 0.9269        | 0.9287        |
|              | Model 2               | 0.8713        | 0.9227        | 0.9271        | 0.9284        |
|              | Model 3               | 0.8674        | 0.9207        | <b>0.9339</b> | 0.9166        |
|              | Model 4               | 0.8734        | <b>0.9239</b> | 0.9221        | <b>0.9367</b> |
|              | Model 5               | 0.8686        | 0.9210        | 0.9336        | 0.9184        |
|              | <b>IGF-Net (ours)</b> | <b>0.8742</b> | <b>0.9239</b> | 0.9331        | 0.9235        |
| S2-WaterSeg  | Model 1               | 0.4918        | 0.6283        | <b>0.7194</b> | 0.6024        |
|              | Model 2               | 0.4893        | 0.6263        | 0.7158        | 0.5957        |
|              | Model 3               | 0.4774        | 0.6145        | 0.7036        | 0.5926        |
|              | Model 4               | 0.4971        | 0.6322        | 0.7006        | 0.6154        |
|              | Model 5               | 0.4601        | 0.5946        | 0.6914        | 0.5740        |
|              | <b>IGF-Net (ours)</b> | <b>0.5009</b> | <b>0.6370</b> | 0.7090        | <b>0.6198</b> |

Overall, IGF-Net achieves relatively strong comprehensive performance on both datasets, which verifies the effectiveness of the IGFM for feature fusion. In particular, on the more challenging S2-WaterSeg dataset, IGF-Net obtains the best results in IoU, Dice, and Recall, indicating that the IGFM may be more conducive to coordinating the complementary information between the visual branch and the spectral branch under complex backgrounds and diverse water-body morphologies.

On TG2-WaterSeg, all models perform well overall, with IoU values above 0.8670 and Dice values above 0.9200. IGF-Net achieves the highest IoU (0.8742) and the joint-highest Dice (0.9239) on this dataset, demonstrating the potential to further improve the overall overlap between predictions and ground-truth annotations. Compared with IGF-Net, Model 1 yields slightly lower results, suggesting that although simple concatenation and linear compression can preserve a certain amount of complementary information, they may still be insufficient in terms of explicit interaction modeling and feature selection. Model 2 brings only marginal performance changes, indicating that iterative attention is more oriented toward enhancing local salient regions, while its effect on modeling semantic consistency between features from different sources is relatively limited. Model 3 achieves the highest Precision of 0.9339, which to some extent indicates that bilinear high-order interaction helps enhance category discriminability and reduce false positives. However, its Recall decreases to 0.9166, suggesting that this method may adopt a relatively conservative prediction strategy, resulting in insufficient coverage of some real water-body regions and ultimately causing its IoU and Dice to be slightly lower than those of IGF-Net. Among all replacement schemes, Model 4 performs most closely to IGF-Net: its Dice is equal to that of IGF-Net, and its Recall reaches the highest value of 0.9367. This indicates that cross-modal attention has certain advantages in mining the complementary relationship between visual features and spectral features and in improving target coverage. However, its relatively lower Precision suggests that while expanding the target response range, this method may also introduce more false detections. Model 5 still exhibits a trend of relatively high Precision but comparatively low Recall, which to some extent indicates that although

dual cross-attention strengthens the semantic correlation between branches, it has not yet achieved a more desirable balance between complete target-region recovery and false-positive control. Overall, IGF-M shows better comprehensive stability on TG2-WaterSeg.

On S2-WaterSeg, the performance advantage of IGF-Net is more significant. Its IoU, Dice, and Recall reach 0.5009, 0.6370, and 0.6198, respectively, outperforming all alternative models on these three metrics. Specifically, although Model 1 achieves the highest Precision (0.7194), its IoU and Dice are both lower than those of IGF-Net. This suggests that simple fusion methods are prone to missed detections in complex water-body regions. From the experimental results, Model 2 does not show a more obvious performance advantage over Model 1, which to some extent indicates that the effectiveness of the iterative attention strategy in this complex scenario is not particularly prominent. The overall performance of Model 3 declines, which may suggest that although bilinear high-order interaction can enhance correlation modeling between features, it is also more likely to introduce noise interference under complex background conditions, thereby affecting the stability of the fusion results. Model 4 remains highly competitive, with its IoU, Dice, and Recall ranking only behind IGF-Net, indicating that the cross-modal interaction mechanism has strong application potential in multi-source feature fusion. Compared with Model 4, IGF-Net achieves slight improvements on all metrics, suggesting that the IGF-M not only constructs effective cross-branch semantic associations, but also achieves a better balance between redundant-information suppression and effective-feature preservation. In contrast, Model 5 does not show obvious overall advantages on this dataset, which may indicate that although its bidirectional cross-coupling mechanism enhances semantic interaction between the two branches, its improvement in information propagation stability and error control remains relatively limited in the current task scenario.

#### 4.3.3. Qualitative Ablation Results and Discussion

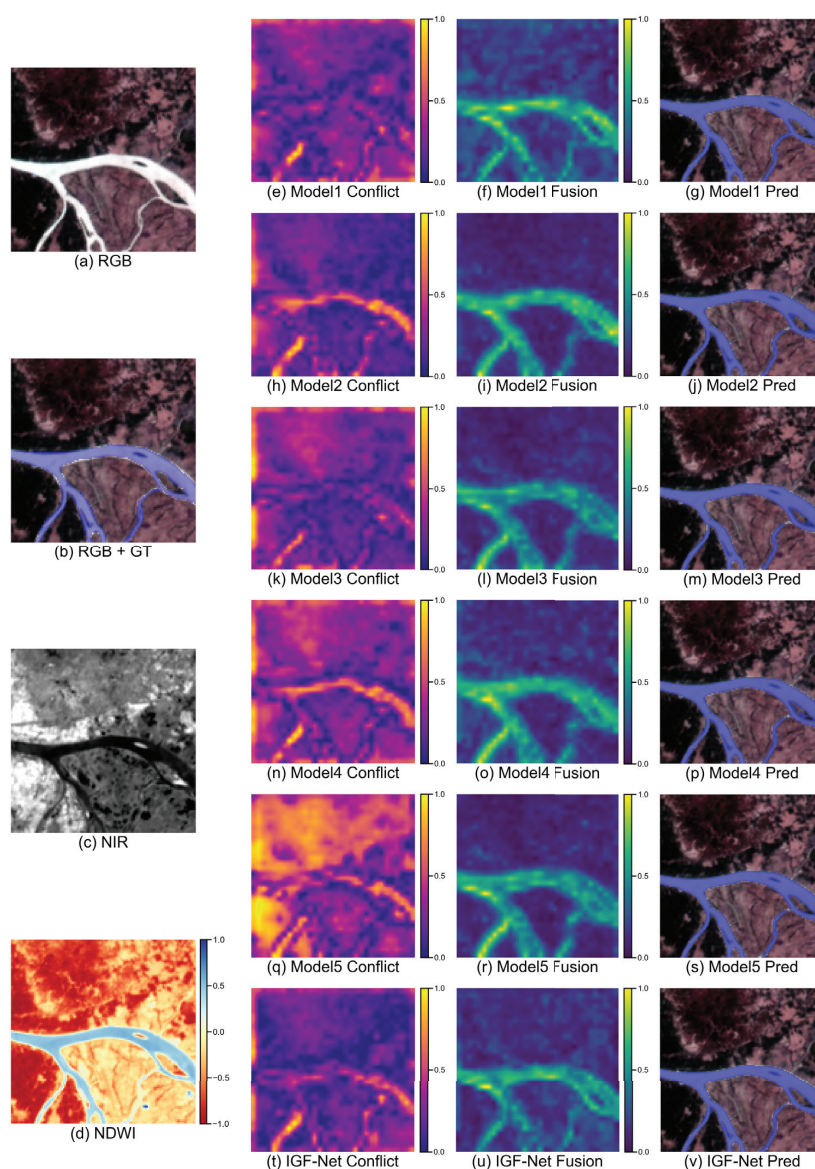
For a visual demonstration of the differences among fusion modules, this section presents two representative samples from the TG2-WaterSeg and S2-WaterSeg test sets. Each image follows a consistent layout: the left column sequentially shows the RGB image, the RGB image with ground truth overlay, the NIR image, and the NDWI image; the right column displays the predictions of benchmark models and IGF-Net, along with their corresponding conflict and fusion maps. Notably, because all models are trained end-to-end, each fusion strategy indirectly shapes the learning of branch features via backpropagation. As a result, distinct conflict patterns emerge even during feature extraction, ultimately affecting the final fusion outcomes. Hence, conflict and fusion maps are best interpreted as references for understanding feature abstraction, rather than as absolute evidence of model performance.

As illustrated in Figure 10, this case study presents a riverine scene from the TG2-WaterSeg dataset. The region exhibits heterogeneous land-cover composition, including riverine water bodies, exposed soil, and vegetated areas. The water network shows a characteristic meandering morphology, with narrow tributary bifurcations in the lower section and shallow shoals along the right bank. Such geomorphological complexity introduces pronounced class ambiguity and spatial uncertainty at the land–water transition zones.

In the conflict maps, all models produce high-magnitude responses along the boundaries of the narrow tributaries in the lower part of the image, indicating that these regions—characterized by slender water morphology and weak textural contrast with the surrounding background—constitute highly uncertain boundaries. Model 2, Model 4, and IGF-Net form relatively continuous conflict-response bands along the boundaries of the main river channel, whereas Model 1 and Model 3 exhibit weaker conflict responses near the main river trunk. In contrast, Model 5 produces extensive patch-like high responses over land areas, reflecting significant discrepancies in the Dual-Branch Feature Encoder Module

when representing background land features. Overall, several comparative methods still retain varying degrees of background conflict residuals. By contrast, IGF-Net concentrates high-conflict responses primarily along genuine land–water boundaries and key structural details while suppressing non-boundary regions to low-value backgrounds. This demonstrates IGF-Net’s stronger capability in conflict localization and noise suppression.

In the fusion activation maps, although all models produce activations along land–water transition zones, clear differences emerge in spatial morphology and background suppression. At the narrow downstream tributaries, IGF-Net’s activations adhere more closely to the boundaries, forming sharper high-activation zones than those of the comparative models and thereby achieving better edge definition and diffusion-noise suppression. In internally inconsistent water regions such as the right-bank shoal, IGF-Net generates more concentrated activation patterns, enabling more precise delineation of local boundary details during post-processing.



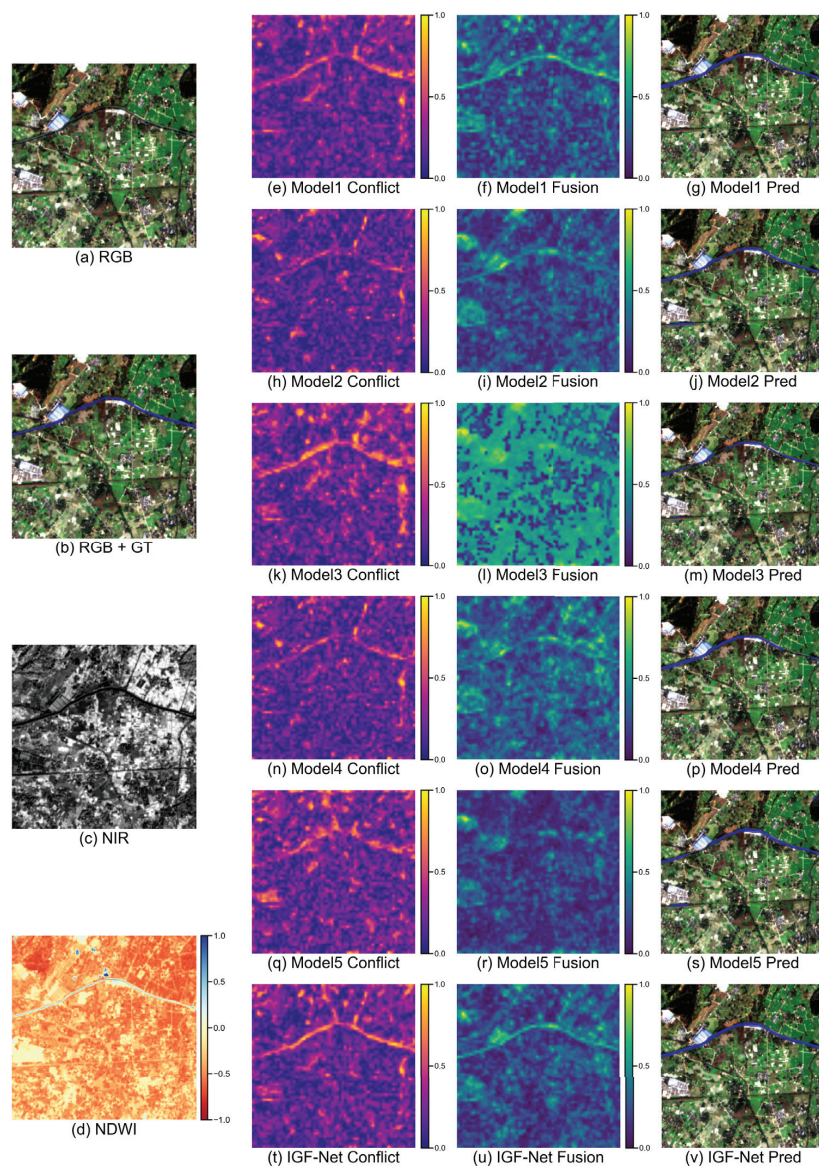
**Figure 10.** Qualitative comparison of conflict maps, fusion activation maps, and segmentation results on a complex riverine scene from the TG2-WaterSeg.

In the final segmentation results, all methods successfully reconstruct the macroscopic contour of the main river channel, demonstrating robust large-scale water-body recognition



capability. However, inter-model differences become evident in uncertain regions such as narrow tributaries and shoal vicinities, mainly manifested as pixel-level boundary deviations and contour discontinuities. Benefiting from accurate conflict localization and low-redundancy fusion representations with enhanced background discrimination, IGF-Net achieves the best performance on this sample, reaching an IoU of 0.8667 compared with 0.8644 (Model 1), 0.8496 (Model 2), 0.8465 (Model 3), 0.8449 (Model 4), and 0.8592 (Model 5). Although these improvements are only subtly visible in thumbnail visualizations, the quantitative metrics confirm IGF-Net's consistent advantages in handling complex boundaries and fine-grained details.

As illustrated in Figure 11, this S2-WaterSeg scene depicts a rural agricultural canal landscape. The main channel arcs across the upper part of the image with a slender, elongated morphology. Surrounding farmland is fragmented into small parcels and interwoven with dense linear structures such as paths and field ridges. These elements create strong spectral and morphological similarities between narrow water bodies and the background, substantially increasing the difficulty of land–water boundary discrimination.



**Figure 11.** Qualitative comparison of conflict maps, fusion activation maps, and segmentation results on a rural agricultural canal scene from the S2-WaterSeg.

The conflict maps show strong responses around the main channel for all models, confirming this region as the primary locus of cross-modal representation discrepancies. However, their spatial distributions differ markedly. Model 1 roughly outlines the channel conflict zone but retains noticeable background noise. Model 2 and Model 4 exhibit more dispersed point-like and strip-like patterns across the scene, including redundant conflicts in non-water areas. Model 3 and Model 5 present patchy and discrete high responses. In contrast, IGF-Net produces high-conflict responses that closely follow the true course and boundaries of the channel, demonstrating more accurate conflict localization.

In the fusion activation maps, all models except Model 3 and Model 5 generate responses along the land–water transition zones. Although Model 1, Model 2, and Model 4 cover the channel area, residual moderate background responses may still cause misclassification. Model 3 produces large, spatially unselective patch-like activations with blurred boundaries, while Model 5 shows generally weak activations accompanied by scattered spurious responses. By comparison, IGF-Net maintains continuous and concentrated activations along the main channel with cleaner background suppression, enabling more reliable discrimination of slender water bodies under complex textures.

The final segmentation results show that all methods recover the overall course of the channel, but two typical failure modes remain: linear infrastructures such as field ridges and paths often induce local false detections, and narrow channel segments may exhibit discontinuities or inaccurate width estimation. Benefiting from more accurate conflict localization and low-redundancy fusion representations, IGF-Net achieves the best performance on this sample, reaching an IoU of 0.6627. This surpasses Model 1–5 (0.6200, 0.6021, 0.4917, 0.5252, and 0.5159, respectively), representing a 4.27-percentage-point improvement over the strongest competing method (Model 1). These results further validate the effectiveness of the IGFM in complex backgrounds with slender water bodies.

## 5. Discussion

The experimental results presented in Chapter 4 demonstrate that IGF-Net achieves superior performance on both TG2-WaterSeg and S2-WaterSeg. Its core module, IGFM, explicitly models modality-specific differences and modality-consistent activations, enabling adaptive fusion of visual and spectral features.

On TG2-WaterSeg, where most models already exhibit competitive performance indicating saturation, IGF-Net still achieves meaningful improvements in key metrics. Notably, IoU reaches 0.8742—a gain of 1.4 percentage points over the best baseline. This gain under saturated conditions validates IGF-Net’s effectiveness in detail preservation and boundary optimization. On the more challenging S2-WaterSeg, IGF-Net shows even greater quantitative improvements, suggesting IGFM’s advantages become more prominent with increased task complexity. Qualitative analyses reveal IGF-Net better preserves fine-grained structures like narrow tributaries and small isolated water bodies compared to other methods. Cross-temporal experiments demonstrate robustness against seasonal variations and imaging condition changes, maintaining main channel continuity through structural/contextual feature learning rather than shallow spectral cues.

However, IGF-Net’s dual-branch feature encoder module introduces higher computational costs: on TG2-WaterSeg, it requires 45.15 M parameters, 51.48 G FLOPs, and achieves 97.92 FPS, showing lower efficiency than lightweight methods. This impacts deployability in resource-constrained scenarios. Additionally, the model lacks systematic validation under extreme atmospheric conditions or with auxiliary data (e.g., DEM and SAR). Ablation studies show while IGFM achieves balanced metric performance, scenario-specific optimization remains necessary for precision/recall requirements.



Despite these limitations, IGF-Net's accurate water extraction under complex conditions holds significant practical value. High IoU/Dice scores translate to reliable land cover products for flood monitoring, water resource management, and environmental change detection. The structure-preserving capability is particularly crucial for monitoring narrow channels, coastal erosion, and small urban/agricultural water bodies where errors impact downstream analyses. Furthermore, IGFM can be extended to other remote sensing tasks that require multispectral or multi-source data fusion.

## 6. Conclusions

This study addresses the challenges of pre-trained model transferability and inadequate multi-band feature fusion in water body segmentation for multispectral remote sensing images, proposing a novel segmentation network and constructing a supporting dataset. The main contributions and conclusions are as follows:

- (1) The Tiangong-2 Remote Sensing Image Water Body Semantic Segmentation Dataset was constructed and made publicly available. This dataset provides 3776 multispectral image groups with pixel-level fine annotations for water body segmentation. To systematically validate model generalization, the public Sentinel-2 Water Segmentation Dataset was also introduced, enabling comprehensive performance evaluation across different sensors, band combinations, and scene complexities.
- (2) This study proposes the IGF-Net architecture to address the challenges of poor transferability in pre-trained models and inadequate multi-source feature fusion. The network employs a dual-branch encoder as its backbone, centered around the IGFM. Through a cascaded mechanism integrating difference-co-occurrence parallel modeling, channel-context prior, and adaptive temperature control, this module achieves adaptive deep fusion of visual and spectral features. This design not only promotes training stability but also improves water body segmentation accuracy.
- (3) Extensive experiments validate the effectiveness and generalization capability of IGF-Net and its core module, IGFM. On TG2-WaterSeg, IGF-Net achieves highly competitive performance (IoU 0.8742), demonstrating its ability to further refine segmentation details in scenarios with relatively limited room for improvement. On the more challenging S2-WaterSeg, IGF-Net shows clearer advantages in key metrics such as IoU (0.5009) and Dice (0.6370), indicating that IGFM is more effective at coordinating complementary information from the visual and spectral branches and achieves a better balance between segmentation completeness and false-positive suppression. Ablation studies comparing five alternative fusion modules further confirm that IGFM provides stronger overall stability, particularly in complex scenarios. Visualization results further show that IGF-Net concentrates cross-modal discrepancies on true boundaries and critical structural regions while suppressing background interference, producing more continuous and clearer responses to complex boundaries and elongated water bodies. Cross-temporal experiments demonstrate that IGF-Net maintains the continuity of major channels despite seasonal variations and imaging condition differences, indicating its ability to learn relatively robust features based on structural and contextual cues.
- (4) Complexity analysis reveals that IGF-Net's dual-branch design leads to higher computational costs compared to lightweight methods, which may limit deployment in resource-constrained scenarios. The model has not been validated under extreme atmospheric conditions or with auxiliary data such as DEM or SAR. Additionally, ablation results suggest that while IGFM achieves the best overall balance, optimal fusion strategies may vary with specific precision-recall requirements.

Building upon the solutions proposed from the perspectives of data, model, and validation, future work will focus on the following directions: (1) expanding data coverage to include more sensor types and more complex scenarios (e.g., heavy cloud/fog, shadows, turbid water and seasonal variations), and investigating stronger domain adaptation strategies; (2) introducing finer boundary constraints and structural priors to further improve pixel-level segmentation accuracy for narrow water bodies and complex shorelines; (3) exploring model light-weighting and inference acceleration methods to enhance practicality for large-scale remote sensing mapping and real-time engineering applications; investigating the transferability of the IGF design philosophy to other multispectral and multi-source fusion tasks in remote sensing.

**Author Contributions:** Conceptualization, T.Z., C.H. and Z.Z. (Zhaofa Zhou); methodology, T.Z.; software, T.Z.; validation, T.Z., C.H. and Z.Z. (Zhaofa Zhou); writing—original draft preparation, T.Z. and C.H.; writing—review and editing, T.Z. and Z.Z. (Zhaofa Zhou); supervision, Z.Z. (Zhili Zhang). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China under grant 62305393; The Science and Technology Innovation Team Projects of Shaanxi Province under grant 2025RSCXTD-046.

**Data Availability Statement:** The source imagery is available from the National Basic Science Data Center (NBSDC) under the accession code CSTR:16666.11.nbsdc.tfpbwtqf (accessed on 25 March 2026). The Sentinel-2 multispectral water-body segmentation dataset employed in this study is publicly available from the project repository described in reference [14]: <https://github.com/SCouly/Sentinel-2-Water-Segmentation> (accessed on 25 March 2026). Furthermore, the derived semantic segmentation annotations and the code used for preprocessing, training, and evaluation in our work are available at: [https://github.com/RFUzt/article\\_IGF-Net\\_code\\_and\\_dataset.git](https://github.com/RFUzt/article_IGF-Net_code_and_dataset.git) (accessed on 25 March 2026).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- McFeeters, S.K. The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]
- Cao, R.; Li, C.; Liu, L. Extracting Miyun Reservoir’s Water Area and Monitoring Its Change Based on a Revised Normalized Difference Water Index. *Sci. Surv. Mapp.* **2008**, *33*, 158–160. (In Chinese)
- Chen, W.; Ding, J.; Li, Y.; Niu, Z. Extraction of Water Information Based on China-Made GF-1 Remote Sensing Image. *Resour. Sci.* **2015**, *37*, 1166–1172. (In Chinese)
- Wang, S.; Baig, M.H.A.; Zhang, L.; Jiang, H.; Ji, Y.; Zhao, H.; Tian, J. A Simple Enhanced Water Index (EWI) for Percent Surface Water Estimation Using Landsat Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 90–97. [CrossRef]
- Chen, C.; Fu, J.Q.; Sui, X.X.; Lu, X.; Tan, A.H. Construction and Application of Knowledge Decision Tree after a Disaster for Water Body Information Extraction from Remote Sensing Images. *J. Remote Sens.* **2018**, *22*, 792–801. (In Chinese) [CrossRef]
- Wang, Z.; Liu, J.; Li, J.; Zhang, D.D. Multi-Spectral Water Index (MuWI): A Native 10-m Multi-Spectral Water Index for Accurate Water Mapping on Sentinel-2. *Remote Sens.* **2018**, *10*, 1643. [CrossRef]
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
- Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20–25 June 2009; pp. 248–255. [CrossRef]
- Zhang, L.; Fan, Y.; Yan, R.; Shao, Y.; Wang, G.; Wu, J. Fine-Grained Tidal Flat Waterbody Extraction Method (FYOLOv3) for High-Resolution Remote Sensing Images. *Remote Sens.* **2021**, *13*, 2594. [CrossRef]
- Weng, Y.; Li, Z.; Tang, G.; Wang, Y. OCNet-Based Water Body Extraction from Remote Sensing Images. *Water* **2023**, *15*, 3557. [CrossRef]

11. Li, L.; Yan, Z.; Shen, Q.; Cheng, G.; Gao, L.; Zhang, B. Water Body Extraction from Very High Spatial Resolution Remote Sensing Data Based on Fully Convolutional Networks. *Remote Sens.* **2019**, *11*, 1162. [CrossRef]
12. Isikdogan, F.; Bovik, A.C.; Passalacqua, P. Surface Water Mapping by Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4909–4918. [CrossRef]
13. Liao, D.; Sun, J.; Deng, Z.; Zhao, Y.; Zhang, J.; Ou, D. A Lightweight Network for Water Body Segmentation in Agricultural Remote Sensing Using Learnable Kalman Filters and Attention Mechanisms. *Appl. Sci.* **2025**, *15*, 6292. [CrossRef]
14. Cao, H.; Tian, Y.; Liu, Y.; Wang, R. Water body extraction from high spatial resolution remote sensing images based on enhanced U-Net and multi-scale information fusion. *Sci. Rep.* **2024**, *14*, 16132. [CrossRef] [PubMed]
15. Li, Y.; Zhou, P.; Wang, Y.; Li, X.; Zhang, Y.; Li, X. Deep Learning Small Water Body Mapping by Transfer Learning from Sentinel-2 to PlanetScope. *Remote Sens.* **2025**, *17*, 2738. [CrossRef]
16. Ngo, P.L.; Pham, V.H.; Bui, N.L.; Phan, H.A.T.; Vo, H.B.; Velavan, T.P.; Tran, D.K. Detection of small water bodies for vector control using deep learning on multispectral imagery from unmanned aerial vehicles. *Discov. Artif. Intell.* **2025**, *5*, 170. [CrossRef]
17. Weng, Z.; Li, Q.; Zheng, Z.; Wang, L. SCR-Net: A Dual-Channel Water Body Extraction Model Based on Multi-Spectral Remote Sensing Imagery—A Case Study of Daihai Lake, China. *Sensors* **2025**, *25*, 763. [CrossRef]
18. Yuan, K.; Zhuang, X.; Schaefer, G.; Feng, J.; Guan, L.; Fang, H. Deep-Learning-Based Multispectral Satellite Image Segmentation for Water Body Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7422–7434. [CrossRef]
19. Hu, H.; He, Z.; Zheng, H. An Algorithm for Multispectral Water Body Detection in Complex Environments. *J. Beijing Univ. Aeronaut. Astronaut.* **2025**, early access (In Chinese) [CrossRef]
20. Yang, S.; Wang, L.; Yuan, Y.; Fan, L.; Wu, Y.; Sun, W.; Yang, G. Recognition of Small Water Bodies under Complex Terrain Based on SAR and Optical Image Fusion Algorithm. *Sci. Total Environ.* **2024**, *946*, 174329. [CrossRef]
21. Wang, R.; Zhang, C.; Chen, C.; Hao, H.; Li, W.; Jiao, L. A Multi-Modality Fusion and Gated Multi-Filter U-Net for Water Area Segmentation in Remote Sensing. *Remote Sens.* **2024**, *16*, 419. [CrossRef]
22. Song, W.; Zhao, Y.; Tu, J.; Chen, M.; Xie, Y.; Cui, X. A Visual Attention-Guided Approach for Concrete Crack Detection in Complex Environments. *Eng. Appl. Artif. Intell.* **2026**, *173*, 114439. [CrossRef]
23. Zhou, Z.; Li, S.; Wu, W.; Guo, W.; Li, X.; Xia, G.; Zhao, Z. NaSC-TG2 (Tiangong-2 Remote Sensing Image Natural Scene Classification Dataset), v1.0; National Basic Science Data Center (NBSDC). 2021. Available online: <https://cstr.cn/CSTR:16666.11.nbsdc.tfpbwtqf> (accessed on 1 January 2026).
24. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [CrossRef]
25. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder–Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
27. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder–Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818. [CrossRef]
28. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019.
29. Loshchilov, I.; Hutter, F. SGDR: Stochastic Gradient Descent with Warm Restarts. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017.
30. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [CrossRef]
31. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890. [CrossRef]
32. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241. [CrossRef]
33. Fu, Y.; Lou, M.; Yu, Y. SegMAN: Omni-scale Context Modeling with State Space Models and Local Attention for Semantic Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 11–15 June 2025; pp. 19077–19087.
34. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. In Proceedings of the Computer Vision—ECCV 2022 Workshops, Tel Aviv, Israel, 23–27 October 2022; pp. 205–218. [CrossRef]

35. Dai, Y.; Giesecke, F.; Oehmcke, S.; Wu, Y.; Barnard, M.; Xing, Y. Attentional Feature Fusion. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Las Vegas, NV, USA, 5–9 January 2021; pp. 3559–3568. [CrossRef]
36. Kim, J.-H.; Jun, J.; Zhang, B.-T. Bilinear Attention Networks. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Montréal, QC, Canada, 3–8 December 2018; pp. 1583–1593.
37. Fang, Q.Y.; Wang, Z.K. Cross-Modality Attentive Feature Fusion for Object Detection in Multispectral Remote Sensing Imagery. *Pattern Recognit.* **2022**, *130*, 108786. [CrossRef]
38. Ates, G.C.; Mohan, P.; Celik, E. Dual Cross-Attention for Medical Image Segmentation. *Eng. Appl. Artif. Intell.* **2023**, *126*, 107139. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## Article

# Hybrid Ensemble Learning for TWSA Prediction in Water-Stressed Regions: A Case Study from Casablanca–Settat Region, Morocco

Youssef Laalaoui <sup>1,\*</sup>, Naïma El Assaoui <sup>1</sup>, Oumaima Ouahine <sup>1</sup>, Thanh Thi Nguyen <sup>2</sup> and Ahmed M. Saqr <sup>3</sup>

<sup>1</sup> Geosciences, Water & Environment Laboratory, Earth Sciences Department, Faculty of Sciences, Mohammed V University in Rabat, Rabat 1014, Morocco; nelassaoui@gmail.com (N.E.A.); ouahineoumama59@gmail.com (O.O.)

<sup>2</sup> Organic Plant Production and Agroecosystems Research in the Tropics and Subtropics, Faculty of Organic Agricultural Sciences, University of Kassel, 37213 Witzenhausen, Germany; thanh.nguyen@uni-kassel.de

<sup>3</sup> Irrigation and Hydraulics Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt; ahmedsaqr@mans.edu.eg

\* Correspondence: laalaoui.youssef@gmail.com

## Abstract

A hybrid machine learning framework has been developed in this study to estimate Terrestrial Water Storage Anomalies (TWSA) in Morocco's Casablanca–Settat region, which faces serious groundwater stress due to rapid urbanization, intensive agriculture, and climate variability. In this study, TWSA is used as an integrated proxy for groundwater-related storage changes, while acknowledging that it also includes contributions from soil moisture and surface water. The approach combines satellite-based observations from the Gravity Recovery and Climate Experiment (GRACE) and GRACE Follow-On (GRACE-FO) with key environmental indicators such as rainfall, evapotranspiration, and land use data to track changes in groundwater availability with improved spatial detail. After preprocessing the data through feature selection, normalization, and outlier handling, the model applies six base learners, i.e., Huber regressor, automatic relevance determination regression, kernel ridge, long short-term memory, k-nearest neighbors, and gradient boosting. Their predictions are aggregated using a random forest meta-learner to improve accuracy and stability. The ensemble achieved strong results, with a root mean square error of 0.13, a mean absolute error of 0.108, and a determination coefficient of 0.97—far better than single-model baselines—based on a temporally independent train-test split. Spatial analysis highlighted clear patterns of groundwater depletion linked to land cover and usage. These results can guide targeted aquifer recharge efforts, drought response planning, and smarter irrigation management. The model also aligns with national goals under Morocco's water sustainability initiatives and can be adapted for use in other regions with similar environmental challenges.

**Keywords:** groundwater prediction; hybrid machine learning; Terrestrial Water Storage Anomaly (TWSA); GRACE satellite data; random forest; water resource management

## 1. Introduction

Groundwater constitutes approximately 30% of the global freshwater supply and plays a critical role in supporting agriculture, maintaining ecological balance, and meeting human consumption needs, particularly in arid and semi-arid zones [1]. However, escalating pressures



from climate change, population growth, and Land Use/Land Cover (LULC) transformations are threatening groundwater sustainability. Effective management strategies require robust systems capable of capturing spatiotemporal fluctuations in groundwater levels [2,3].

Satellite-based missions such as Gravity Recovery and Climate Experiment (GRACE) and GRACE Follow-On (GRACE-FO) have significantly advanced large-scale monitoring of Terrestrial Water Storage Anomalies (TWSA), offering insights into the combined behavior of groundwater, soil moisture, and surface water. Despite their global utility, the coarse spatial resolution of these datasets limits their applicability in localized groundwater planning. To bridge this gap, machine learning (ML) techniques are increasingly utilized to extract meaningful correlations between hydro-climatic variables and groundwater dynamics. While single-model approaches often lack the flexibility to generalize across heterogeneous geographies, hybrid ML frameworks integrate multiple algorithms and deep learning architectures. Architectures such as Long Short-Term Memory (LSTM) have shown promising results in capturing both short-term variability and long-term trends, thereby enhancing the accuracy and operational relevance of groundwater predictions.

### *1.1. Relevance to the Casablanca–Settat Region*

The Casablanca–Settat region of Morocco, positioned along the Atlantic coast (Figure 1), represents a strategic and economically vital zone that faces mounting groundwater challenges. It hosts Morocco’s largest urban hub (Casablanca) and extensive agricultural hinterlands (e.g., Berrechid, Settat, and El Jadida), resulting in complex and competing water demands. While cities rely heavily on groundwater for industrial and domestic use, adjacent farmlands extract substantial volumes for irrigation, especially during the long dry seasons. This imbalance, combined with irregular rainfall, high evapotranspiration, and coastal saline intrusion, has led to widespread aquifer stress and declining groundwater tables.

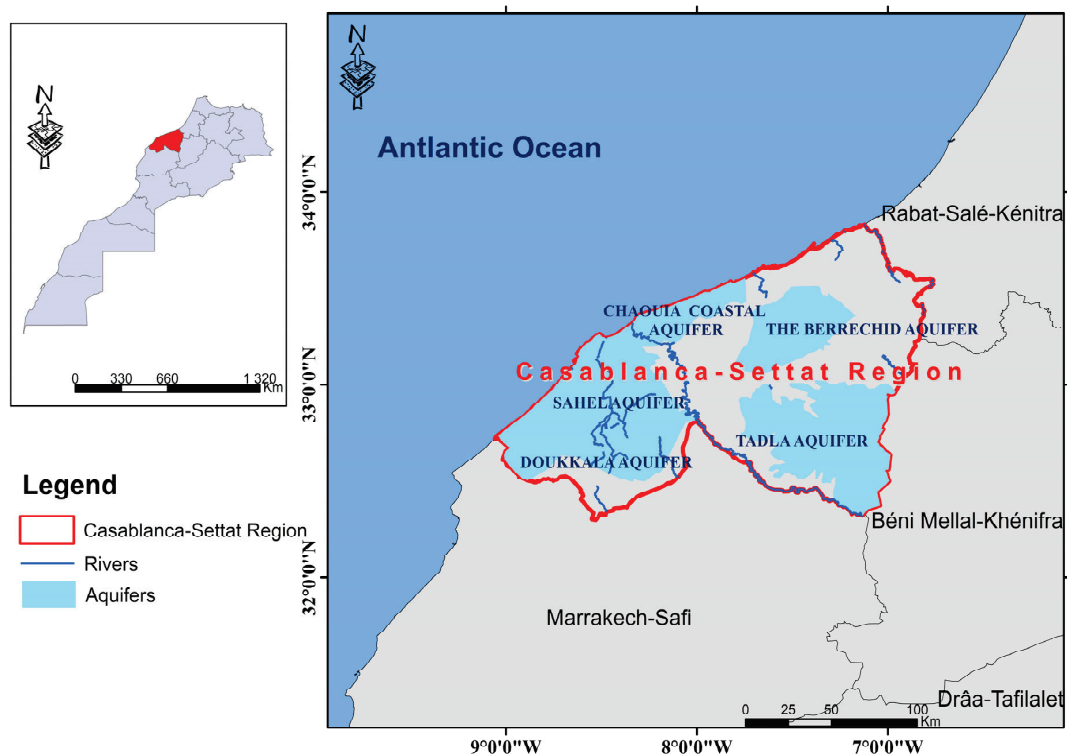
Traditional monitoring approaches in the region suffer from low spatial resolution and infrequent data updates, making it difficult to detect local depletion trends or inform timely policy actions. In this context, the proposed hybrid ML model offers a robust solution by integrating satellite-derived Terrestrial Water Storage Anomaly (TWSA) data with regional hydro-climatic and LULC variables. This enables high-resolution, spatially explicit forecasting of Ground Water Storage (GWS) anomalies across diverse landscapes and time periods. The model helps identify high-risk zones such as peri-urban agricultural belts prone to over-extraction and inland regions with limited natural recharge. For instance, it can flag potential depletion near Berrechid or highlight increased storage following rainfall in coastal El Jadida. These predictions enable evidence-based decision-making for aquifer recharge projects, groundwater licensing, irrigation scheduling, and drought preparedness. This approach aligns with Morocco’s national water strategy and supports Sustainable Development Goal (SDG) targets related to clean water access and climate adaptation, making it a timely and regionally impactful contribution to groundwater governance.

### *1.2. Aims and Objectives*

This study aims to build a robust hybrid ML model for predicting groundwater anomalies in the Casablanca–Settat region. The core objectives are as follows:

- **Data Integration and Model Development:** To compile a multi-source dataset (e.g., GRACE/GRACE-FO TWSA and LULC classifications) and construct a hybrid ensemble model combining Huber Regressor (HR), Automatic Relevance Determination (ARD) regression, Kernel Ridge (KR), K-Nearest Neighbors (KNN), Gradient Boosting (GB), and LSTM, with Random Forest (RF) as the final aggregator.

- **Model Evaluation:** To evaluate the hybrid model's performance in predicting TWSA using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and determination coefficient ( $R^2$ ), and to compare its results against individual baseline regression models trained on the same input features.
- **Insight Extraction:** To generate region-specific insights—such as LULC impact, spatial anomaly clustering, and extreme groundwater zones—to support data-driven hydrological planning.



**Figure 1.** Location of the Casablanca–Settat region, Morocco. Source: Aquifer boundaries digitized by the authors based on the official regional monograph of Casablanca–Settat (Ministry of Interior, Morocco).

## 2. Literature Review

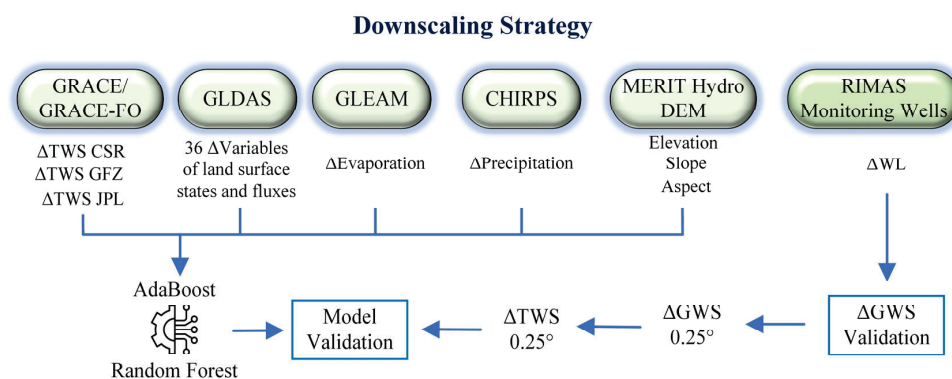
This research has explored the integration of GRACE/GRACE-FO satellite observations with ML and hydrological modeling to better capture spatiotemporal dynamics in GWS. The following literature is categorized into four thematic areas based on methodological approaches and geographic relevance.

### 2.1. ML-Based Downscaling of GRACE Data

ML models have been widely applied to refine the spatial resolution of GRACE observations. An XGBoost-based model was developed by the authors in [4], which was used for the Upper Indus Plain, utilizing the Famine Early Warning Systems Network Land Data Assimilation System (FLDAS) and terrain features to enhance GWS anomaly (GWSA) detection. But notably, this model lacked Normalized Difference Vegetation Index (NDVI) and LULC inputs, limiting ecological interpretation. A similar three-stage framework built in [5], integrating RF and XGBoost, achieved an  $R^2$  of 0.8674 but did not model spatial geolocation features such as latitude and longitude.

In Hamou-Ali [6], the authors worked on the dataset of the Huang–Huai–Hai (HJH) Plain of China and used a 4 km downscaling model utilizing RF, XGBoost, and LightGBM

along with TerraClimate variables. They achieved up to 23% improvement in correlation. Palmer Drought Severity Index (PDSI) and Actual EvapoTranspiration (AET) were identified as key predictors via SHapley Additive exPlanations (SHAP) analysis. In Morocco, an RF model incorporating NDVI and Land Surface Temperature (LST) inputs successfully generated 1 km TWSA maps validated against 63% of wells, although it excluded runoff dynamics [7]. Moreover, the Amazon Basin saw the application of RF and AdaBoost for downscaling from 1° to 0.25°, capturing a net TWSA gain of +22.24 km<sup>3</sup>/year while revealing sub-basin heterogeneity. However, it focused primarily on humid regions, limiting broader applicability [8]. Figure 2 illustrates the general GRACE downscaling methodology used in the referenced study, shown here for explanatory purposes.



**Figure 2.** GRACE downscaling methodology (reproduced from [8] for illustrative purposes). The diagram summarizes the main steps, including coarse-resolution GRACE/GRACE-FO TWSA data as input, integration of auxiliary high-resolution hydro-climatic and land-surface variables, supervised learning-based disaggregation, and generation of spatially refined TWSA outputs.

Similarly, regression-based downscaling using a GWR model in the Songhua River Basin achieved high accuracy (Nash–Sutcliffe Efficiency (NSE) = 0.989), though generalization to agricultural LULC transitions remained unaddressed [9].

## 2.2. Temporal Forecasting and Gap-Filling of GRACE Time Series

Efforts to resolve temporal discontinuities in GRACE datasets have driven hybrid modeling innovation. A Conventional Neural Network–Long Short-term Memory (CNN–LSTM) model developed and trained in [10], across Indian basins, achieved an  $R^2$  score of 92%, and a Mean Squared Error (MSE) = 0.91, offering reliable GWSA reconstruction. Notably, this work lacked LULC awareness. For the Tehran-based database, a non-stationary LSTM model was integrated by the authors in [11]. They used ERA5 and Global Land Data Assimilation System (GLDAS) data to predict monthly TWSA with an  $R^2$  = 0.93 and RMSE = 4.75. But on the downside, the spatial heterogeneity was not modeled in this work. For Shandong Province, LSTM, Autoregressive Moving Average (ARMA), Singular Spectrum Analysis (SSA), and Support Vector Machine (SVM) models were tested using GRACE and WaterGAP Global Hydrology Model (WGHM) inputs; SVM yielded the best RMSE = 4.42, but the study did not explore spatial interpolation [12]. For the South Korea-based dataset, CNN–LSTM networks outperformed standalone LSTM models in predicting TWSA based on GLDAS, Tropical Rainfall Measuring Mission (TRMM), and Landsat inputs, although localized feedback processes were underrepresented [13].

A combination of CNN, LSTM, and CNN–LSTM was used for the Ganga Basin-based dataset, with LSTM showing superior performance (RMSE = 0.292), yet ecological and LULC variables were excluded here as well [14]. Meanwhile, a feature attribution study

across 254 basins used XGBoost and SHAP values to distinguish spatial versus temporal importance, offering insight. But this work also lacked spatial downscaling [15].

### *2.3. Groundwater Trend Assessment, Validation, and Sustainability Monitoring*

To assess long-term GWS dynamics, GRACE data have been validated using in situ and quality-based indicators. In Iran's Yazd region, a strong inverse correlation of  $-0.79$  between GRACE-derived GWS and electrical conductivity levels confirmed aquifer depletion. Notably, spatial modeling was not performed in this work [16]. A large-scale analysis across India detected GWS losses as high as  $-0.75$  cm/year in the Indo-Gangetic plain using Mann–Kendall (MK) and Sen's Linear Regression (SLR) statistical tests, but these are limited by GRACE's resolution [17].

A 21-year analysis across the Middle East used GRACE to derive a sustainability index, revealing an alarming 5.93 mm/year GWSA depletion and severe stress in 59% of basins; however, the study lacked predictive modeling or input feature disaggregation [18]. In North China, integrated GRACE, groundwater extraction (GWE), and in situ observations were used to assess South-to-North Water Transfer Project (SNWTP) impacts and sustainability indices without incorporating spatial downscaling [19]. Validation against groundwater modeling and observational wells in five Iranian provinces showed GRACE-FO outperformed GRACE, although it remained coarse in resolution [20]. In the Central Kabul River Basin (CKRB) spanning Pakistan and Afghanistan, long-term depletion rates of  $-7.83$  mm/month were observed via GRACE-FO and GLDAS data, though validation yielded only moderate  $R^2$  (0.47) with well data [21].

### *2.4. Integrative and Regional Approaches for GRACE Enhancement*

More recent approaches adopt region-specific frameworks integrating GRACE data with ancillary inputs for enhanced monitoring. In Türkiye's Kizilirmak Basin, integration of FLDAS data with hydrological indices such as the Warm Spell Duration Index (WSDI) successfully detected localized drought periods. Yet, the model did not incorporate NDVI or LULC predictors [22]. Meanwhile, an integrated groundwater sustainability index for North China was developed using GRACE, GRACE-FO, GWE, and hydrological indicators to evaluate SNWTP impacts. However, the study did not apply ML for enhanced spatial granularity [23]. Another SHAP-based feature analysis using XGBoost across Iranian basins provided interpretability into spatial versus temporal drivers, although predictive downscaling and LULC integration were not implemented [24].

### *2.5. Regional Perspectives on Water Challenges and Modeling Approaches in Morocco*

Recent Moroccan studies reflect a growing urgency to address the multifaceted environmental stresses affecting water resources, particularly in the semi-arid and agriculturally critical regions. One such research study [25] highlighted the severity of soil degradation in the Tamdoust watershed, where average erosion rates reach up to 800 T/ha/year, largely driven by topographic and rainfall factors. Their integration of the Revised Universal Soil Loss Equation (RUSLE) with Geographic Information System (GIS) and remote sensing offered thematic maps to guide erosion control in threatened zones, demonstrating how geospatial modeling supports environmental decision-making. Meanwhile, two-dimensional modeling efforts in the Bouregreg estuary [26] have underscored the benefits of hydrodynamic simulations for managing sediment transport and water circulation. These models showed that dredging operations can reduce erosion impacts by approximately 30%, improving both ecological balance and structural resilience in coastal estuarine systems.

In flood-prone Mediterranean regions, deterministic hydrological models have been used with success, achieving an  $R^2$  of 0.81 and Nash–Sutcliffe performance indices averaging 0.69, although initial hydrographs remain difficult to predict [27]. Urbanization's encroachment on natural water systems forms another vital axis of Moroccan water research. Another study emphasized how urban growth exacerbates desertification and groundwater depletion, calling for the integration of remote sensing, geophysical investigations, and water quality assessments to quantify urban runoff contamination and aquifer drawdown [28]. Complementing these technical studies, a bibliometric analysis of Moroccan water research revealed that, despite substantial international collaboration—particularly with France and Spain—there is a relative lack of emphasis on the development of advanced GIS-based solutions for groundwater mapping, pollution tracking, and agricultural planning [29].

### 2.6. Research Gap

While recent research has made great strides in applying GRACE/GRACE-FO data and ML to groundwater monitoring, there are still some clear gaps. Many existing models tend to focus either on improving spatial resolution or predicting temporal trends—but rarely both together. Even in well-known studies that use strong algorithms like RF or LSTM, key environmental factors such as LULC, NDVI [30–32], or evapotranspiration are often left out. This makes it harder to understand how surface conditions actually relate to what's happening underground. Moreover, even when models perform well, the insights they generate are not always transformed into region-specific maps or tools that can directly help policymakers on the ground.

In the case of Morocco, this gap becomes even more noticeable. Although researchers have performed important work on issues like soil erosion, urban sprawl, and flood modeling, few have developed models that combine satellite data with advanced ML to predict groundwater changes in a specific region. For example, in areas like Casablanca–Settat, where both cities and farms are putting pressure on water resources, there is a real need for robust tools that can not only forecast groundwater trends but also show where the problems are emerging [33,34]. Thus, this study addresses this by building a hybrid model that brings together satellite data, climate records, and LULC details to produce accurate, high-resolution predictions. The goal is to turn complex data into practical insights that support smarter, more sustainable groundwater management.

## 3. Methodology

The methodology adopted in this study follows a structured framework. This framework is designed to predict TWSA across the Casablanca–Settat region. TWSA serves as a proxy for groundwater storage anomalies. The process begins with the acquisition and preparation of satellite and environmental datasets. It continues with data cleaning, normalization, and feature engineering [33,35].

### 3.1. Dataset Collection and Pre-Processing

Publishing after 2020 reflects an enhanced comprehension of sustainable development. The primary variable analyzed in this study is the TWSA, which reflects changes in the total volume of water—including groundwater, soil moisture, and surface water—within a specific region. TWSA values were sourced from NASA's GRACE and GRACE-FO satellite missions, which use gravity measurements to monitor large-scale hydrological shifts globally. These datasets were accessed through the publicly available GRACE portal (<https://grace.jpl.nasa.gov>), accessed on 10 September 2025. The GRACE and GRACE-FO missions are separated by a known observational gap between mid-2017 and mid-



2018. In this study, this temporal discontinuity was explicitly accounted for during data preprocessing by treating the two missions as a continuous but segmented time series, without imposing artificial interpolation across the gap. Model training and evaluation were therefore conducted only on periods with available observations, preserving the physical integrity of the time series. To build a reliable prediction model, the following environmental and geographic features were incorporated [6]:

- LULC: Categorical data indicating surface types (e.g., urban, forest, and agricultural), later one-hot encoded to reflect their influence on local water retention.
- Climate and Hydrological Indicators: These include Precipitation\_mm, Evapotranspiration\_mm, Temperature\_C, Soil\_Moisture\_mm, and Runoff\_mm, sourced from satellite-based datasets like GLDAS. These indicators capture water balance and hydrological dynamics.
- NDVI: Representing vegetation health and land cover variation, it is included as a key predictor due to its influence on evapotranspiration and soil water retention.
- Topographic elevation data were obtained from a Digital Elevation Model (DEM) and used as a static predictor to account for spatial variability in groundwater recharge and storage conditions.
- Geospatial Attributes: Latitude and Longitude are retained for spatial mapping, while Spatial\_Resolution\_km and Downscale\_Method provide metadata for spatial fidelity.
- All spatial data points were mapped to a grid covering the Casablanca–Settat region and resampled to monthly intervals. Each data record represents a unique combination of location and time, along with its associated environmental features and corresponding TWSA values. While the spatial domain consists of approximately 200 grid cells, the final dataset contains multiple spatio-temporal samples derived from monthly observations across all grid cells.

For each monthly time step, the GRACE-derived TWSA value corresponding to the parent coarse-resolution grid cell was assigned to all fine-resolution grid cells contained within that footprint, while spatial differentiation was learned through the associated high-resolution auxiliary predictors. The combination of remotely sensed water data with climate and LULC variables enables the model to capture complex interactions affecting groundwater dynamics. Using open-access and globally trusted sources ensures transparency and reproducibility of the work.

### 3.2. Spatial Downscaling Strategy

The native spatial resolution of GRACE and GRACE-FO satellite data is approximately  $1^\circ \times 1^\circ$ , which translates to a surface area of around  $110 \text{ km} \times 110 \text{ km} \approx 12,100 \text{ km}^2$  per grid cell near the equator, shown in Equation (1) [36].

$$110 \text{ km} \times 110 \text{ km} \approx 12,100 \text{ km}^2 \quad (1)$$

However, this resolution is too coarse for regional decision-making in water-scarce zones such as Casablanca–Settat. To make the groundwater anomaly predictions actionable at a local level, a spatial downscaling framework was employed to translate GRACE-derived TWSA data into a  $10 \text{ km} \times 10 \text{ km}$  resolution, resulting in  $100 \text{ km}^2$  per grid cell. The total study area spans  $\sim 20,000 \text{ km}^2$  and was subdivided into 200 downscaled grid cells, calculated in Equation (2).

$$\frac{20,000 \text{ km}^2}{100 \text{ km}^2} = 200 \text{ grid cells} \quad (2)$$

To achieve this, a supervised regression-based disaggregation approach was adopted. Coarse TWSA values from GRACE were spatially mapped and redistributed using high-resolution auxiliary variables, including precipitation, evapotranspiration, soil moisture, elevation, and LULC. These variables were resampled or natively available at or near the target resolution (10 km), sourced from the Moderate Resolution Imaging Spectrometer (MODIS) dataset described in the next section [37]. Each high-resolution covariate was assigned to one or more of the 200 grid cells using spatial mapping techniques, enabling fine-scale feature extraction. The GRACE-based coarse TWSA values were then used as labels in a training dataset, where each downscaled grid point became a sample, and the high-resolution variables became the predictors [38,39]. Although the same coarse-resolution GRACE TWSA value is assigned to multiple fine-resolution grid cells within a given month, these samples remain distinct due to their different auxiliary predictor values. During model evaluation, temporal independence is preserved by applying a chronological train–test split, ensuring that all samples from the same time period are assigned exclusively to either the training or testing subset. This prevents information leakage despite the presence of repeated labels within individual time steps [40].

A hybrid ML model was used to learn the non-linear mapping as shown in the following Equation (3):

$$\text{TWSA}_{10\text{ km}} = f(\text{P}, \text{ET}, \text{SM}, \text{LULC}, \text{Elev}) \quad (3)$$

where P is precipitation, ET is evapotranspiration, SM is soil moisture, and  $f$  represents the trained hybrid model function.

The resulting model generated 200 localized TWSA estimates, each representing a  $10\text{ km} \times 10\text{ km}$  ( $100\text{ km}^2$ ) spatial unit across the Casablanca–Settat region, allowing more precise monitoring and forecasting of groundwater fluctuations.

### 3.3. Data Pre-Processing

Preparing the dataset for hybrid modeling is an essential step that directly affects the accuracy and stability of the model. This process involves several key stages: data cleaning, encoding, normalization, reshaping for sequence models, and constructing feature interactions. These steps collectively ensure that all the input variables are numerically coherent and aligned with the expectations of the downstream algorithms.

The raw data may contain missing values in the input datasets, primarily arising from occasional gaps in satellite retrievals, reanalysis assimilation issues, or temporal inconsistencies during data aggregation, rather than from sensor malfunction. Given the relatively low proportion of missing data and the use of monthly aggregated hydroclimatic variables, mean imputation was applied as a simple and stable strategy to preserve temporal continuity without introducing artificial variability. These missing values are addressed using mean imputation, as depicted in Equation (4) [41–43].

$$x_i = \frac{1}{n} \sum_{j=1}^n x_{ij} \quad (4)$$

where  $x_i$  is the imputed value for feature  $i$ ,  $x_{ij}$  is the  $j$ th value of feature  $i$ , and  $n$  is the total number of samples.

Outliers are handled by clipping values beyond the  $1.5 \times$  interquartile range (IQR), ensuring the model is not biased by extreme data points, as shown in Equations (5) and (6), respectively.

$$\text{IQR} = Q_3 - Q_1 \quad (5)$$

$$x_{\text{clipped}} = \min(\max(x, Q_1 - 1.5 \cdot \text{IQR}), Q_3 + 1.5 \cdot \text{IQR}) \quad (6)$$

where  $x$  is the original value,  $Q_1$  and  $Q_3$  are the 25th and 75th percentiles, and  $x_{\text{clipped}}$  is the adjusted value within the IQR threshold.

The IQR-based clipping defined in Equations (5) and (6) was applied only to continuous hydroclimatic predictors (e.g., precipitation, evapotranspiration, soil moisture, and vegetation indices), while categorical variables such as LULC classes were excluded. Thresholds were examined to ensure that physically plausible extreme values were preserved and that only statistically anomalous outliers were limited. Further, the LULC variable is a categorical input. It is encoded using one-hot encoding, where a feature  $x(i)$  with  $k$  distinct categories is transformed into a  $k$ -dimensional binary vector. To align the scale of input variables, min-max normalization is applied as shown in Equation (7). It rescales each value  $x$  to a range of  $[0, 1]$ :

$$x_{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (7)$$

where  $x$  is the original feature value, and  $x_{\text{norm}}$  is its normalized form within  $[0, 1]$  based on its column's min and max. This improves model convergence and stability, especially for gradient-based learners and neural networks [44–46].

#### 3.4. Baseline Models and Hybrid Model Architecture

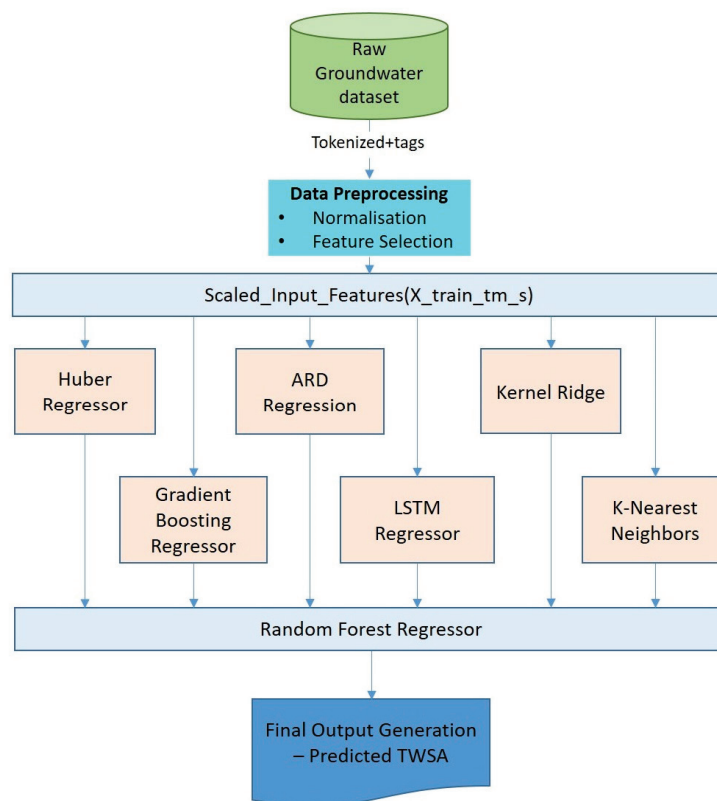
To benchmark the performance of the proposed hybrid model, the following four baseline regression algorithms were implemented: linear regression, ridge regression, decision tree, and gradient boosting base. Each model was independently trained and evaluated using the same feature set to ensure consistency. These baselines helped in establishing a reference point for comparison and in understanding the individual contribution of different learning paradigms [33]. The hybrid regression model integrates several ML algorithms that operate on different statistical principles to improve the robustness and accuracy of groundwater storage prediction. By combining linear, non-linear, tree-based, and temporal models, this architecture captures a wide variety of data patterns and relationships. The final TWSA prediction is obtained using a stacking ensemble strategy, where the outputs of multiple base learners are combined through an RF meta-learner [47].

$$\hat{y} = f_{\text{RF}}(\hat{y}_1(x), \hat{y}_2(x), \dots, \hat{y}_k(x)) \quad (8)$$

where  $\hat{y}$  denotes the final predicted TWSA,  $\hat{y}_k(x)$  represents the prediction of the  $k$ -th base learner for input feature vector  $x$ ,  $K = 6$ ,  $K$  is the number of base learners (HR, ARD, KR, KNN, GB, and LSTM), and  $f_{\text{RF}}(\cdot)$  denotes the RF meta-learner. In this framework, the meta-learner learns a non-linear mapping between the base-model predictions and the target TWSA, allowing adaptive combination of individual model outputs rather than a fixed or voting-based aggregation.

The proposed hybrid model pipeline shown in Figure 3 is designed to leverage a diverse ensemble of ML algorithms to accurately predict TWSA based on groundwater-related features. The pipeline begins with a raw, open-source groundwater dataset comprising various hydro-meteorological indicators [48]. This dataset is first subjected to semantic tokenization and contextual tagging, enabling a structured mapping of variables such as precipitation, evapotranspiration, and soil moisture with their corresponding timestamps and spatial characteristics. This annotated data then undergoes a preprocessing phase where the normalization process is used to scale all the data such that all features are in the same range; thus, the high values of the features have a greater numerical range. Further,

the correlation-based filtering and reduction in multicollinearity are employed to choose relevant features that can be used to strengthen the generalizability of the model [38,39,49].



**Figure 3.** Architecture of the proposed stacked ensemble regression framework, in which predictions generated by the base learners (HR, ARD, KR, KNN, GB, and LSTM) are stacked and provided as input features to a Random Forest meta-learner to produce the final TWSA estimate.

After that, the scaled input feature set is introduced into a hybrid ensemble with six different models, including HR, ARD, KR, KNN, LSTM, and GB. All these models add their own mathematical lens to the data distribution: ARD adds a Bayesian view of automatic determination of relevance, while LSTM adds the dynamics of temporal dependencies in between sequences. The process of outputs of the six learners, in the embodiment of regional and global patterns, as well as linear and nonlinear features, is sequentially transferred to an RF Regressor. The RF acts as a meta-learner in a stacked ensemble regression framework, learning a non-linear mapping between the base learners' predictions and the target TWSA values [50,51]. The last result is the statistical prediction of the TWSA, and it is learned as a non-linear combination of statistical inference, time learning, and stacked ensemble regression.

The importance of each model and its role in the hybrid framework is explained in Table 1.

**Table 1.** Role of each model in the hybrid framework. Model descriptions are based on standard definitions of the corresponding algorithms in the machine learning literature (e.g., scikit-learn documentation and foundational ML tutorials).

| Model | Role in Hybrid Framework                         |
|-------|--------------------------------------------------|
| HR    | Handles data outliers and stabilizes predictions |
| ARD   | Discovers relevant features for regression       |

Table 1. Cont.

| Model | Role in Hybrid Framework                                     |
|-------|--------------------------------------------------------------|
| KR    | Introduces non-linear mapping using kernel functions         |
| GB    | Refines errors through sequential decision tree learning     |
| KNN   | Adds local instance-based learning                           |
| LSTM  | Captures temporal dependencies and sequential patterns       |
| RF    | Aggregates the final prediction using deep ensemble learning |

### 3.5. Model Training and Evaluation Strategy

Data Partitioning and Validation Approach: All the data was divided into training and test segments with a temporal division of 80:20. The temporal split was performed chronologically, with the earliest 80% of the monthly time series used for model training and the most recent 20% reserved for independent testing, ensuring that no future information leaked into the training process. The 5-fold cross-validation was applied exclusively within the training subset to tune model parameters and assess robustness, while the test subset remained completely unseen until the final evaluation. This method minimizes the variability in the estimation of the model and confidently estimates the stability of the models in unrevealed data sets [52].

For LSTM training, the Adaptive Moment Estimation (Adam) optimizer was employed. This optimizer is a gradient-based algorithm that adaptively adjusts learning rates to ensure efficient and stable neural network training. The input data was reshaped to meet the 3D input format expected by LSTM layers—structured as [samples, time\_steps, features]. A batch size of eight was chosen to promote stable learning, while the number of epochs was limited to three due to computational constraints, with early stopping mechanisms applied to prevent overfitting. Although the maximum number of epochs was set to three to limit overfitting and computational cost, early stopping was included as a safeguard mechanism to terminate training if no improvement in validation loss was observed between successive epochs [53,54]. For this model, temporal ordering was strictly preserved, and the chronological train–test split ensured that model evaluation was performed on future time periods relative to training, thereby preventing temporal information leakage. The RF and GB models were trained using all available CPU cores to maximize scalability and reduce training time. To measure model performance comprehensively, the following evaluation metrics were used:

- MAE: Measures the average magnitude of prediction errors.
- RMSE: Penalizes large errors more heavily, important for hydrological extremes.
- $R^2$  Score: Indicates how well the model explains the variance in the target variable.

This structured training and evaluation pipeline guarantees fairness in comparison across models, stability in prediction, and transparency in model decisions.

## 4. Results

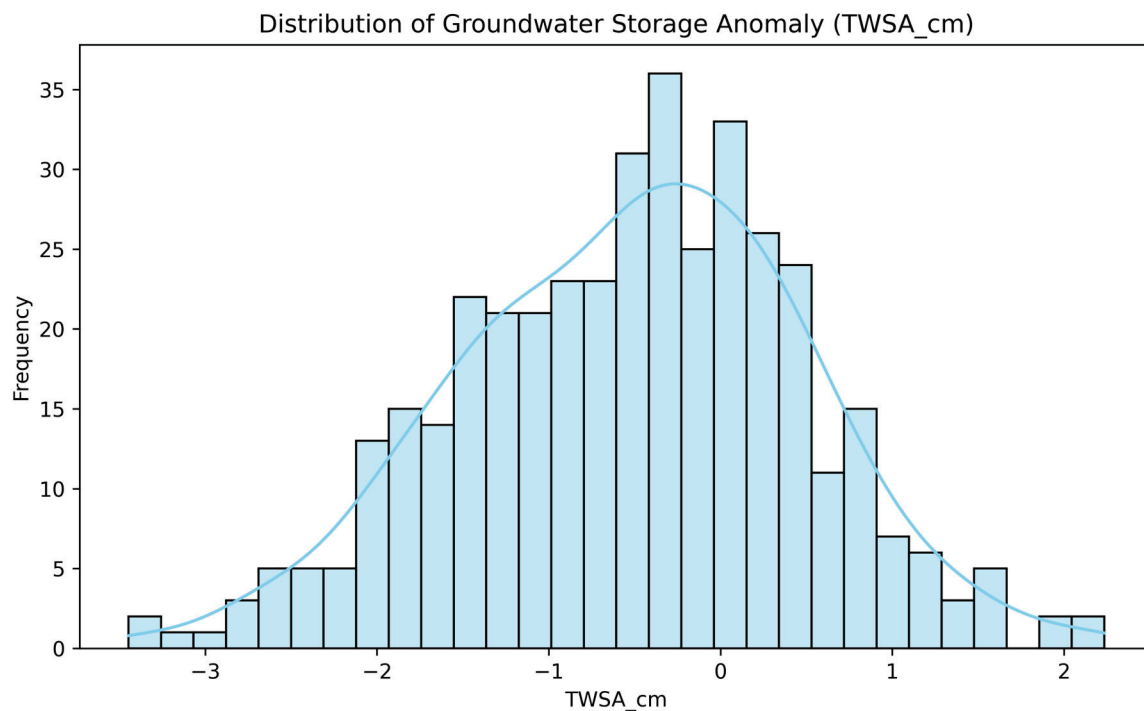
### 4.1. Exploratory Data Analysis

Exploratory Data Analysis (EDA) plays a critical role in understanding the structure, distribution, and interactions within the dataset before model development. In this study, EDA was conducted on the merged satellite, climatic, and LULC dataset to examine the underlying patterns influencing groundwater fluctuations, particularly the TWSA.



#### 4.1.1. Distribution of TWSA

The plot in Figure 4 illustrates the frequency distribution of the TWSA values, range between  $-1$  to  $1$ . It shows that TWSA measurements are approximately normally distributed, with a mild skew toward negative values, indicating that groundwater depletion episodes are more frequent than excess storage periods.



**Figure 4.** Histogram of TWSA values showing a skewed distribution with more frequent groundwater deficits.

#### 4.1.2. Correlation Heatmap of Features

The correlation heatmap (Figure 5) visualizes the pairwise Pearson correlation coefficients among all features in the dataset. The TWSA has highly correlated with hydro-climatic, herein, temperature (0.96) and between then such as soil moisture and runoff (0.98).

#### 4.1.3. Three-Dimensional Scatter Plot (TWSA vs. Latitude vs. Longitude)

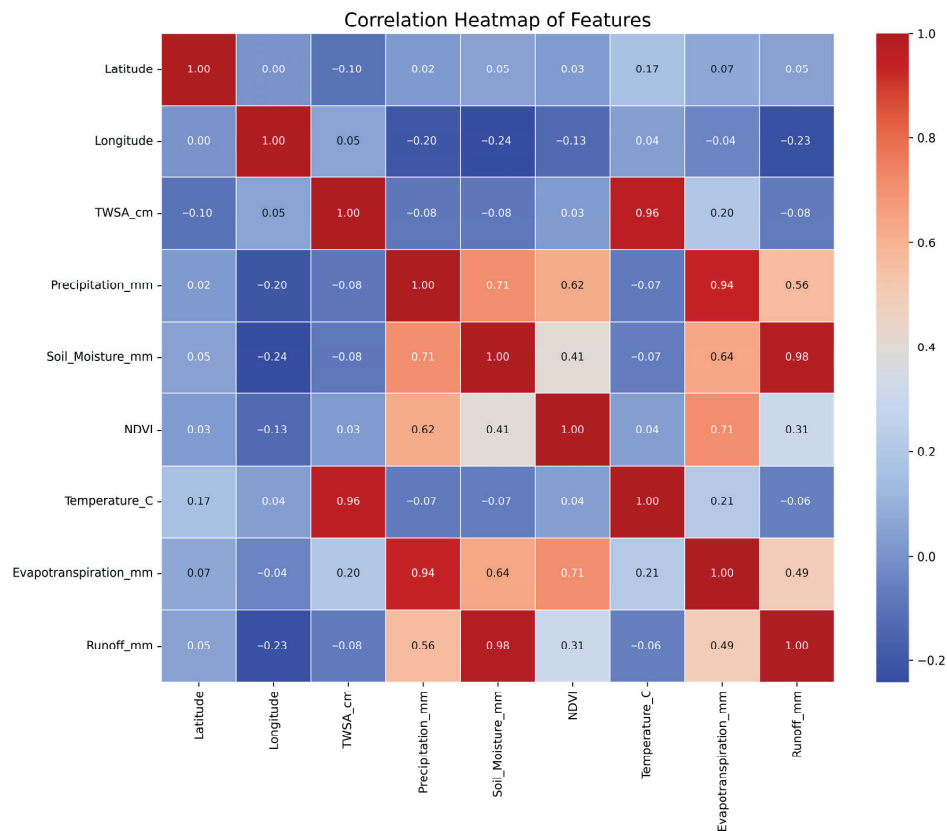
The three-dimensional (3D) scatter plot in Figure 6 provides a spatial-temporal view of TWSA variation across geographic coordinates. This plot is particularly helpful in identifying spatial outliers and regional extremes in groundwater trends.

#### 4.1.4. Pairwise Relationships Between Hydro-Climatic Variables

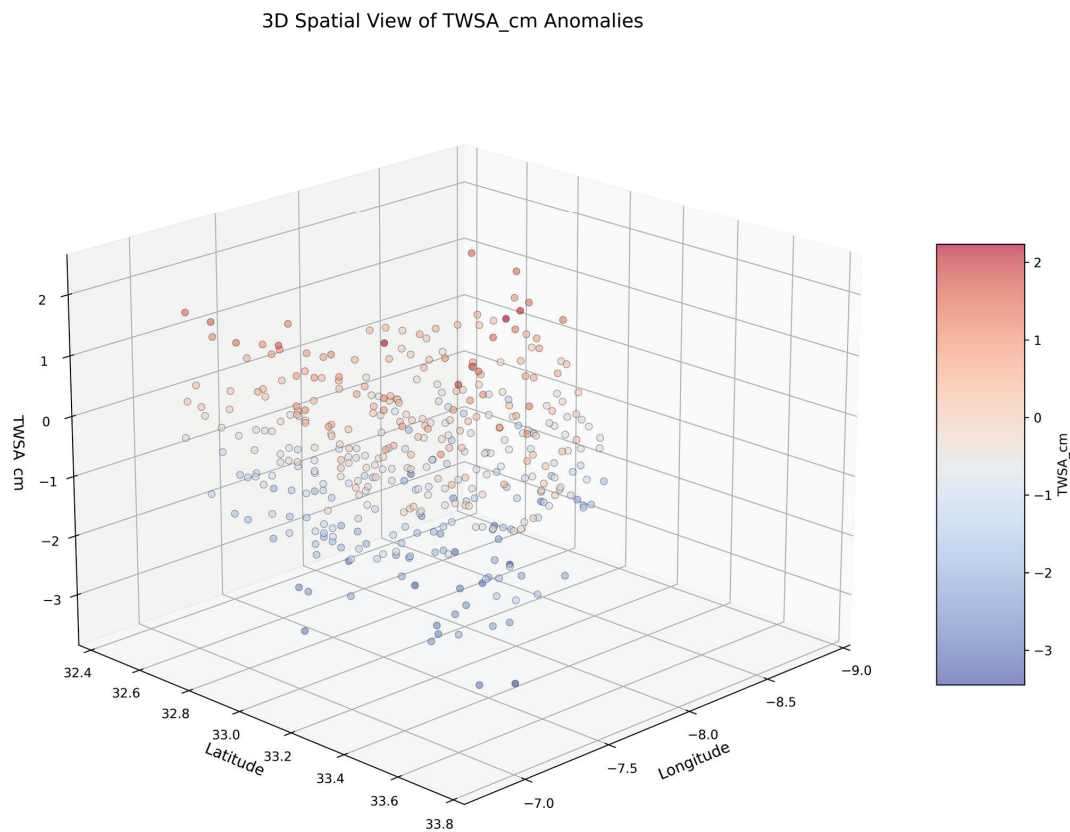
The pairwise plot in Figure 7 visualizes the joint distributions and correlations between TWSA, precipitation, evapotranspiration, temperature, and LULC indices. It reveals linear trends between TWSA and precipitation, as well as non-linear associations with evapotranspiration and land cover. These relationships guide feature selection and model tuning.

#### 4.1.5. Spatial Plot of TWSA\_cm (Bubble Map)

This spatial visualization (plotted in Figure 8) maps TWSA values over the Casablanca–Settat region. It demonstrates geographical disparities in groundwater availability, with coastal and lowland zones showing greater depletion. This helps assess the spatial generalizability of the model.

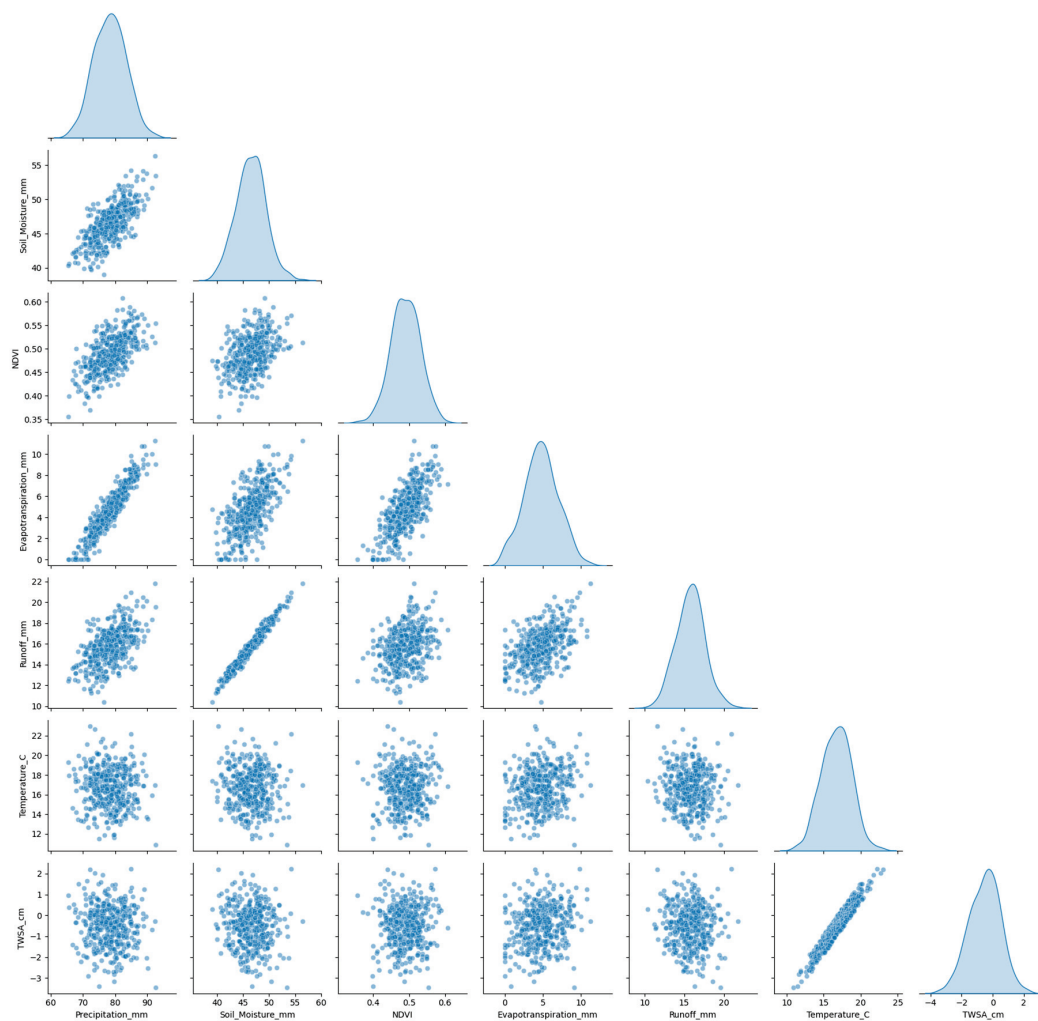


**Figure 5.** Correlation heatmap revealing key relationships between TWSA and hydro-climatic variables.



**Figure 6.** Three-dimensional scatter plot of TWSA against latitude and longitude, revealing spatial patterns.

Pairwise Relationships Between Hydro-Climatic Variables



**Figure 7.** Pairwise plots showing interactions between TWSA and key environmental features.

Together, these visualizations provide a comprehensive understanding of the dataset's statistical and spatial properties.

#### 4.2. Model Evaluation

The baseline models were selected to represent a broad spectrum of statistical learning strategies, including linear, kernel-based, decision-tree, and boosting-based methods. The proposed hybrid ensemble model was evaluated using the same metrics as the baseline models—RMSE, MAE, and  $R^2$  Score [32,55]. Table 2 summarizes the evaluation thus made. A visual comparison of model performance across these metrics is provided in Figure 9.

**Table 2.** Model evaluation comparison.

| Model Type      | Model             | RMSE   | MAE    | $R^2$ Score |
|-----------------|-------------------|--------|--------|-------------|
| Baseline Models | Linear Regression | 0.2682 | 0.2060 | 0.9241      |
|                 | Ridge Regression  | 0.3379 | 0.2601 | 0.8796      |
|                 | Decision Tree     | 0.4206 | 0.3228 | 0.8136      |
|                 | Gradient Boosting | 0.5025 | 0.3678 | 0.7339      |

Table 2. Cont.

| Model Type            | Model                                                | RMSE   | MAE    | R <sup>2</sup> Score |
|-----------------------|------------------------------------------------------|--------|--------|----------------------|
| Proposed Hybrid Model | HR + ARD + GB + KR + KNN + LSTM with RF meta learner | 0.1300 | 0.1080 | 0.9700               |

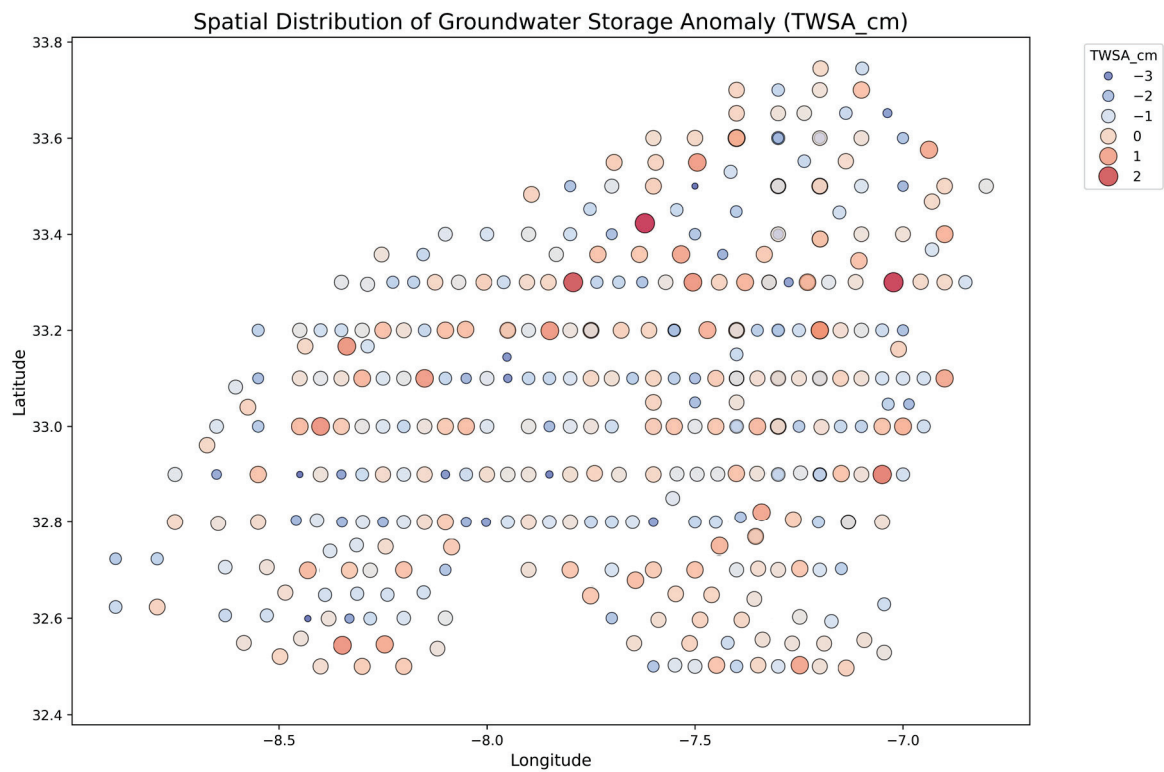
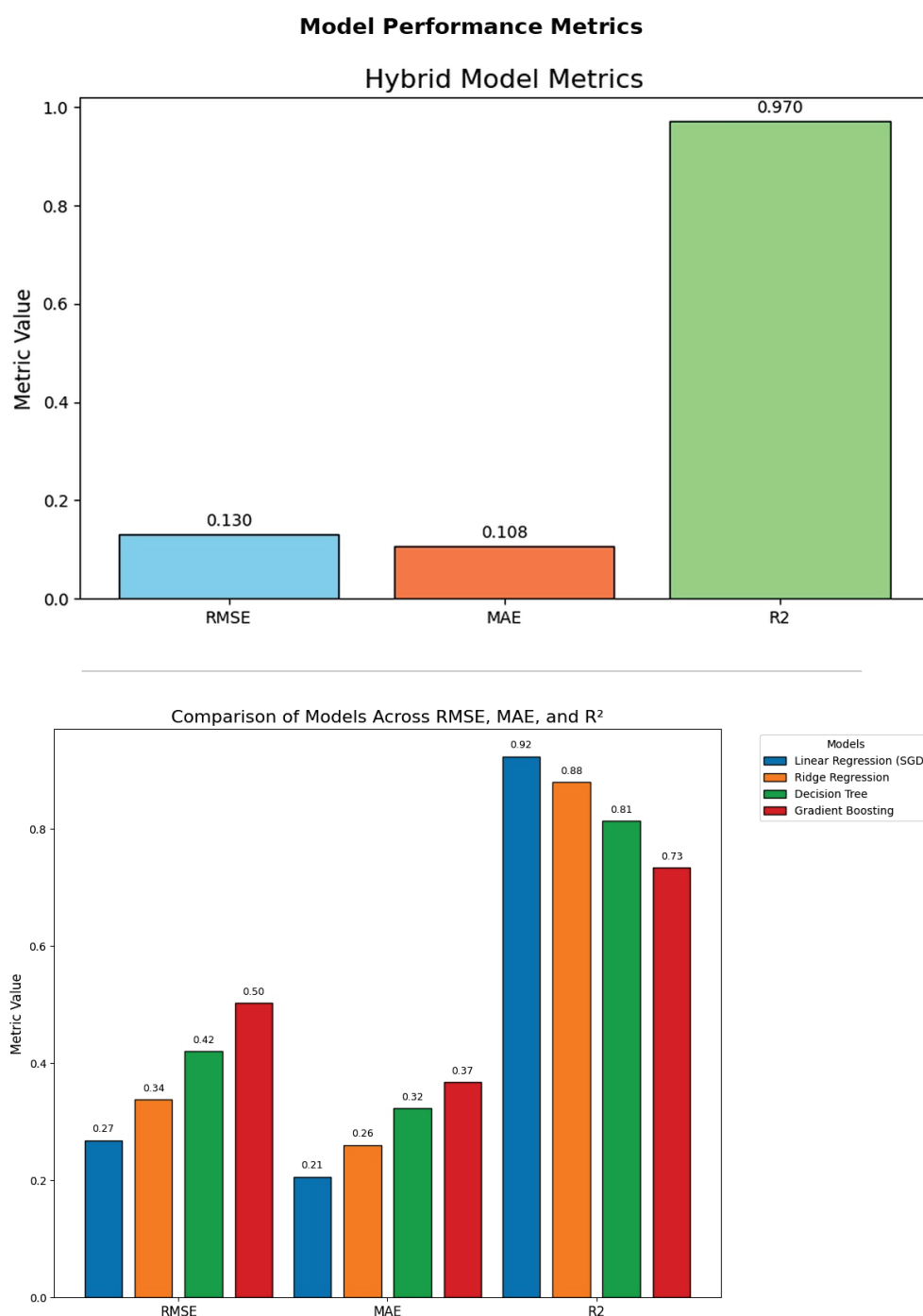


Figure 8. Spatial bubble map illustrating regional variations in groundwater storage.

Baseline Models’ Performance Insights: linear regression outperformed other baseline models, achieving the highest R<sup>2</sup> score and lowest RMSE/MAE. This suggests that the underlying data may have strong linear components that are effectively captured by the stochastic gradient optimization. Ridge regression, while slightly less accurate, still performed well. Decision tree performance declined due to its tendency to overfit on small local patterns, especially in datasets with temporal dependencies and continuous variables. The gradient boosting base, surprisingly, showed a lower R<sup>2</sup>. This may be attributed to shallow tree depths or insufficient learning iterations, which limited the model’s capacity to learn complex interactions. In this study, the gradient boosting model was configured with a limited number of estimators and shallow tree depth (`n_estimators = 50`, `max_depth = 2`, `learning_rate = 0.02`, `subsample = 0.8`), which may restrict its ability to capture complex non-linear relationships compared to more flexible ensemble architectures. The gradient boosting base was intentionally configured as a conservative baseline with limited depth and learning capacity, and its lower performance is therefore attributed to constrained model flexibility rather than inherent deficiencies in the input data.

Hybrid Model Performance: The hybrid model demonstrated superior performance due to the synergistic integration of multiple learning paradigms. Unlike baseline models that specialize in capturing either linear, non-linear, or local relationships in isolation, the hybrid model combines these strengths into a single predictive system. The key factors contributing to this improvement include the following:

- **Multi-Perspective Learning:** By incorporating diverse regressors (HR, ARD, KNN, KR, and LSTM), the hybrid model captures a wider spectrum of statistical patterns—from temporal dependencies to local clusters and global trends.
- **Error Compensation:** Weaknesses in individual models are balanced by strengths in others. For instance, LSTM can model temporal dependencies missed by KNN, while ARD enhances feature selection that benefits kernel methods.
- **RF Aggregation:** The meta-learner (RF) combines model outputs in a non-linear way, applying internal bootstrapping and meta-learning aggregation. This reduces prediction variance and improves generalization, especially in the presence of noisy or correlated features [35].



**Figure 9.** Comparison of models across RMSE, MAE, and R<sup>2</sup>, and model performance metrics.



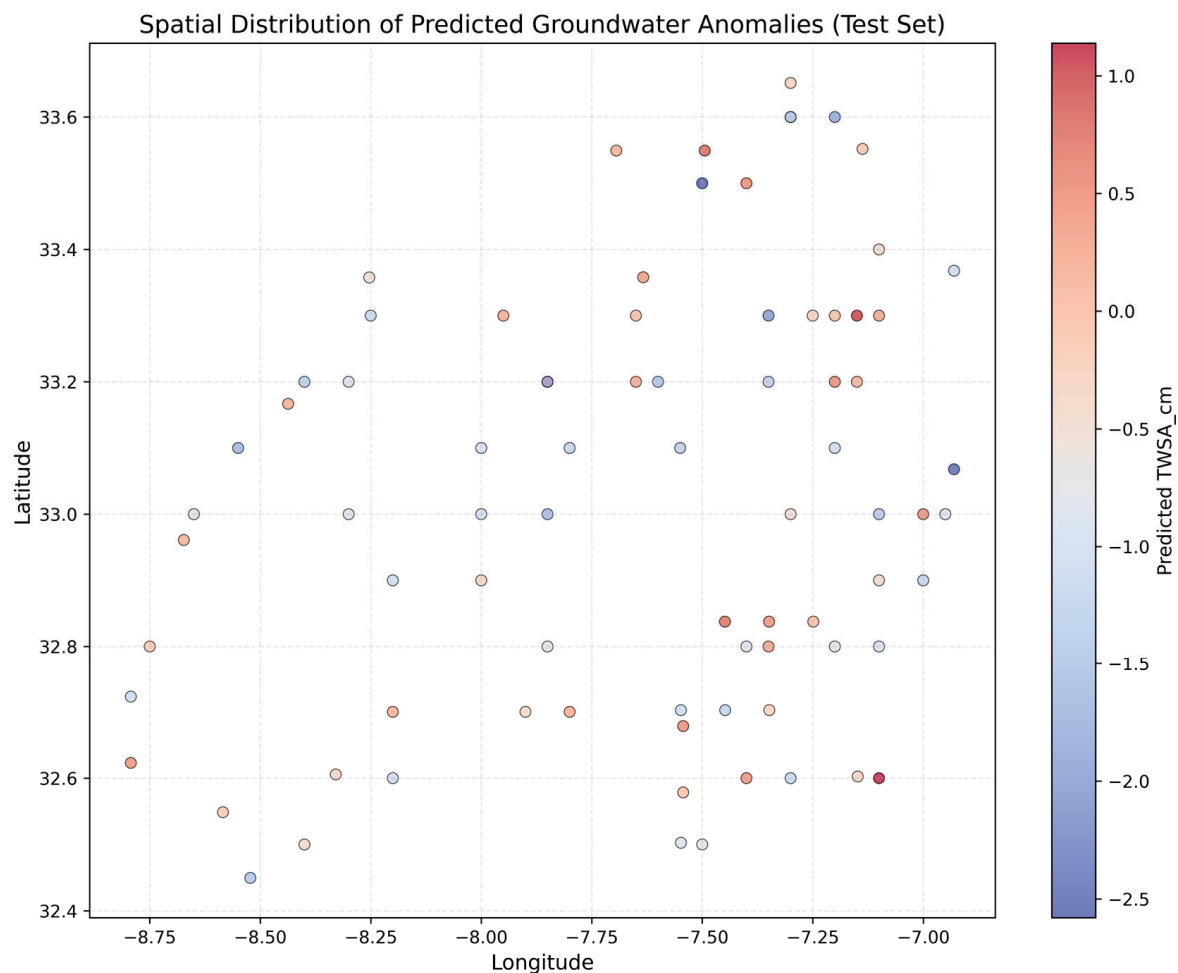
**Feature Stability via Preprocessing:** Prior feature normalization and interaction construction ensured consistent scaling across learners, improving convergence and learning quality.

#### 4.3. Insights from TWSA Analysis

To further understand the spatial and contextual implications of groundwater storage patterns in the Casablanca–Settat region, multiple visualization techniques were employed using the predicted TWSA values. These visualizations not only validate model behavior but also uncover geographic trends, LULC influences, and extreme anomaly zones that are critical for sustainable groundwater resource planning [56].

**Spatial Scatter Plot of Predicted Anomalies:** The scatter plot of predicted TWSA values in Figure 10 reveals clear spatial heterogeneity across the region. Each point represents a georeferenced prediction of groundwater storage deviation from the climatological norm:

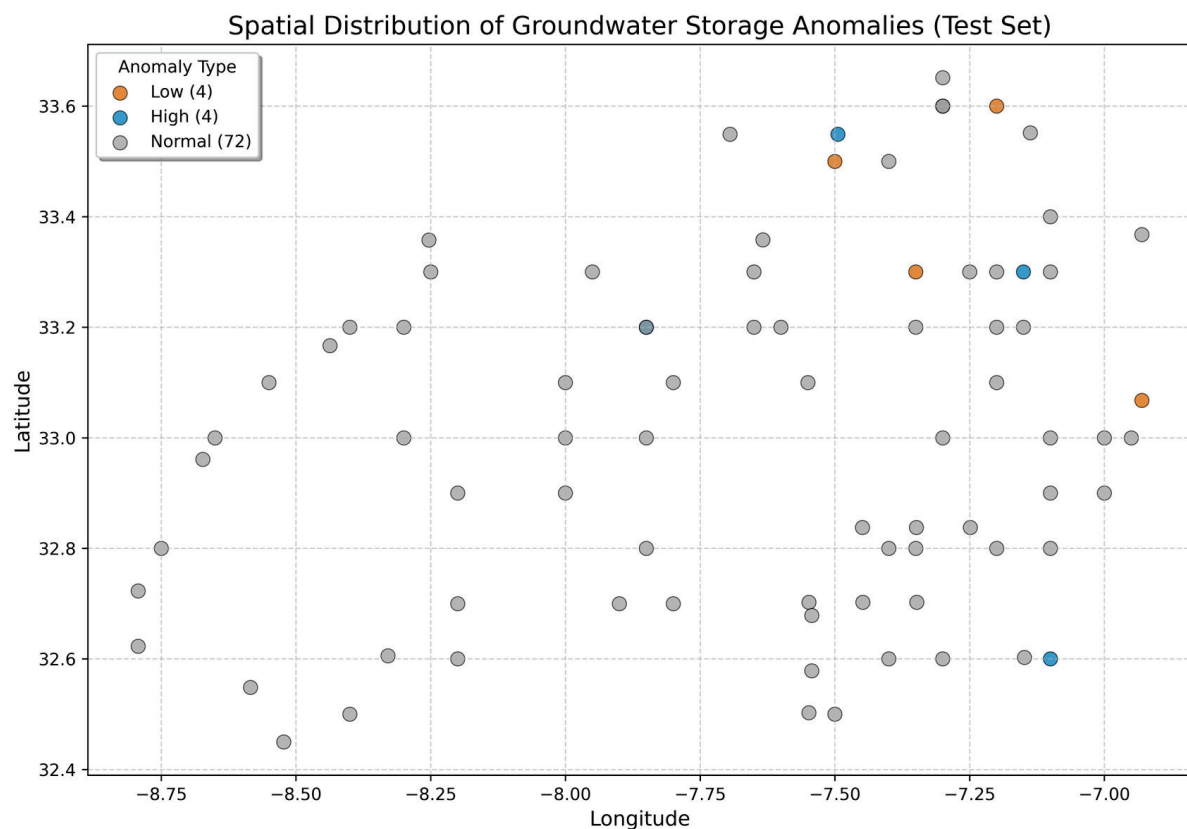
- Blue markers indicate negative anomalies, highlighting zones under stress or depletion, which may require immediate intervention or investigation.
- Red markers denote positive anomalies, possibly driven by recent rainfall, surface recharge, or low extraction activity.



**Figure 10.** Spatial scatter plot of predicted TWSA (cm).

The visual clustering of these zones supports the identification of geographically coherent patterns, aiding hydrological agencies in allocating monitoring resources efficiently. Additionally, Anomaly Detection and Extreme Zone Identification: To highlight

critical groundwater conditions, the predicted TWSA values were categorized and plotted in Figure 11.



**Figure 11.** Classified TWSAs: orange (deficit), blue (surplus), and gray (normal).

- Red zones: The lowest 5% of predicted values, representing areas at significant risk of groundwater depletion.
- Blue zones: The highest 5%, suggesting recharge-rich regions or underutilized aquifers.
- Gray zones: Represent normal storage behavior, serving as a baseline for temporal comparisons.

This classification aids anomaly tracking and early warning systems for water scarcity, helping stakeholders initiate localized action plans or awareness campaigns [13].

## 5. Discussion

### 5.1. Summary of Findings

The hybrid model proposed in this study significantly outperforms traditional baseline regression models in predicting TWSA for the Casablanca–Settat region [57]. The achieved  $R^2$  of 0.97, with an RMSE of 0.13 and an MAE of 0.108, marks a substantial improvement over earlier models like linear regression ( $R^2 = 0.924$ ) and decision tree regression ( $R^2 = 0.813$ ). This highlights the model's superior ability to capture both linear and non-linear hydro-climatic dependencies. Compared to earlier studies referenced in the literature review, our model demonstrates marked advantages in several dimensions. Table 3 below compares the key performance outcomes of our model with the most relevant previous approaches, strictly derived from the cited works.

Previous studies have leveraged deep learning or spatio-temporal models (Table 3), they generally focus on specific patterns (either temporal or spatial) without extensive

ensemble integration. Our approach focuses the benefits of diverse model types, leading to significantly higher prediction accuracy and robust generalizability.

**Table 3.** Comparative performance of existing studies and the proposed hybrid model.

| Study     | Model/Approach                          | Region/Context                          | Reported R <sup>2</sup> /Accuracy | Comment                                                                         |
|-----------|-----------------------------------------|-----------------------------------------|-----------------------------------|---------------------------------------------------------------------------------|
| [10]      | CNN-LSTM Hybrid                         | Global GRACE-based regions              | R <sup>2</sup> = 0.91             | Spatial patterns modeled, but high variance in prediction                       |
| [11]      | LSTM + Attention Mechanism              | Temporal groundwater anomaly prediction | R <sup>2</sup> = 0.93             | Focused on the temporal trend but lacks ensemble integration                    |
| This Work | LSTM + Ensemble ML (RF, KRR, ARD, etc.) | Casablanca–Settat, Morocco              | R <sup>2</sup> = 0.97             | Combines spatial-temporal and ensemble learning for fine-grained local accuracy |

5.2. Regional Relevance: Casablanca–Settat Groundwater Concerns

The Casablanca–Settat region, home to over 7 million residents and accounting for a significant portion of Morocco’s economic activity, faces acute groundwater stress due to its semi-arid climate and rapidly urbanizing footprint. The region exhibits pronounced seasonality, where rainfall is irregular and often confined to a few months, while the dry periods are prolonged. This climatic setup causes erratic recharge patterns, often failing to replenish deep aquifers fully [34,56,58]. Moreover, the region’s agricultural belt—especially in Settat and El Jadida—relies heavily on groundwater for irrigation. Coupled with expanding industrial zones and population clusters around Casablanca, the stress on aquifers is mounting. Over-extraction in shallow aquifers, absence of controlled recharge programs, and LULC changes further exacerbate depletion. The hybrid modeling framework presented in this study offers several region-specific benefits as follows:

- **Localized Anomaly Detection:** By detecting micro-level TWSAs, the model supports future zoning of groundwater stress at district and sub-district levels.
- **Temporal Sensitivity:** Integration of LSTM allows the model to capture seasonal recharge–discharge cycles critical in Morocco’s agro-ecological planning.
- **LULC Driven Forecasting:** The model incorporates land cover data to detect how different uses (urban, forest, cropland) impact local groundwater storage.
- **Data-Informed Policy:** High-resolution forecasts generated from the model can guide dynamic water allocation, irrigation scheduling, and artificial recharge decisions.

Although a formal feature attribution analysis was not conducted, the predictive behavior of the ensemble is primarily driven by hydroclimatic and land–surface variables that are physically linked to terrestrial water storage dynamics. Precipitation and evapotranspiration directly control water inputs and losses, while soil moisture and vegetation indices reflect surface and root-zone water availability. LULC and elevation provide spatial context related to groundwater abstraction intensity and recharge potential. The consistency of the downscaled patterns with known hydrogeological conditions suggests that the ensemble leverages physically meaningful relationships rather than purely statistical correlations.

This work fills a critical technological gap by equipping local water governance bodies with a fine-grained, interpretable, and spatially aware groundwater forecasting tool—a much-needed step in promoting water security across the Casablanca–Settat region.

5.3. Limitations and Future Scope

Although the proposed hybrid model has shown an acceptable level of predictive accuracy, there are some limitations that need to be noted to give an idea of how it should be enhanced. The dependence on GRACE (and the related GRACE-FO satellites) is one of the

main limitations of the work. This interferes with the sensitivity of the model to localized movements of groundwater, particularly in heterogeneous terrains like the subregion in the bush or topography-differing regions. Moreover, ensemble architecture, where several ML and deep learning algorithms are combined, presents a certain level of complexity of computations that might limit its usage in real-time monitoring or operations based in environments with scarce digital infrastructure [59,60].

The fact that the model demonstrates some form of dependence on past trends poses yet another challenge. It can be effective, during normal seasonal and climatic conditions, but can have less adaptability in case of extreme events, like abrupt LULC changes, non-controlled groundwater pumping, or prolonged drought periods. There is also a lack of dense, field-level validated data, especially in less-monitored districts of the Casablanca–Settat region, which also constrains the capacity to make a granular accuracy evaluation at the zone level [39,61,62].

This framework can be further enhanced by combining near-real-time data streams of the Internet-of-Things (IoT)–based groundwater monitoring systems. This would improve the time responsiveness of the model so that there will be timely interventions. Also, to render the insights of this model more accessible to the practitioners, including the water boards and agricultural planners, an interactive and multilingual decision-support interface, specifically for the local administrative units, would need to be developed. It would also be beneficial to explore transfer learning methods so that the adaptation of this framework to other regions of Morocco and basins in North Africa can be achieved. In addition, integrating the model with climate modeling tools can enhance long-term scenario planning, whereby authorities can model the resilience of groundwater under different climate change scenarios [43,63].

While the downscaled outputs are provided at a 10 km spatial grid, corresponding to approximately 200 spatial grid cells across the study region, independent validation of the spatial fidelity of these patterns using dense in situ groundwater observations or high-resolution hydrological models was not feasible due to data limitations in the study region. Consequently, model evaluation focuses on temporal consistency with GRACE-derived TWSA signals using standard performance metrics. The resulting spatial patterns should therefore be interpreted as physically informed, relative variations constrained by auxiliary predictors, rather than exact point-scale groundwater measurements.

#### 5.4. Applications of the Work

The hybrid model proposed has a substantial operational worth in regulating groundwater in the Casablanca–Settat region, wherein there are increasing demands on the resources as a result of intense agricultural production scenarios, fast urbanization processes, and climate change. With Settat and Berrechid provinces in particular and high water demand to cultivate crops, there lies the requirement in the Casablanca urban area, where the water footprint is also continuing to spread and strain the local aquifers [64].

Regional decisions can be made by the spatial predictions of TWSA that are highlighted by the model. For instance, consistently positive TWSA zones may be earmarked for recharge zone protection, while areas exhibiting persistent negative anomalies—such as ‘Sidi Bennour’ or ‘El Jadida’—could be prioritized for artificial recharge programs or stricter extraction regulations. In agricultural planning, the model offers a data-driven foundation for optimizing irrigation schedules, particularly during moisture-deficit periods, enabling agencies like the “Office Régional de Mise en Valeur Agricole (ORMVA), Agence de bassin hydraulique (ABH)” to balance water availability with crop productivity.

To manage the urban areas, the predictive information can assist authorities to better measure extraction stresses and think about groundwater licensing, well-boring zoning, or groundwater recharge parks within the peri-urban areas where infiltration has lessened due to the development of impervious surfaces. Furthermore, by aligning the model's outputs with national programs like the Plan Maroc Vert, the framework can contribute to broader sustainable water governance strategies in Morocco's most economically significant region [65,66].

## 6. Conclusions

This paper establishes a hybrid ML model to anticipate TWSA over the Casablanca–Settat area, where multi-source data was used to combine satellite-observed gravity anomalies, climatic variables, and LULC data. The architecture of the models used was an ensemble of six different algorithms: HR, ARD, KR, LSTM, KNN, and GB, along with an ensemble model consisting of RF as the meta-learner. The best predictive performance was revealed in evaluation results, where the ensemble model delivered an RMSE of 0.13, an MAE of 0.108, and an  $R^2$  score of 0.97, which was way out of the baseline models.

In addition to its precision, which is in the numerical domain, the model provided useful spatial information on how the groundwater behaved in different regions and the uncovered relationships between water storage behavior and the types of land cover. Such functionalities render the framework as a feasible decision-aiding framework for water resource planning and drought monitoring, priority assignments to aquifer recharge, and adaptive irrigation management. The fact that it is modular and interpretable also means that it is scalable to other regions of comparable data availability. Integrating the empirical results of the expectations about groundwater through data and human needs of water management (in practical). This research works to bring about the contemporary theme of sustainable groundwater management in the semi-arid areas.

**Author Contributions:** Conceptualization, Y.L., N.E.A. and O.O.; methodology, Y.L., N.E.A. and O.O.; software, Y.L.; validation, N.E.A., T.T.N. and A.M.S.; formal analysis, Y.L.; investigation, O.O.; resources, O.O. and N.E.A.; data curation, Y.L. and O.O.; writing—original draft preparation, Y.L. and O.O.; writing—review and editing, Y.L., O.O., N.E.A. and A.M.S.; visualization, Y.L., N.E.A. and T.T.N.; supervision, N.E.A., T.T.N. and A.M.S.; project administration, Y.L. and N.E.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The original data presented in the study, with all analysis code (Jupyter notebook), are archived on Zenodo at <https://doi.org/10.5281/zenodo.18017503>. The record is currently restricted to avoid premature public release during peer review, but access can be granted to editors and reviewers on request, and it will be made fully public upon article acceptance.

**Acknowledgments:** The authors express their sincere gratitude to the Moroccan Institute of Scientific and Technical Information (Institut Marocain de l'Information Scientifique et Technique, IMIST-CNRST) for providing access to international scientific databases such as Scopus and Web of Science through the e-resources platform, which was essential for conducting the bibliometric analysis and literature review of the present study.

**Conflicts of Interest:** The authors declare no conflicts of interest.



## References

1. Mukherjee, A.; Scanlon, B.R.; Aureli, A.; Langan, S.; Guo, H.; McKenzie, A. Chapter 1—Global Groundwater: From Scarcity to Security through Sustainability and Solutions. In *Global Groundwater*; Mukherjee, A., Scanlon, B.R., Aureli, A., Langan, S., Guo, H., McKenzie, A.A., Eds.; Elsevier: Amsterdam, The Netherlands, 2021; pp. 3–20, ISBN 978-0-12-818172-0.
2. Velis, M.; Conti, K.I.; Biermann, F. Groundwater and Human Development: Synergies and Trade-Offs within the Context of the Sustainable Development Goals. *Sustain. Sci.* **2017**, *12*, 1007–1017. [CrossRef] [PubMed]
3. Scanlon, B.R.; Fakhreddine, S.; Rateb, A.; De Graaf, I.; Famiglietti, J.; Gleeson, T.; Grafton, R.Q.; Jobbagy, E.; Kebede, S.; Kolusu, S.R.; et al. Global Water Resources and the Role of Groundwater in a Resilient Water Future. *Nat. Rev. Earth Environ.* **2023**, *4*, 87–101. [CrossRef]
4. Ali, S.; Ran, J.; Khorrami, B.; Wu, H.; Tariq, A.; Jehanzaib, M.; Khan, M.M.; Faisal, M. Downscaled GRACE/GRACE-FO Observations for Spatial and Temporal Monitoring of Groundwater Storage Variations at the Local Scale Using Machine Learning. *Groundw. Sustain. Dev.* **2024**, *25*, 101100. [CrossRef]
5. Shilengwe, C.; Banda, K.; Nyambe, I. Machine Learning Downscaling of GRACE/GRACE-FO Data to Capture Spatial-Temporal Drought Effects on Groundwater Storage at a Local Scale under Data-Scarcity. *Environ. Syst. Res.* **2024**, *13*, 38. [CrossRef]
6. Hamou-Ali, Y.; Karmouda, N.; Mohsine, I.; Bouramtane, T.; Kacimi, I.; Tweed, S.; Tahiri, M.; Kassou, N.; El Bilali, A.; Chafki, O.; et al. Downscaling GRACE Total Water Storage Data Using Random Forest: A Three-Round Validation Approach under Drought Conditions. *Front. Water* **2025**, *7*, 1545821. [CrossRef]
7. Khorrami, B.; Ali, S.; Gündüz, O. Investigating the Local-Scale Fluctuations of Groundwater Storage by Using Downscaled GRACE/GRACE-FO JPL Mascon Product Based on Machine Learning (ML) Algorithm. *Water Resour. Manag.* **2023**, *37*, 3439–3456. [CrossRef]
8. Satizábal-Alarcón, D.A.; Suhogusoff, A.; Ferrari, L.C. Characterization of Groundwater Storage Changes in the Amazon River Basin Based on Downscaling of GRACE/GRACE-FO Data with Machine Learning Models. *Sci. Total Environ.* **2024**, *912*, 168958. [CrossRef] [PubMed]
9. Liu, C.; Zhang, Z.; Xu, C.; Zhang, W. Reconstructing Long-Term, High-Resolution Groundwater Storage Changes in the Songhua River Basin Using Supplemented GRACE and GRACE-FO Data. *Remote Sens.* **2024**, *16*, 4566. [CrossRef]
10. Moudgil, P.S.; Rao, G.S.; Heki, K. Bridging the Temporal Gaps in GRACE/GRACE-FO Terrestrial Water Storage Anomalies over the Major Indian River Basins Using Deep Learning. *Nat. Resour. Res.* **2024**, *33*, 571–590. [CrossRef]
11. Mousavimehr, S.M.; Kavianpour, M.R. A Non-Stationary Downscaling and Gap-Filling Approach for GRACE/GRACE-FO Data under Climatic and Anthropogenic Influences. *Appl. Water Sci.* **2025**, *15*, 91. [CrossRef]
12. Li, W.; Bao, L.; Yao, G.; Wang, F.; Guo, Q.; Zhu, J.; Zhu, J.; Wang, Z.; Bi, J.; Zhu, C.; et al. The Analysis on Groundwater Storage Variations from GRACE/GRACE-FO in Recent 20 Years Driven by Influencing Factors and Prediction in Shandong Province, China. *Sci. Rep.* **2024**, *14*, 5819. [CrossRef] [PubMed]
13. Seo, J.Y.; Lee, S.-I. Predicting Changes in Spatiotemporal Groundwater Storage Through the Integration of Multi-Satellite Data and Deep Learning Models. *IEEE Access* **2021**, *9*, 157571–157583. [CrossRef]
14. Moudgil, P.S.; Rao, G.S. Groundwater Levels Estimation from GRACE/GRACE-FO and Hydro-Meteorological Data Using Deep Learning in Ganga River Basin, India. *Environ. Earth Sci.* **2023**, *82*, 441. [CrossRef]
15. Tariq, A.; Ali, S.; Basit, I.; Jamil, A.; Farmonov, N.; Khorrami, B.; Khan, M.M.; Sadri, S.; Baloch, M.Y.J.; Islam, F.; et al. Terrestrial and Groundwater Storage Characteristics and Their Quantification in the Chitral (Pakistan) and Kabul (Afghanistan) River Basins Using GRACE/GRACE-FO Satellite Data. *Groundw. Sustain. Dev.* **2023**, *23*, 100990. [CrossRef]
16. Amiri, V.; Ali, S.; Sohrabi, N. Estimating the Spatio-Temporal Assessment of GRACE/GRACE-FO Derived Groundwater Storage Depletion and Validation with in-Situ Water Quality Data (Yazd Province, Central Iran). *J. Hydrol.* **2023**, *620*, 129416. [CrossRef]
17. Hasan, M.S.U.; Saif, M.M.; Ahmad, N.; Rai, A.K.; Khan, M.A.; Aldrees, A.; Khan, W.A.; Mohammed, M.K.A.; Yaseen, Z.M. Spatiotemporal Analysis of Future Trends in Terrestrial Water Storage Anomalies at Different Climatic Zones of India Using GRACE/GRACE-FO. *Sustainability* **2023**, *15*, 1572. [CrossRef]
18. Nikraftar, Z.; Parizi, E.; Saber, M.; Hosseini, S.M.; Ataie-Ashtiani, B.; Simmons, C.T. Groundwater Sustainability Assessment in the Middle East Using GRACE/GRACE-FO Data. *Hydrogeol. J.* **2024**, *32*, 321–337. [CrossRef]
19. Qu, W.; Zhang, P.; Chen, P.; Li, J.; Gao, Y. Spatiotemporal Variations and Sustainability Characteristics of Groundwater Storage in North China from 2002 to 2022 Revealed by GRACE/GRACE Follow-On and Multiple Hydrologic Data. *Remote Sens.* **2024**, *16*, 1176. [CrossRef]
20. Hadavi, D.; Mousavi, S.M.; Rahimzadegan, M. An Intercomparison of the Groundwater Level Estimations by GRACE and GRACE-FO Satellites and Groundwater Modeling in Iran. *Acta Geophys.* **2024**, *72*, 3609–3629. [CrossRef]
21. Wu, C.; Lu, C.; Hu, J.; Liu, B.; Shu, L.; Zhang, Y. Machine Learning-Based Downscaling of GRACE Data to Enhance Assessment of Spatiotemporal Evolution of Coastal Plain Groundwater Storage. *Water Resour. Manag.* **2025**, *39*, 6377–6397. [CrossRef]

22. Khorrami, B.; Ali, S.; Gündüz, O. An Appraisal of the Local-scale Spatio-temporal Variations of Drought Based on the Integrated GRACE/GRACE-FO Observations and Fine-resolution FLDAS Model. *Hydrol. Process.* **2023**, *37*, e15034. [CrossRef]
23. Hamdi, M.; El Alem, A.; Goita, K. Groundwater Storage Estimation in the Saskatchewan River Basin Using GRACE/GRACE-FO Gravimetric Data and Machine Learning. *Atmosphere* **2025**, *16*, 50. [CrossRef]
24. Szabó, V. Comparison Features Importance for Temporal and Spatial-Temporal Approaches in GRACE and GRACE-FO Signal Reconstruction Based on GLDAS Data. *Int. J. Hydrol. Sci. Technol.* **2023**, *16*, 370–389. [CrossRef]
25. El Assaoui, N.; Bouiss, C.; Sadok, A. Assessment of Water Erosion by Integrating RUSLE Model, GIS and Remote Sensing—Case of Tamdoust Watershed (Morocco). *Ecol. Eng. Environ. Technol.* **2023**, *24*, 43–53. [CrossRef] [PubMed]
26. El Assaoui, N.; Sadok, A.; Bendaraa, A.; Charafi, M.M. Two-Dimensional Numerical Modeling of Morphodynamic Evolution in Bouregreg Estuary (Morocco). *Ecol. Eng. Environ. Technol.* **2023**, *24*, 217–230. [CrossRef] [PubMed]
27. Karmouda, N.; El Assaoui, N.; Kacimi, I.; Mahe, G.; Bouramtane, T.; Brirhet, H.; Idrissi, A.; Kassou, N. Hydrological Modelling of Extreme Events in Ouergha Mediterranean Basin, Northern Morocco, Using a Deterministic Model and Gridded Precipitations. *IGJ* **2023**, *56*, 1–20. [CrossRef]
28. Laalaoui, Y.; Ellassaoui, N.; Ouahine, O. Balancing Urban Growth and the Sustainability of Groundwater and Agricultural Land: Case of the Berrechid-Settat Area. In *Proceedings of the E3S Web of Conferences*; EDP Sciences: Les Ulis, France, 2024; Volume 489, p. 04012. [CrossRef]
29. Lrhoul, H.; Assaoui, N.E.; Turki, H. Mapping of Water Research in Morocco: A Scientometric Analysis. In *Proceedings of the Materials Today: Proceedings*; Elsevier Ltd.: Amsterdam, The Netherlands, 2021; Volume 45, pp. 7321–7328.
30. Belgiu, M.; Drăguț, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
31. Achahboun, C.; Chikhaoui, M.; Naimi, M.; Bellafkih, M. Crops Classification Using Machine Learning and Google Earth Engine. In *Proceedings of the SITA 2023: 2023 14th International Conference on Intelligent Systems: Theories and Applications*, Casablanca, Morocco, 22–23 November 2023; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2023.
32. Abuzied, S.M.; Pradhan, B. Hydro-Geomorphic Assessment of Erosion Intensity and Sediment Yield Initiated Debris-Flow Hazards at Wadi Dahab Watershed, Egypt. *Georisk Assess. Manag. Risk Eng. Syst. Geohazards* **2021**, *15*, 221–246. [CrossRef]
33. Elmotawakkil, A.; Moumane, A.; Zahi, A.; Sadiki, A.; Karkouri, J.A.; Batchi, M.; Bhagat, S.K.; Enneya, N. HydroPredictor a Hybrid Machine Learning Model for Addressing Data Scarcity in Groundwater Prediction. *Sci. Rep.* **2025**, *15*, 44069. [CrossRef]
34. Ben Driss, M.A.; Ifilllis, A.; Ettazarini, S.; Hahou, Y.; Boudad, L.; El Amrani, M.; Courba, S. Groundwater Potential Mapping in Fractured Aquifers Using Remote Sensing and GIS Technology in the Moulouya Region, Morocco. *Iraqi Geol. J.* **2024**, *57*, 172–190. [CrossRef]
35. Snousy, M.G.; Elshafie, H.M.; Abouelmagd, A.R.; Hassan, N.E.; Abd-Elmaboud, M.E.; Mohammadi, A.A.; Elewa, A.M.T.; EL-Sayed, E.; Saqr, A.M. Enhancing the Prediction of Groundwater Quality Index in Semi-Arid Regions Using a Novel ANN-Based Hybrid Arctic Puffin-Hippopotamus Optimization Model. *J. Hydrol. Reg. Stud.* **2025**, *59*, 102424. [CrossRef]
36. Rodell, M.; Reager, J.T. Water Cycle Science Enabled by the GRACE and GRACE-FO Satellite Missions. *Nat. Water* **2023**, *1*, 47–59. [CrossRef]
37. Yan, Z.; Ran, J.; Xiao, Y.; Xu, Z.; Wu, H.; Deng, X.; Du, L.; Zhong, M. The Temporal Improvement of Earth's Mass Transport Estimated by Coupling GRACE-FO With a Chinese Polar Gravity Satellite Mission. *JGR Solid Earth* **2023**, *128*, e2023JB027157. [CrossRef]
38. Forkuor, G.; Hounkpatin, O.K.L.; Welp, G.; Thiel, M. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. *PLoS ONE* **2017**, *12*, e0170478. [CrossRef]
39. Bhungeni, O.; Ramjatan, A.; Gebreslasie, M. Evaluating Machine-Learning Algorithms for Mapping LULC of the uMngeni Catchment Area, KwaZulu-Natal. *Remote Sens.* **2024**, *16*, 2219. [CrossRef]
40. Mallick, J.; Alqadhi, S.; Talukdar, S.; Hang, H.T. Evaluating groundwater sustainability and vegetation dynamics in arid regions: Advanced remote sensing and spatiotemporal analysis in Saudi Arabia. *Environ. Technol. Innov.* **2025**, *38*, 104203. [CrossRef]
41. Sabri, A.; Bahi, H.; Bounoua, L.; Tahiri, M.; Tweed, S.; LeBlanc, M.; Bouramtane, T.; Malah, A.; Kacimi, I. Predictive Modelling on Spatial-Temporal Land Use and Land Cover Changes at the Casablanca-Settat Region in Morocco. *Model. Earth Syst. Environ.* **2024**, *10*, 6691–6714. [CrossRef]
42. Guellaf, A.; Kettani, K. Assessment of Seasonal Variations in Water Quality and Pollution Sources in the Coastal Water Bodies between Casablanca and Rabat (Northwest Morocco). *Ecol. Eng. Environ. Technol.* **2025**, *26*, 323–336. [CrossRef]

43. Amazirh, A.; Ouassanouan, Y.; Bouimouass, H.; Baba, M.W.; Bouras, E.H.; Rafik, A.; Benkirane, M.; Hajhouji, Y.; Ablila, Y.; Chehbouni, A. Remote Sensing-Based Multiscale Analysis of Total and Groundwater Storage Dynamics over Semi-Arid North African Basins. *Remote Sens.* **2024**, *16*, 3698. [CrossRef]
44. Bao, F.; Martek, I.; Wu, Q.; Wang, Z.; Yang, Y.; Chen, C.; Chan, A.P.C. Scientometric Review of Smart Water Management Literature from the Sustainable Development Goal Perspective. *Int. J. Strateg. Prop. Manag.* **2023**, *27*, 218–232. [CrossRef]
45. Ding, C.; Guo, Z.; Chen, K.; Fan, L.; Zhan, Y.; Kuang, X.; Cui, B.; Zheng, C. The Effect of Seasonally Frozen Ground on Rainfall Infiltration and Groundwater Discharge in Qinghai Lake Basin, China. *Front. Water* **2024**, *6*, 1495763. [CrossRef]
46. El Assaoui, N.; Sadok, A.; Merimi, I. Impacts of Climate Change on Moroccan's Groundwater Resources: State of Art and Development Prospects. In *Proceedings of the Materials Today: Proceedings*; Elsevier Ltd.: Amsterdam, The Netherlands, 2021; Volume 45, pp. 7690–7696.
47. Hossain, M.; Wiegand, B.; Reza, A.; Chaudhuri, H.; Mukhopadhyay, A.; Yadav, A.; Patra, P.K. A machine learning approach to investigate the impact of land use land cover (LULC) changes on groundwater quality, health risks and ecological risks through GIS and response surface methodology (RSM). *J. Environ. Manag.* **2024**, *366*, 121911. [CrossRef]
48. Assaoui, N.E.; Sadok, A.; Charafi, M. Analysis of a Water Supply Intake from a Silted Dam Using Two-Dimensional Horizontal Numerical Modeling: Case of Mechraa Hammadi Dam (Morocco). *Mater. Today Proc.* **2021**, *45*, 7718–7724. [CrossRef]
49. Saqr, A.M.; Kartal, V.; Karakoyun, E.; Abd-Elmaboud, M.E. Improving the Accuracy of Groundwater Level Forecasting by Coupling Ensemble Machine Learning Model and Coronavirus Herd Immunity Optimizer. *Water Resour. Manag.* **2025**, *39*, 5415–5442. [CrossRef]
50. Fentaw, A.E.; Abegaz, A. Analyzing Land Use/Land Cover Changes Using Google Earth Engine and Random Forest Algorithm and Their Implications to the Management of Land Degradation in the Upper Tekeze Basin, Ethiopia. *Sci. World J.* **2024**, *2024*, 3937558. [CrossRef] [PubMed]
51. Dharumarajan, S.; Hegde, R. Digital Mapping of Soil Texture Classes Using Random Forest Classification Algorithm. *Soil Use Manag.* **2022**, *38*, 135–149. [CrossRef]
52. Lu, D.; Weng, Q.; Li, G. Evaluation of Training Sample Sizes and Class Distribution on Machine Learning Classifiers for Land Cover Classification Using Google Earth Engine. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *100*, 102327. [CrossRef]
53. Nazari, A.; Jamshidi, M.; Roozbahani, A.; Golparvar, B. Groundwater Level Forecasting Using Empirical Mode Decomposition and Wavelet-Based Long Short-Term Memory (LSTM) Neural Networks. *Groundw. Sustain. Dev.* **2025**, *28*, 101397. [CrossRef]
54. Barzegar, M.; Gharehdash, S.; Chowdhury, F.; Liu, M.; Timms, W. Hybrid Machine Learning for Predicting Groundwater Level: A Comparison of Boosting Algorithms with Neural Networks. *Groundw. Sustain. Dev.* **2025**, *31*, 101508. [CrossRef]
55. Chomani, K.; Pshdari, S. Evaluation of Different Classification Algorithms for Land Use Land Cover Mapping. *KJAR* **2024**, *9*, 13–22. [CrossRef]
56. Bouzekraoui, H.; El Asri, Z.; Ouikhalan, M.; Benyaich, A. Urban Sprawl and Land Degradation in the Casablanca-Settat Region of Morocco: Analysis and Implications. *Urban Plan. Dev.* **2020**, *146*, 05020011. [CrossRef]
57. Jadouane, A.; Chaouki, A. Simulation of the Flood of El Maleh River by GIS in the City of Mohammedia-Morocco. In *Proceedings of the Climate Change and Water Security: Select Proceedings of VCDRR 2021*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 93–104.
58. El Haou, M.; Ourribane, M.; Ismaili, M.; Abdelrahman, K.; Fnais, M.S.; Krimissa, S.; El Oudi, H.; Hajji, S.; El Bouzkraoui, M.; Tarchi, F.; et al. Advanced GIS-Based Modeling for Flood Hazards Mapping in Urban Semi-Arid Regions: Insights from Beni Mellal, Morocco. *Front. Environ. Sci.* **2025**, *13*, 1585926. [CrossRef]
59. Feizizadeh, B.; Omarzadeh, D.; Kazemi Garajeh, M.; Lakes, T.; Blaschke, T. Machine Learning Data-Driven Approaches for Land Use/Cover Mapping and Trend Analysis Using Google Earth Engine. *J. Environ. Plan. Manag.* **2023**, *66*, 665–697. [CrossRef]
60. Talukdar, S.; Singha, P.; Mahato, S.; Shahfahad, P.; Pal, S.; Liou, Y.-A.; Rahman, A. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sens.* **2020**, *12*, 1135. [CrossRef]
61. Suryawanshi, A.; Sharnagat, N. *Google Earth Engine-Application in LULC Classification*; ICAR Research Complex for NEH Region: Meghalaya, India, 2024. [CrossRef]
62. Riche, A.; Ricci, R.; Melgani, F.; Drias, A.; Souissi, B. Machine Learning Approach to LULC Forecasting. In *Proceedings of the 2024 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*, Oran, Algeria, 15–17 April 2024; IEEE: Oran, Algeria, 2024; pp. 59–63.
63. Landini, F.; Malerba, F.; Mavilia, R. The Structure and Dynamics of Networks of Scientific Collaborations in Northern Africa. *Scientometrics* **2015**, *105*, 1787–1807. [CrossRef]
64. Bouabid, H.; Martin, B.R. Evaluation of Moroccan Research Using a Bibliometric-Based Approach: Investigation of the Validity of the h-Index. *Scientometrics* **2009**, *78*, 203–217. [CrossRef]

65. Xie, S.; Liu, L.; Zhang, X.; Yang, J.; Chen, X.; Gao, Y. Automatic Land-Cover Mapping Using Landsat Time-Series Data Based on Google Earth Engine. *Remote Sens.* **2019**, *11*, 3023. [CrossRef]
66. Sharnagat, N.; Nema, A.K.; Mishra, P.K.; Patidar, N.; Kumar, R.; Suryawanshi, A.; Radha, L. State-of-the-Art Status of Google Earth Engine (GEE) Application in Land and Water Resource Management: A Scientometric Analysis. *J. Geovis. Spat. Anal.* **2025**, *9*, 16. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## Article

# Fusion and Evaluation of Multi-Source Satellite Remote Sensing Precipitation Products Based on Transformer Machine Learning

Qingyuan Luo <sup>1,2,3</sup>, Dongzhi Wang <sup>2</sup>, Lina Liu <sup>4</sup>, Caihong Hu <sup>1</sup> and Chengshuai Liu <sup>1,\*</sup>

<sup>1</sup> School of Water Conservancy and Transportation, Zhengzhou University, Zhengzhou 450001, China

<sup>2</sup> Henan Provincial Hydrology and Water Resources Monitoring and Forecasting Center, Zhengzhou 450003, China

<sup>3</sup> Anyang Branch of Hydrology and Water Resources Monitoring and Forecasting Center in Henan Province, Anyang 455099, China

<sup>4</sup> National Supercomputing Center in Zhengzhou, Zhengzhou University, Zhengzhou 450001, China

\* Correspondence: liuchengshuai@zzu.edu.cn; Tel.: +86-15237157043

## Abstract

Satellite precipitation products offer great potential for acquiring reliable precipitation data in data-sparse areas, yet they have inherent uncertainties and errors as indirect observations. This study evaluated the accuracy of multi-source satellite precipitation products from daily and precipitation magnitude perspectives and discussed the spatiotemporal variation in their inversion errors. Based on ground rainfall observations, satellite products, and environmental factors, a Transformer-based multi-source precipitation fusion method was proposed, with its effectiveness preliminarily analyzed for daily precipitation in the Jingle River Basin. The main conclusions are as follows: (1) Compared with the observed precipitation data, the GSMaP\_Gauge satellite remote sensing precipitation product showed the closest agreement with the observations, ranking first in all indicators except the Probability of Detection (POD). The MSWEP satellite remote sensing precipitation product followed in performance, while the CHIRPS satellite product performed the poorest. Satellite products showed distinct error characteristics across seasons and rainfall intensities, as well as general overestimation of light rain frequency and insufficient heavy rain capture; however, these products also showed better detection capability in flood seasons. Error spatial distribution was consistent with topography, vegetation coverage, and temperature. (2) Verification demonstrated that the Transformer fusion algorithm effectively reduced relative bias and improved correlation with ground data. The scheme which incorporated environmental factors outperformed the other, which only considered precipitation characteristics, achieving higher estimation accuracy and fusion stability.

**Keywords:** multi-source precipitation fusion; machine learning; transfer learning; Transformer model; satellite remote sensing; Jingle River Basin

## 1. Introduction

High-quality and high-spatial-temporal resolution precipitation information plays an important role in medium-scale weather numerical forecasting and verification, hydrological monitoring and forecasting, water resource planning and management, as well as flood and drought monitoring [1,2]. It can also be used for disaster prevention and mitigation, such as monitoring small-scale extreme rainfall events and the resulting floods, geological



hazards, and so on [3]. However, precipitation exhibits significant spatiotemporal variability due to various factors, including atmospheric conditions, geographical location, and underlying surface conditions. Traditional point observations from ground-based rain gauges have limitations in spatiotemporal continuity, making them one of the main sources of uncertainty in hydrometeorological research [4,5]. Satellite remote sensing precipitation data provides a new approach for studying the spatial variations in precipitation and the complexity of regional precipitation. Currently, there are dozens of mainstream satellite remote sensing precipitation products which differ in terms of monitoring sensor types and inversion algorithms. Algorithms for satellite-based precipitation estimation include the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) [6], Climate Prediction Center Morphing Technique (CMORPH) [7], Tropical Rainfall Measuring Mission (TRMM), and the Multi-satellite Precipitation Analysis (TMPA) [8]. Different satellite remote sensing precipitation products have different error characteristics and sources of uncertainty, and their precipitation estimation accuracy and error characteristics often exhibit local or regional features.

Regarding satellite remote sensing precipitation products, their most studied aspects have been the evaluation or validation of precipitation estimation accuracy, as well as the analysis of error characteristics [9–11]. Sharifi et al. [12] conducted accuracy assessments for three satellite remote sensing precipitation products, namely IMERG Final Run, TRMM, and Multi-satellite Precipitation Analysis (TMPA-3B42), in four different basins in Iran with varying terrain and climate conditions. The results showed that all three products exhibited underestimation at the daily scale, with IMERG performing the best overall and having the best monitoring capability for heavy precipitation events. Liu et al. [13] evaluated the applicability of the TMPA-3B42 product at different time scales using observed precipitation data in China. The study found that the accuracy of TMPA was unstable, though the accuracy improved with an increasing time scale. Guo et al. [14] evaluated four satellite remote sensing precipitation products, namely CHIRPS, CMORPH, TRMM, and PERSIANN-CDR, in the Yangtze River Basin. The results showed that TRMM and CMORPH performed better overall at the temporal scale, but there were significant differences in accuracy among the four products in terms of spatial distribution. Helmi et al. [9] evaluated five satellite remote sensing precipitation products, namely CHIRPS V2.0, CMORPH, TMPA 3B42 V7, PERSIANN-CDR, and GPM IMERG V6, at the daily scale in Saudi Arabia. The study found that different precipitation products exhibited different performances in different regions and even at different time periods.

Different precipitation observation sources have their own characteristics of estimation errors in terms of spatiotemporal scale. How to scientifically and reasonably utilize multi-source precipitation observations to improve the reliability of precipitation estimation is a key issue in studying the spatiotemporal variability of hydrometeorological processes and improving the spatiotemporal accuracy of precipitation estimation [1,15,16]. For example, Sreeparvathy et al. [15] fused remote sensing precipitation products, such as CHIRPS and IMERG, with ground-based observation networks using Bayesian fuzzy clustering techniques, significantly improving the precipitation estimation capability. Li et al. [17] proposed a multi-source precipitation information fusion framework considering the identification of rainy and non-rainy conditions. They fused the MSWEP V2.1 product with daily precipitation data from ground-based observation networks, and the proposed method effectively improved the detection capability of satellite remote sensing precipitation estimation.

Common data fusion methods rely on ground-based precipitation observations to correct remote sensing precipitation estimates. These methods include co-kriging [18], Bayesian correction [19], Kalman filtering [20], optimal interpolation-probabilistic matching [21], geo-

graphically weighted regression [22], and so on. Although the expression and applicability conditions of precipitation fusion methods vary, the core idea is to construct a background field (or initial field) of precipitation based on weighted averaging, regression, analysis, filtering analysis, and other mathematical concepts. Then, the ground-based observed precipitation information is used to correct the background field under the assumed optimization criteria, providing a relative direct treatment of errors in remote sensing precipitation observations. However, due to the diversity of precipitation data sources and the uncertainty of their accuracy at different spatiotemporal scales, it is still challenging to find a multi-source precipitation information fusion algorithm that is suitable for various types of data and applications and to establish a unified mathematical fusion model [1,23–25]. Traditional precipitation data fusion methods mainly rely on statistical approaches for massive observation data. However, these methods often have limitations and deficiencies in practical applications. Deep learning, with its powerful ability to explore deep features in data, strong nonlinear fitting capability, and adaptability to different task scenarios, has great potential for improving the accuracy and reliability of precipitation estimation. Du et al. [26] proposed a precipitation forecasting model based on particle swarm optimization support vector machine (PSO-SVM) to replace the linear threshold of traditional precipitation. Various parameters such as temperature, humidity, wind speed, and direction were considered. Bhuiyan [27] used a tree-based model quantile regression forest (QRF) to combine multiple global precipitation datasets and characterize the uncertainty of the combined product. Ehsan Bhuiyan [28] evaluated a machine learning-based precipitation ensemble technique (MLPET) in three tropical mountainous regions. This technique was based on quantile regression forest and integrated global satellite precipitation datasets from CMORPH, PERSIANN, GSMaP (V6), and 3B42 (V7), as well as atmospheric reanalysis precipitation products (El\_GPCC) with daily soil moisture, specific humidity, and terrain elevation datasets. Pham [29] developed an adaptive network-based fuzzy inference system that used particle swarm optimization (PSOANFIS), artificial neural network (ANN), and support vector machine (SVM) to predict daily rainfall in the Binh Phuoc province in Vietnam. The aforementioned studies on rainfall fusion based on machine learning models have achieved good results. However, most of the previous research on prediction has focused on fused precipitation based on environmental factors as feature variables, and there are fewer studies investigating fuse precipitation based on its own characteristics. The Transformer neural network, with its transfer learning capabilities, can comprehensively consider the intrinsic features of rainfall data and perform transfer learning to maximize the retention of key information.

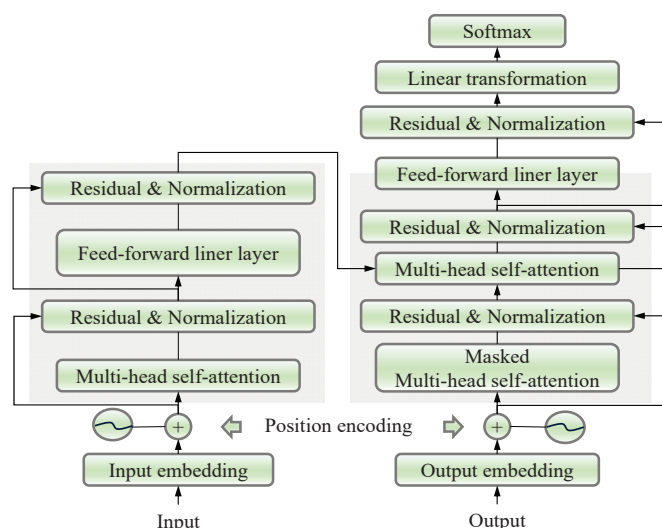
In this study, six satellite remote sensing precipitation products were selected, and accuracy assessment was conducted based on continuity and classification indicators to reveal the spatiotemporal error characteristics of satellite precipitation products. This study proposes a Transformer-based multi-source satellite remote sensing precipitation data fusion method that incorporates additional feature inputs of environmental factors, and conducts a comparative analysis between the fusion results and the precipitation observation data from ground rainfall stations. The method was applied to the fusion of daily precipitation data in the Jingle Basin to evaluate its effectiveness.

## 2. Materials and Methods

### 2.1. Transformer Model

Transformer is a model based on the encoder–decoder structure with the attention mechanism [30]. The basic structure of the Transformer network is shown in Figure 1. The core structural components of the Transformer deep learning model include an input pro-

cessing module, an encoder, a decoder, and an output layer which work synergistically to achieve efficient modeling of long-sequence data. The input processing module consists of an embedding layer and positional encoding: the embedding layer converts discrete inputs (such as text tokens and time-series data points) into continuous vectors of fixed dimensions, while positional encoding generates vectors through sine and cosine functions to inject temporal information, making up for the lack of temporal perception in the self-attention mechanism. The encoder is formed by stacking multiple identical sublayers, each containing a multi-head self-attention mechanism and a feed-forward neural network; the former captures multi-dimensional sequence dependencies in parallel, while the latter enhances nonlinear fitting capabilities through linear transformations and activation functions, with each sublayer equipped with residual connections and layer normalization to ensure stable training. On the basis of the encoder's sublayers, the decoder adds a masked multi-head self-attention mechanism to mask future position information and avoid information leakage, and simultaneously associates the global features output by the encoder through encoder–decoder attention. The output layer projects the decoder's output into the target space via linear mapping and then outputs prediction results through the Softmax function. With the advantages of parallel computing and global dependency capture, the overall architecture significantly improves the efficiency and accuracy of long-sequence modeling [31]. After the input sequence undergoes embedding mapping and position encoding, the Transformer model can be used for tasks such as translation. However, there are still challenges in applying the Transformer model to regression prediction tasks for meteorological elements [32]. This is because meteorological elements exhibit strong spatiotemporal dependencies, which the original Transformer model struggles to handle. Therefore, it is necessary to incorporate domain knowledge in meteorology to improve the model.



**Figure 1.** Basic Structure of Transformer Network.

In the encoder, the process of encoding is calculated as Equations (1) and (2). The input data first goes through an embedding layer consisting of a single-layer fully connected network. In order to obtain information about the temporal association of the data, the network uses positional encoding to calculate the relative position of data in the sequence [30].

$$PE(pos, 2i) = \sin\left(\frac{pos}{10^{8i/d_{model}}}\right) \quad (1)$$

$$PE(pos, 2i + 1) = \cos\left(\frac{pos}{10^{8i/d_{model}}}\right) \quad (2)$$

where  $pos$  is the position and  $i$  is the dimension, this represents the position encoding vector of the  $i$ -th dimension at the  $pos$  position.

In addition, when the Transformer model is used for translation tasks, it is necessary to use positional encoding to capture and add the positional information of data points in the sequence to the input vectors [33]. Otherwise, it will affect the translation results. However, multi-source precipitation sequence data is a type of multivariate spatiotemporal sequence data. It uses multivariate inputs to perform rolling regression on individual precipitation values. In this task, the positional encoding and decoder structure are not needed because there is no need to output a sequence result from the decoder. The adopted Transformer model consists of an encoder and a liner layer directly, where the observed values of multi-source precipitation products at the same location and related environmental factors are used as inputs to the encoder part, and the results are directly outputted through the liner layer [34].

## 2.2. Transformer-Based Multi-Source Precipitation Fusion Model

The adopted Transformer model mainly consists of multiple encoding layers and liner layers. The encoding layers include multi-head attention layers, residual and normalization layers, and convolutional layers. The convolutional layers consist of two layers of one-dimensional convolution. The input data includes multi-source satellite precipitation products and environmental factor data (such as Digital Elevation Model, Normalized Difference Vegetation Index, and temperature) that have spatiotemporal dependencies with precipitation. The convolutional layers extract deep features from the data. They reduce the training parameters in a single convolutional layer through weight sharing and sparse connections, improving the efficiency during forward propagation. Dropout layers are added for regularization to prevent overfitting. The output layer uses a global average pooling layer to globally average the vectorized data, transforming the multidimensional input into one-dimensional. Then, the liner layers output the fusion results of the multi-source precipitation inputs. The framework of the Transformer model for multi-source satellite precipitation fusion is shown in Figure 2. The model adopts the multi-head scaled dot-product attention mechanism from the original Transformer. The attention mechanism can be described as mapping a Query ( $Q$ ) and a set of Key-Value pairs ( $K$ - $V$ ) to an output, where the output is the weighted sum of  $V$ , with weights determined by the similarity between  $Q$  and  $V$ . The final output is obtained by fusing the outputs of  $h$  ( $h$  represents the number of heads in the multi-head attention mechanism) scaled dot-product attentions, with each attention output referred to as a head. Figure 2 includes the structure of multi-head scaled dot-product attention, whose mathematical principles are as follows [27]:

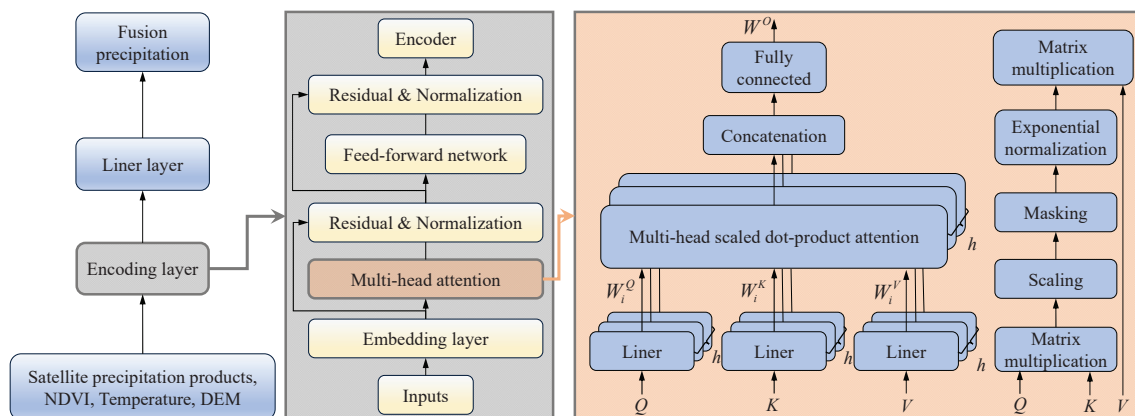
$$Attention(Q, K, V) = \text{soft max}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

$$MultiHead(Q, K, V) = \text{Concat}(head_1, \dots, head_h)W^O \quad (4)$$

where  $W_i^Q \in R^{d_{model} \times dk}$ ,  $W_i^K \in R^{d_{model} \times dk}$ ,  $W_i^V \in R^{d_{model} \times dk}$ ,  $W^O \in R^{hdv \times d_{model}}$ ,  $h$  represents the number of attention heads, with set  $d_k = d_{model}/h$ ,  $d$  is the dimension of the vectors, and  $T$  denotes the transpose of a matrix.

Before training the Transformer model, it is necessary to set hyperparameters that can affect the training results. Through multiple experiments, it was found that four parameters, namely the dimension of attention heads, the number of attention heads, the dimension of liner layers, and the number of encoding blocks, have a significant impact on the training performance of the Transformer model. The hyperparameter combinations were optimized using the grid search method with the *NSE* (Nash–Sutcliffe efficiency coefficient) as the

objective function [35]. Through multiple experiments, it was observed that the dimension and number of attention heads have a greater impact on the model, while the number of encoding blocks mainly affects the training time. Additionally, a comparison experiment was conducted using a Transformer model that considered environmental factors as covariates and another model that did not consider them. The model parameters were set as follows: number of encoding blocks = 4, dimension of attention heads = 64, number of attention heads = 16, dimension of liner layers = 64, learning rate = 0.001, loss function = MAE (Mean Absolute Error), optimizer = Adam, dropout rate = 0.1, and training window size = 1.



**Figure 2.** Transformer model for multi-source satellite precipitation fusion.

Building a multi-source precipitation fusion model based on the deep learning framework TensorFlow. The specific structure and process of the model are as follows:

(1) Input Data: The input precipitation data can be represented as  $U = [G, P, S]$ , where  $G$  represents the observed precipitation data,  $P$  is the matrix of multi-source satellite precipitation data  $[P_1, P_2, \dots, P_n]$  ( $n$  represents the number of different types of satellite precipitation products, which is 6), and  $S$  is the matrix of environmental factor data  $[NDVI, T_e, E]$  (where  $NDVI$  represents the vegetation coverage index,  $T_e$  represents the temperature, and  $E$  represents the elevation at the location). The input shape is  $[(None, 1, 7)]$  or  $[(None, 1, 10)]$ .

(2) Multi-head Attention Layer: The multi-head attention layer maps the input data to multiple different subspaces. The original high-dimensional feature vectors are split into multiple mutually independent low-dimensional feature spaces through different linear transformation matrices. The output shape is  $[(None, 1, 14)]$  or  $[(None, 1, 20)]$ .

(3) Residual and Normalization: The original input is added to the output of the previous step to prevent gradient explosion. The output shape is  $[(None, 1, 14)]$  or  $[(None, 1, 20)]$ .

(4) Convolutional Layer and Dropout Layer: Two layers of one-dimensional convolutional layers are used instead of liner layers for convolutional calculations. This is performed for nonlinear mapping, and Dropout is added to prevent overfitting. The output shape is  $[(None, 1, 14)]$  or  $[(None, 1, 20)]$ .

(5) Global Average Pooling and Liner Layer: Data dimensionality reduction is performed through global average pooling, and regularization is applied to the entire network structure to prevent overfitting. The result is output through a liner layer. The output shape is  $[(None, 1)]$ .

### 2.3. Accuracy Evaluation Indicators

To comprehensively reflect the accuracy characteristics of satellite precipitation products on a daily scale, both continuity and categorical indicators are used to evaluate various



satellite remote sensing precipitation products. The continuity indicators are used to describe the overall agreement and error characteristics of satellite remote sensing precipitation products with daily precipitation amount and processes. They include correlation coefficient (*CC*), bias (*BIAS*), mean absolute error (*MAE*), root mean square error (*RMSE*), and Nash–Sutcliffe efficiency coefficient (*NSE*) [35,36]. The categorical indicators reflect the estimation ability of satellite precipitation products for the occurrence of daily precipitation events. They include bias score (*BS*), false alarm rate (*FAR*), Probability of Detection (*POD*), and critical success index (*CSI*) [37]. Based on 24 h of rainfall, the precipitation is divided into five levels: no rain (precipitation < 0.1 mm), light rain (0.1–9.9 mm), moderate rain (10–24.9 mm), heavy rain (25–49.9 mm), and torrential rain ( $P \geq 50$  mm). The detection ability of each satellite remote sensing precipitation product for different precipitation levels is considered. The calculation methods for each indicator are as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - S_i| \quad (5)$$

$$BIAS = \frac{S - P}{P} \quad (6)$$

$$RMSE = \left[ \frac{1}{n} \sum_{i=1}^n (S_i - P_i)^2 \right]^{1/2} \quad (7)$$

$$CC = \frac{\sum_{i=1}^n (P_i - \bar{P})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (P_i - \bar{P})^2} \times \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}} \quad (8)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (P_i - S_i)^2}{\sum_{i=1}^n (P_i - \bar{P})^2} \quad (9)$$

where  $P_i$  represents the observed precipitation at the ground-based rainfall station for the  $i$ -th period (mm),  $S_i$  represents the estimated precipitation from satellite remote sensing for the  $i$ -th period (mm), and  $\bar{P}$  and  $\bar{S}$  represent the average values of observed precipitation at the ground-based rainfall station and estimated precipitation from satellite remote sensing, respectively.

$$BS_i = \frac{H + M}{H + F} \quad (10)$$

$$FAR = \frac{F}{H + F} \quad (11)$$

$$CSI = \frac{H}{H + M + F} \quad (12)$$

$$POD = \frac{H}{H + M} \quad (13)$$

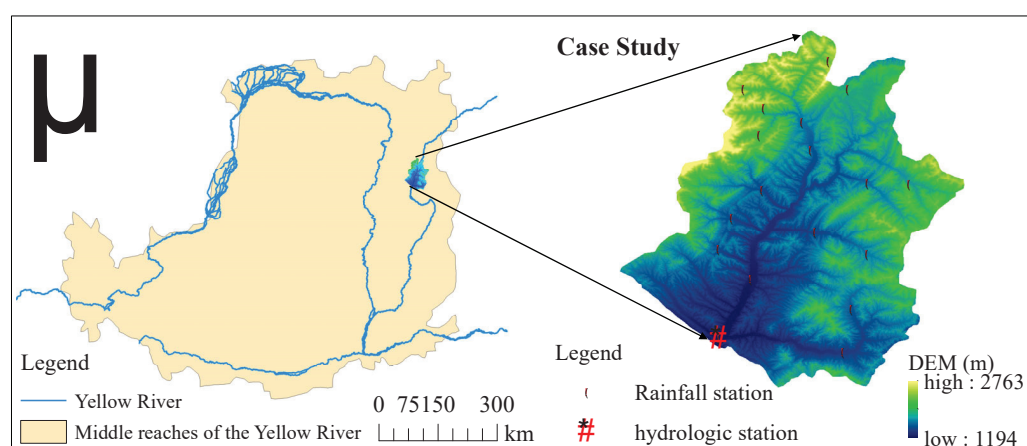
Among the above:  $H$  represents the number of accurately captured precipitation events, which refers to the frequency of precipitation events detected by both satellite remote sensing precipitation data and ground-based rainfall station observations.  $M$  represents the number of missed precipitation events, which refers to the frequency of precipitation events detected by ground-based rainfall station observations but were not detected by satellite precipitation data.  $F$  represents the number of false precipitation events, which refers to the frequency of precipitation events detected by satellite remote sensing precipitation data but were not detected by ground-based rainfall station observations. It should be noted that based on the measured precipitation data, precipitation greater than 0.1 mm is considered as rainfall. The  $BS$  is a parameter that measures the deviation of

satellite monitoring of precipitation events. The ideal value of *BS* is 1. When the value is greater than 1, it indicates positive bias, and when it is less than 1, it indicates negative bias. The value range of three indicators, *POD*, *FAR*, and *CSI*, is 0 to 1. A higher value of both *POD* and *CSI* and a lower value of *FAR* indicate a higher ability of satellite remote sensing precipitation data to detect precipitation events, and vice versa.

### 3. Case Study

#### 3.1. Study Area Information

The Fen River is the second-largest tributary of the Yellow River. The geographical location of the Fen River Basin is between 110.5° to 113.5° E and 35.3° to 39° N. The main stream of the Fen River has a total length of 713 km, and the controlled basin area is 39,721 km<sup>2</sup>. The study area includes the Jingle Basin, which is located in the upper reaches of the Fen River. The total length of the river channel in the Jingle Basin is 83.9 km, with an average slope of 6.7‰, and the controlled basin area is 2799 km<sup>2</sup>. The location of the Jingle Basin and the monitoring sites are shown in Figure 3. The Jingle Basin is located in the Loess Plateau region in the middle reaches of the Yellow River, in a transitional zone between semi-humid and semi-arid climates. The average annual precipitation in the Jingle Basin is 498 mm (1971–2014), with a maximum annual precipitation of 749 mm and a minimum of 216 mm. The highest recorded peak discharge in history was 2267 m<sup>3</sup>/s, with an average annual maximum peak discharge of 596 m<sup>3</sup>/s. Precipitation is concentrated in the summer, with rainfall from June to September accounting for over 70% of the total annual precipitation. The basin experiences high temperatures and frequent heavy rainfall during the summer, leading to frequent flash floods and regional mountain flood disasters [38]. The Jingle Basin is one of the areas in the middle reaches of the Yellow River where the contradiction between water supply and demand is more prominent. Therefore, the Jingle Basin urgently needs accurate rainfall forecasts. Accurate rainfall forecasts can provide early warning of potential flood events, enabling management authorities and residents of the Jingle Basin to take necessary flood prevention measures and reduce flood losses. It can also assist decision-makers in better planning and managing water resources. Based on the forecasted rainfall, water reservoirs can be managed and water resources can be allocated in a rational manner to meet the needs of agriculture, industry, and residential water supplies.



**Figure 3.** Map of the Jingle River Basin and station locations (The map was created using ArcMap 10.1 software, and the drawing boundaries were sourced from <https://www.gpsov.com/cn2/>, accessed on 1 December 2024).

### 3.2. Data Source

(1) DEM Data: Download the 30 m resolution ASTER Global Digital Elevation Model (DEM) provided by Geospatial Data Cloud. Download link: <http://www.gscloud.cn/>.

(2) NDVI Data: Source from the China Resources and Environmental Science Data Center, the dataset provides spatial distribution of daily Normalized Difference Vegetation Index (NDVI) for China region. Download link: <https://www.resdc.cn/>.

(3) Hydrometeorological Data: Sourced from the China Meteorological Data Service Center, the dataset includes meteorological elements such as precipitation, evapotranspiration, and temperature. Download link: <http://data.cma.cn>.

(4) Observed Rainfall-Runoff Data: Sourced from hydrological data in the Yellow River Basin, this dataset includes daily rainfall data from 17 rainfall stations and daily runoff data from the Jingle Hydrological Station for the period from 2006 to 2014 (excluding missing daily runoff data for the year 2011). Additionally, annual runoff data from the Jingle Hydrological Station is available from 1971 to 2014.

(5) Multi-source Satellite Precipitation Products: Six satellite precipitation products, namely CHIRPS (<https://data.chc.ucsb.edu/products/CHIRPS-2.0/>, accessed on 1 December 2024), CMORPH-CRT (<https://www.ncei.noaa.gov/products/climate-data-records/precipitation-cmorph>, accessed on 1 December 2024), GSMaP\_Gauge (<https://sharaku.eorc.jaxa.jp/GSMaP/>, accessed on 1 December 2024), IMERG Final Run (<https://gpm.nasa.gov/data/directory>), PERSIANN-CCS-CDR (<https://chrsdata.eng.uci.edu/>), and MSWEP (<http://www.gloh2o.org/mswep/>, accessed on 1 December 2024), were selected for research. The suitability of these products as substitutes for ground-based data is discussed. Refer to Table 1 for their basic information.

**Table 1.** Basic Information of Six Multi-Source Satellite Precipitation Products.

| Satellite Precipitation Products | Source | Time Resolution | Space Resolution                   | Coverage Area                           | Start Time | Http Link                                                                                                                                                               |
|----------------------------------|--------|-----------------|------------------------------------|-----------------------------------------|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CHIRPS                           | USGS   | 1 d             | $0.05^{\circ} \times 0.05^{\circ}$ | $50^{\circ}\text{S}–50^{\circ}\text{N}$ | 1981–now   | <a href="https://data.chc.ucsb.edu/products/CHIRPS-2.0/">https://data.chc.ucsb.edu/products/CHIRPS-2.0/</a>                                                             |
| CMORPH-CRT                       | NOAA   | 3 h             | $0.25^{\circ} \times 0.25^{\circ}$ | $60^{\circ}\text{S}–60^{\circ}\text{N}$ | 1998–now   | <a href="https://www.ncei.noaa.gov/products/climate-data-records/precipitation-cmorph">https://www.ncei.noaa.gov/products/climate-data-records/precipitation-cmorph</a> |
| GSMaP_Gauge                      | JP/JMA | 1 h/3 d         | $0.1^{\circ} \times 0.1^{\circ}$   | $60^{\circ}\text{N}–60^{\circ}\text{S}$ | 2000–now   | <a href="https://sharaku.eorc.jaxa.jp/GSMaP/">https://sharaku.eorc.jaxa.jp/GSMaP/</a>                                                                                   |
| IMERG Final Run                  | NASA   | 0.5 h/1 d       | $0.1^{\circ} \times 0.1^{\circ}$   | $60^{\circ}\text{N}–60^{\circ}\text{S}$ | 2000–now   | <a href="https://gpm.nasa.gov/data/imerg">https://gpm.nasa.gov/data/imerg</a>                                                                                           |
| PERSIANN-CCS-CDR                 | CHRS   | 3 h/6 h/1 d     | $0.04^{\circ} \times 0.04^{\circ}$ | $60^{\circ}\text{S}–60^{\circ}\text{N}$ | 1983–now   | <a href="https://chrsdata.eng.uci.edu/">https://chrsdata.eng.uci.edu/</a>                                                                                               |
| MSWEP                            | EC/JRC | 3 h             | $0.1^{\circ} \times 0.1^{\circ}$   | Global                                  | 1979–now   | <a href="http://www.gloh2o.org/mswep/">http://www.gloh2o.org/mswep/</a>                                                                                                 |

Considering that the rainfall observations from ground-based rain gauge stations are available for the period from 2006 to 2014, we have selected six satellite-based rainfall products with the corresponding time-series lengths from 2006 to 2014. Additionally, the time resolution has been standardized to 24 h, which is equivalent to one day.

### 3.3. Model Parameter Settings

Model parameters are divided into hardware and software parameters. Hardware parameters refer to the computer environment configuration parameters on which the simulation in this paper relies. The computer environment configuration parameters are fixed, with an Intel® Core TM i9-13900 CPU (Santa Clara, CA, USA), an NVIDIA GeForce GTX 4060 GPU (Santa Clara, CA, USA), and 32 GB of RAM. Software parameters refer to the programming language environment and machine learning model hyperparameters. The programming language environment for program execution is Python 3.9. In this study, the *NSE* is used as the objective function, and the model hyperparameters are calibrated using the particle swarm optimization algorithm. The deep learning hyperparameters include the number of neurons, the number of hidden layers, the learning rate, the batch size, the time step, and the number of iterations, which are calibrated to values of 256, 2, 0.01, 64, 6, and 300, respectively.

## 4. Results

### 4.1. Evaluation of Single Satellite Remote Sensing Rainfall Products

#### 4.1.1. Daily Scale Evaluation

We used observed precipitation data from various stations within the Jingle River Basin to calculate the basin-average rainfall using the Thiessen polygon method. This basin-average rainfall was then used as the evaluation benchmark for satellite products. From the perspectives of continuity and classification indicators, the statistical results of rainfall evaluation for the six satellite remote sensing precipitation products are shown in Table 2. The results indicate that, considering all the indicators, GSMaP\_Gauge performed the best, leading in seven out of the nine indicators compared to other satellite rainfall products. MSWEP exhibited the highest *POD* rate and performed slightly lower than GSMaP\_Gauge in other indicators, followed by IMERG Final Run. CMORPH-CRT, CHIRPS, and PERSIANN-CCS-CDR have varying strengths and weaknesses, but all of them were at the bottom level. In terms of the *CC*, GSMaP\_Gauge reached a high value of 0.872, while PERSIANN-CCS-CDR was only 0.474, indicating a poor correlation with actual precipitation. Regarding the *MAE* and *RMSE* indicators, PERSIANN-CCS-CDR and CHIRPS exhibited the worst performance, with errors more than 50% higher than other satellite products. For the *NSE*, PERSIANN-CCS-CDR and CHIRPS were only 0.036 and 0.168, respectively, which were insufficient to reflect actual precipitation processes. Other precipitation products had *NSE* values greater than 0.5, which fall within an acceptable range.

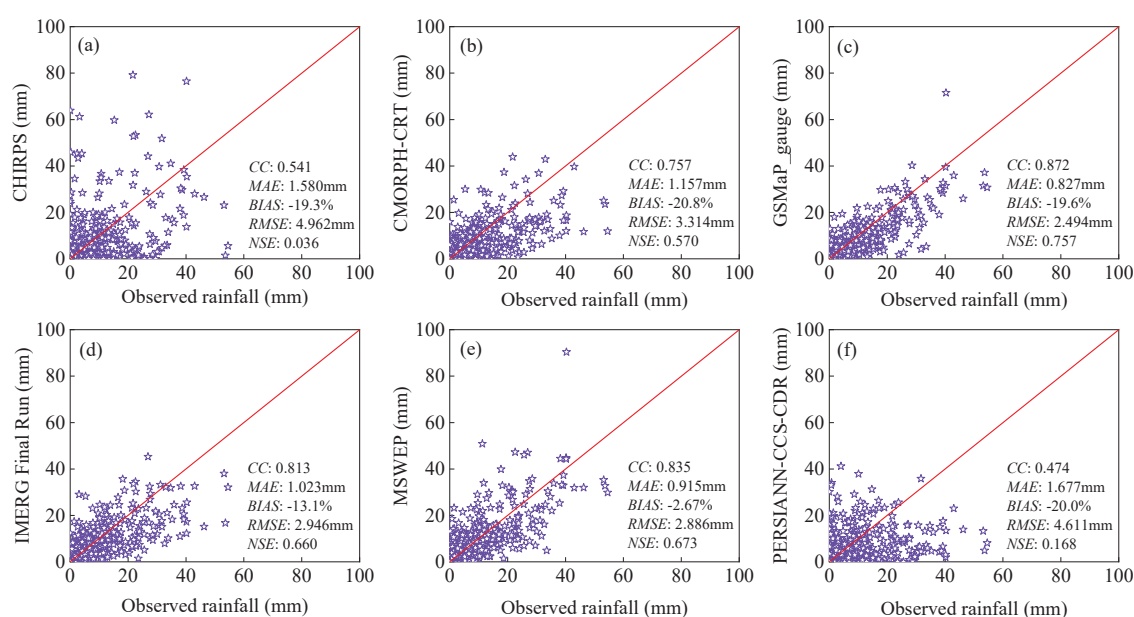
**Table 2.** Statistical Results of Surface Rainfall Evaluation for Satellite Remote Sensing Precipitation Products.

| Satellite Precipitation Products | Continuity Index |             |            |             |            | Classification Index |            |            |            |
|----------------------------------|------------------|-------------|------------|-------------|------------|----------------------|------------|------------|------------|
|                                  | <i>CC</i>        | <i>BIAS</i> | <i>MAE</i> | <i>RMSE</i> | <i>NSE</i> | <i>BS</i>            | <i>FAR</i> | <i>POD</i> | <i>CSI</i> |
| CHIRPS                           | 0.541            | −19.25%     | 1.580      | 4.962       | 0.036      | 0.766                | 0.364      | 0.487      | 0.381      |
| CMORPH-CRT                       | 0.757            | −20.79%     | 1.157      | 3.314       | 0.570      | 1.309                | 0.418      | 0.762      | 0.493      |
| GSMaP_Gauge                      | 0.872            | −19.64%     | 0.827      | 2.494       | 0.757      | 0.924                | 0.156      | 0.780      | 0.681      |
| IMERG Final Run                  | 0.813            | −13.12%     | 1.023      | 2.946       | 0.660      | 1.237                | 0.343      | 0.812      | 0.570      |
| PERSIANN-CCS-CDR                 | 0.474            | −19.96%     | 1.677      | 4.611       | 0.168      | 1.225                | 0.435      | 0.692      | 0.451      |
| MSWEP                            | 0.835            | −2.669%     | 0.915      | 2.886       | 0.673      | 1.116                | 0.243      | 0.845      | 0.665      |

Regarding the classification indicators, apart from the *POD* indicator where MSWEP performed the best with a value of 0.845, GSMaP\_Gauge consistently exhibited the best performance with values of 0.924 for *BS*, 0.156 for *FAR*, and 0.681 for *CSI*. Looking at the two indicators, *POD* and *CSI*, which represented the satellite precipitation detection

capability, CHIRPS had the lowest performance with values of 0.487 and 0.381, respectively, indicating the poorest detection capability. PERSIANN-CCS-CDR followed with values of 0.692 and 0.451, and CMORPH-CRT showed relatively improved detection capability with values of 0.762 and 0.493 compared to CHIRPS. IMERG Final Run achieved a *CSI* value of 0.570, indicating over 50% monitoring accuracy. In terms of the *FAR*, CMORPH-CRT and PERSIANN-CCS-CDR had higher values, both exceeding 0.4, indicating a false alarm rate of over two-fifths throughout the year. CHIRPS and IMERG Final Run had *FAR* values of 0.364 and 0.343, respectively, meaning the false alarm rate was close to one-third. Regarding the *BS*, CHIRPS and GSMaP\_Gauge underestimated the occurrence of precipitation events, while others showed a tendency of excessive monitoring. However, overall, GSMaP\_Gauge was closer to the actual precipitation situation. Considering all the indicators, the ranking of inversion accuracy for each satellite remote sensing precipitation product was as follows: GSMaP\_Gauge > MSWEP > IMERG Final Run > CMORPH-CRT > PERSIANN-CCS-CDR > CHIRPS.

Scatter plots of the observed calculated surface rainfall and satellite remote sensing precipitation estimates for the six products in the Jingle Basin were plotted, as shown in Figure 4. Based on the scatter distribution of precipitation in Figure 4, it can be observed that GSMaP-Gauge, IMERG-Final Run, and MSWEP exhibited good correlation with the observed precipitation data, with their points closely distributed around the 45-degree line. On the other hand, the precipitation estimates of the remaining products showed more scattered distribution. Specifically, PERSIANN-CCS-CDR underestimated moderate to heavy rainfall events, while CHIRPS exhibited the highest level of divergence and high uncertainty, indicating significant systematic bias.



**Figure 4.** Scatter plot of satellite precipitation products and observed data at daily scale.

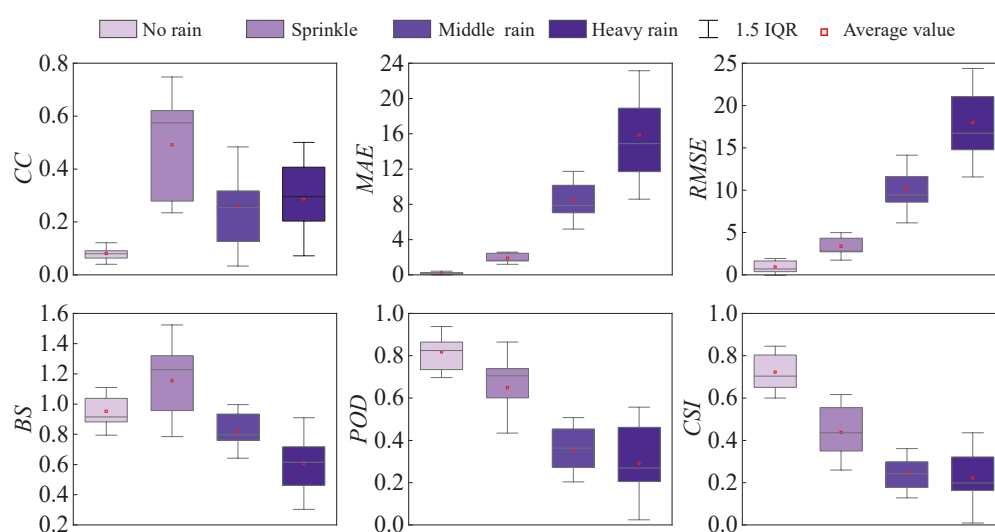
#### 4.1.2. Evaluation of Different Precipitation Levels

According to the regional meteorological industry standard “Grade of Precipitation” (GB/T 28592-2012 [39]), precipitation was divided into five levels: no rain or trace rainfall (0~0.1 mm), light rain (0.1~9.9 mm), moderate rain (10~24.9 mm), heavy rain (25~49.9 mm), and torrential rain ( $P \geq 50$  mm). Due to the low frequency of torrential rainfall events in the Jingle Basin, they were not discussed separately. Figure 5 presents the variations in various indicators for each satellite precipitation product within different precipitation threshold



ranges. Since dividing the data by precipitation thresholds has a significant impact on *NSE* and *FAR*, only *CC*, *MAE*, *RMSE*, *BS*, *POD*, and *CSI* indicators are shown.

Analyzing the results from Figure 5 and Table 3, it can be observed that the overall *CC* between satellite precipitation data and observed precipitation data decreased as the rainfall amount increased. Specifically, for no rainfall and light rainfall conditions, the correlation coefficients were 0.08 and 0.49, respectively. As the rainfall amount increased, the values of the correlation coefficients became lower, with the coefficients only reaching 0.26 and 0.29 under the condition of high precipitation magnitude. This indicates that the accuracy of satellite precipitation in capturing rainfall decreased with increasing rainfall amount, and the low correlation in no rainfall conditions was influenced by the prevalent zero values. Additionally, within a certain range of rainfall, the correlation between satellite precipitation data and observed precipitation data strengthened with increasing rainfall amount. However, when the rainfall amount exceeded a certain threshold, satellite precipitation data could not accurately reflect the actual precipitation, leading to a decrease in correlation. This can be attributed to the influence various factors had on satellite precipitation data, such as cloud depth and cloud top temperature. As the rainfall level increased, the values of *MAE* and *RMSE* continued to rise, indicating an increasing deviation between the estimated values and the actual observed values. For no rainfall and light rainfall conditions, the *RMSE* values were 0.93 and 3.36, respectively, while for moderate and heavy rainfall conditions, the *RMSE* values increased to 10.15 and 17.98, respectively.



**Figure 5.** Accuracy evaluation of satellite precipitation products for different rainfall levels.

Moreover, a comprehensive analysis based on *BS*, *POD*, and *CSI* indicated that these indicators decrease with increasing rainfall levels. For instance, under light rainfall conditions, the average values of *BS*, *POD*, and *CSI* were 1.03, 0.54, and 0.37, respectively. However, under heavy rainfall levels, these values decreased to 0.61, 0.29, and 0.22, respectively. This suggests that the accuracy of satellite precipitation was lower for higher rainfall levels compared to lower rainfall levels. The *BS* showed a decreasing trend with increasing rainfall, indicating an increasing discrepancy between the estimated precipitation and the observed precipitation frequency, leading to a more severe negative bias. The *POD* and *CSI* also showed a decreasing trend with increasing rainfall, indicating weaker satellite capture capability for extreme rainfall events. Considering these indicators collectively, satellite precipitation products provided relatively accurate estimates for small to moderate rainfall events, but they may have had some errors and limitations during heavy rainfall

events. Additionally, the trends of these indicators can contribute to the evaluation and improvement of satellite precipitation data.

**Table 3.** Statistical indicators of satellite precipitation products for different rainfall levels.

| Rainfall Level | Satellite Precipitation Products | CC   | BIAS    | MAE   | RMSE  | BS   | POD  | CSI  |
|----------------|----------------------------------|------|---------|-------|-------|------|------|------|
| No rain        | CHIRPS                           | 0.08 | 38.61%  | 0.26  | 1.85  | 1.12 | 0.86 | 0.69 |
|                | CMORPH-CRT                       | 0.05 | 35.51%  | 0.23  | 0.83  | 0.85 | 0.73 | 0.65 |
|                | GSMaP_Gauge                      | 0.09 | 7.066%  | 0.05  | 0.34  | 1.04 | 0.93 | 0.84 |
|                | IMERG Final Run                  | 0.08 | 22.81%  | 0.15  | 0.56  | 0.88 | 0.79 | 0.72 |
|                | PERSIANN-CCS-CDR                 | 0.06 | 65.96%  | 0.43  | 1.64  | 0.89 | 0.73 | 0.64 |
|                | MSWEP                            | 0.13 | 10.84%  | 0.08  | 0.37  | 0.94 | 0.87 | 0.80 |
| Light rain     | CHIRPS                           | 0.27 | −12.83% | 2.50  | 5.10  | 0.78 | 0.38 | 0.27 |
|                | CMORPH-CRT                       | 0.56 | −14.49% | 1.65  | 2.71  | 1.42 | 0.69 | 0.40 |
|                | GSMaP_Gauge                      | 0.62 | −12.93% | 1.48  | 2.42  | 0.96 | 0.72 | 0.58 |
|                | IMERG Final Run                  | 0.59 | 0.27%   | 1.65  | 2.77  | 1.31 | 0.74 | 0.47 |
|                | PERSIANN-CCS-CDR                 | 0.28 | −1.40%  | 2.44  | 4.31  | 1.32 | 0.60 | 0.35 |
|                | MSWEP                            | 0.63 | 10.99%  | 1.58  | 2.84  | 1.15 | 0.77 | 0.55 |
| Middle rain    | CHIRPS                           | 0.13 | −37.67% | 11.85 | 14.63 | 0.65 | 0.21 | 0.15 |
|                | CMORPH-CRT                       | 0.21 | −34.69% | 8.45  | 9.98  | 0.79 | 0.32 | 0.22 |
|                | GSMaP_Gauge                      | 0.51 | −27.81% | 5.98  | 7.12  | 0.76 | 0.46 | 0.36 |
|                | IMERG Final Run                  | 0.30 | −25.84% | 7.28  | 8.58  | 0.93 | 0.40 | 0.26 |
|                | PERSIANN-CCS-CDR                 | 0.09 | −49.67% | 10.16 | 11.63 | 0.81 | 0.27 | 0.18 |
|                | MSWEP                            | 0.32 | −13.98% | 7.04  | 8.94  | 0.98 | 0.45 | 0.30 |
| Heavy rain     | CHIRPS                           | 0.37 | −38.23% | 18.90 | 21.06 | 0.72 | 0.28 | 0.20 |
|                | CMORPH-CRT                       | 0.20 | −45.38% | 16.03 | 17.69 | 0.46 | 0.21 | 0.16 |
|                | GSMaP_Gauge                      | 0.41 | −25.12% | 11.02 | 13.65 | 0.69 | 0.51 | 0.43 |
|                | IMERG Final Run                  | 0.08 | −37.98% | 13.71 | 15.81 | 0.54 | 0.26 | 0.20 |
|                | PERSIANN-CCS-CDR                 | 0.22 | −72.77% | 23.83 | 24.92 | 0.33 | 0.03 | 0.02 |
|                | MSWEP                            | 0.44 | −16.06% | 11.71 | 14.78 | 0.90 | 0.46 | 0.32 |

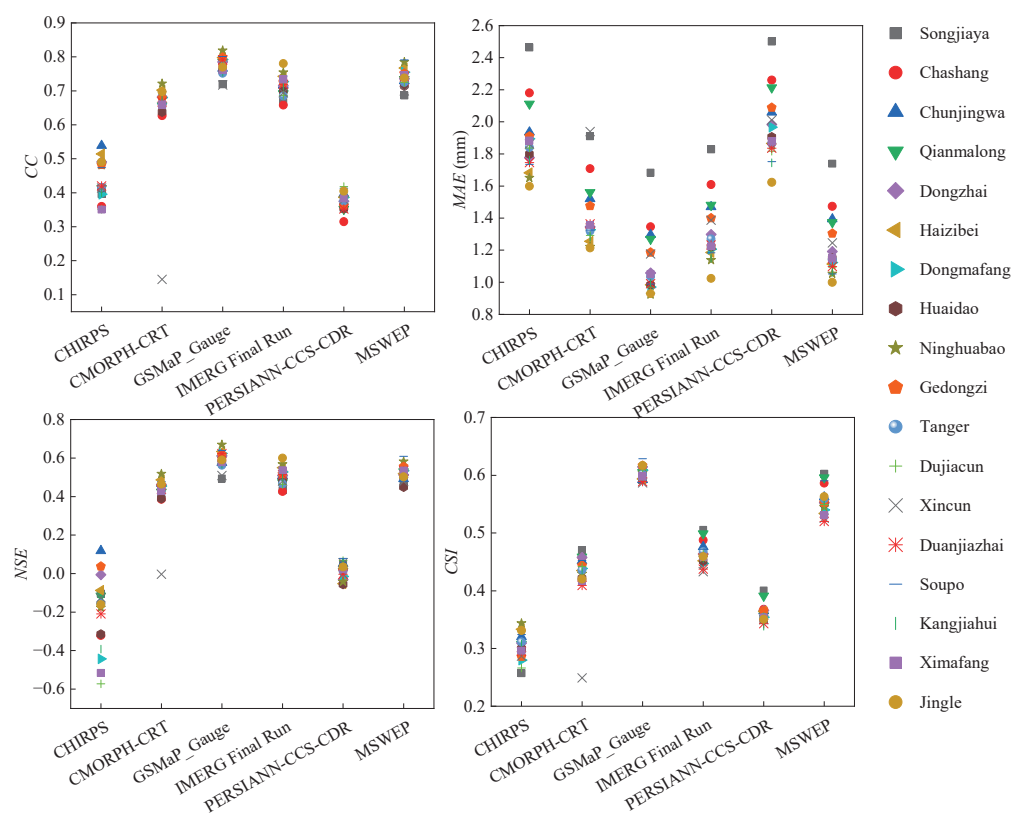
#### 4.1.3. Site Evaluation Results

There were significant differences in the accuracy of satellite precipitation products at different stations which may be attributed to various factors, such as regional precipitation characteristics, topography and elevation, and the sensitivity of different satellite sensors. Exploring the differences in retrieval accuracy of different satellite precipitation products in different regions can help to identify their strengths and limitations more accurately. To further analyze the spatial distribution characteristics of satellite errors, Figure 6 presents scatter plots of the four evaluation indicators at each station. From Figure 6, it can be observed that there was significant variability in the evaluation indicators for different products, indicating high uncertainty in satellite precipitation data. Specifically, the CC of the six products were mostly above 0.5 at most stations, with the best performance observed at the Ninghuabao station. The MAE of different satellite remote sensing precipitation products showed the highest level of dispersion, with the Songjiaya station exhibiting significantly higher precipitation estimation errors than other stations. For the NSE, CHIRPS was generally below 0 at most stations. The CSI indicator of GSMaP\_Gauge was the highest overall, with relatively small differences and high concentration among different stations.

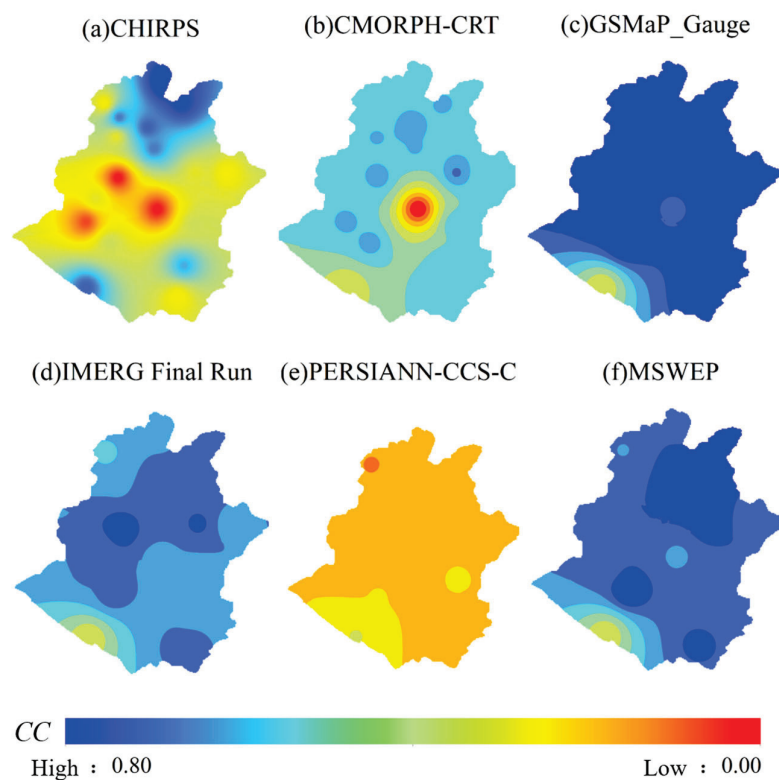
#### 4.2. Spatial Distribution of Error Characteristics

Taking the CC from the continuity evaluation indicators as an example, the spatial distribution characteristics can be discussed. From Figure 7, it can be observed that all six products exhibited higher accuracy in the northeast region, while lower accuracy was observed in the southwest region. Furthermore, the southwest region was closer to the

center of precipitation extremes, indicating a more pronounced deficiency in detecting precipitation extremes by satellite precipitation products.

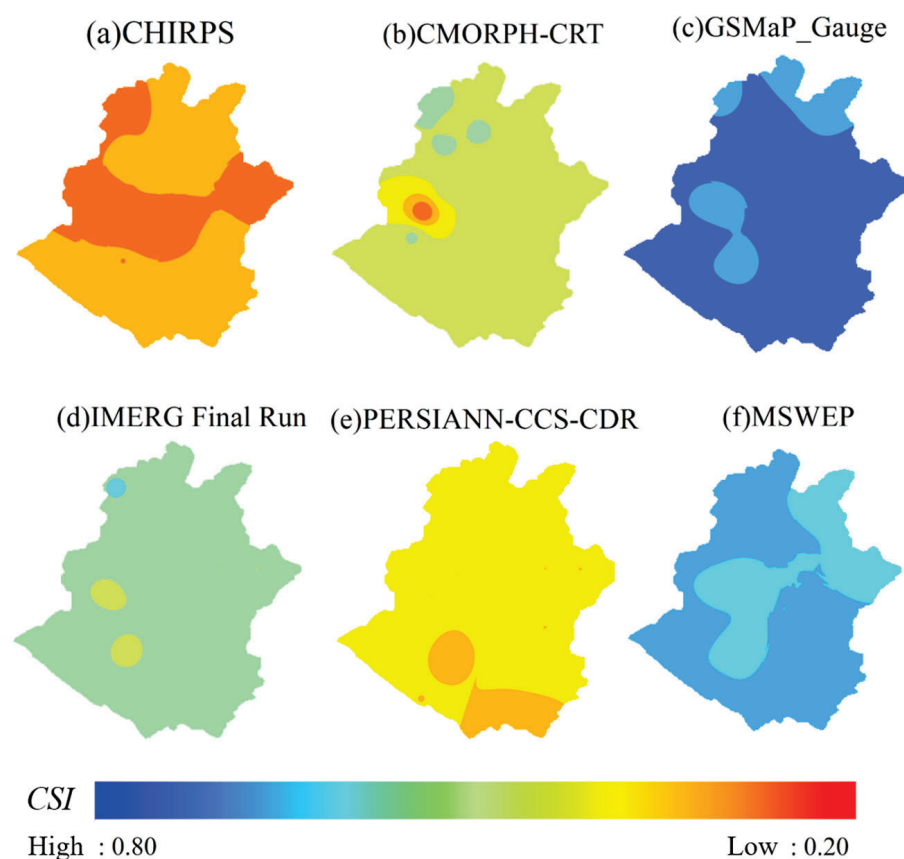


**Figure 6.** Distribution of accuracy indicators for satellite precipitation products at different stations.



**Figure 7.** Distribution map of CC of six satellite remote sensing precipitation products for daily precipitation estimation.

Figure 8 presents the spatial distribution characteristics of the CSI evaluation indicator, which reflects the satellite detection capability. GsMaP-Gauge has a CSI value greater than 0.6 in most grid cells. The lowest CSI values for the six satellite precipitation products were located in areas with significant elevation differences in the study area, indicating that the satellite's ability to capture precipitation events in mountainous regions still needs improvement.

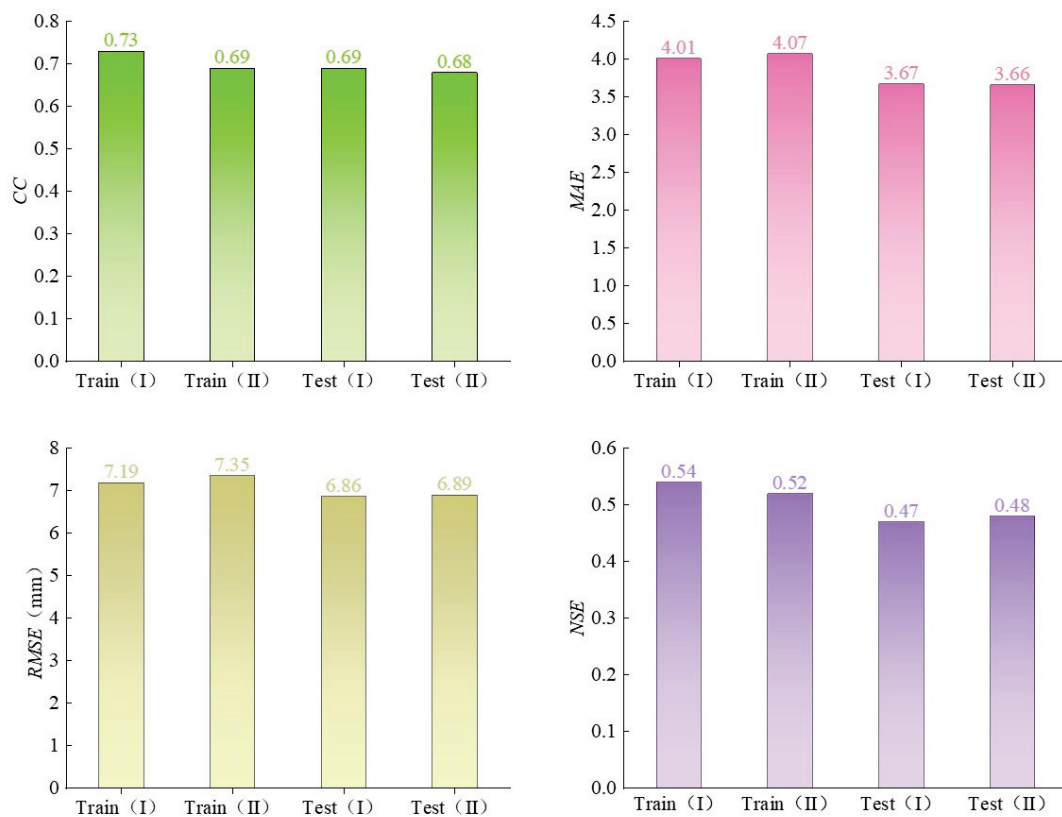


**Figure 8.** Distribution map of CSI for daily precipitation estimation of six satellite remote sensing precipitation products.

#### 4.3. Evaluation of Multi-Source Precipitation Data Fusion Products

Based on the discussion in the previous text, the Transformer-based multi-source precipitation fusion model was used to study the application effect of the Transformer model in precipitation estimation in the Jingle Basin. The model utilized six high-quality satellite remote sensing precipitation products (CHIRPS, CMORPH-CRT, GsMaP\_Gauge, IMERG Final Run, PERSIANN-CCS-CDR, and MSWEP), ground observation precipitation data in the Jingle Basin, and environmental factor datasets. To avoid the impact of matrix sparsity caused by zero values, only daily data from four or more of the six satellite remote sensing precipitation products that simultaneously monitor precipitation events were selected for training. Two schemes were adopted in the study: Scheme I considered environmental factors in the Transformer model, while Scheme II only considered satellite remote sensing precipitation products in the Transformer model. The fusion effect was evaluated at both daily scales, and the corresponding indicators are shown in Figure 9. Figure 9 presents the fusion results for 17 stations, divided into training and validation sets for statistical analysis. It can be observed that the fusion effect of the training set and validation set of the Transformer model was consistent, indicating good robustness of the model. Moreover, the

scheme considering environmental factors had certain advantages compared to the scheme that only considered satellite remote sensing precipitation products.

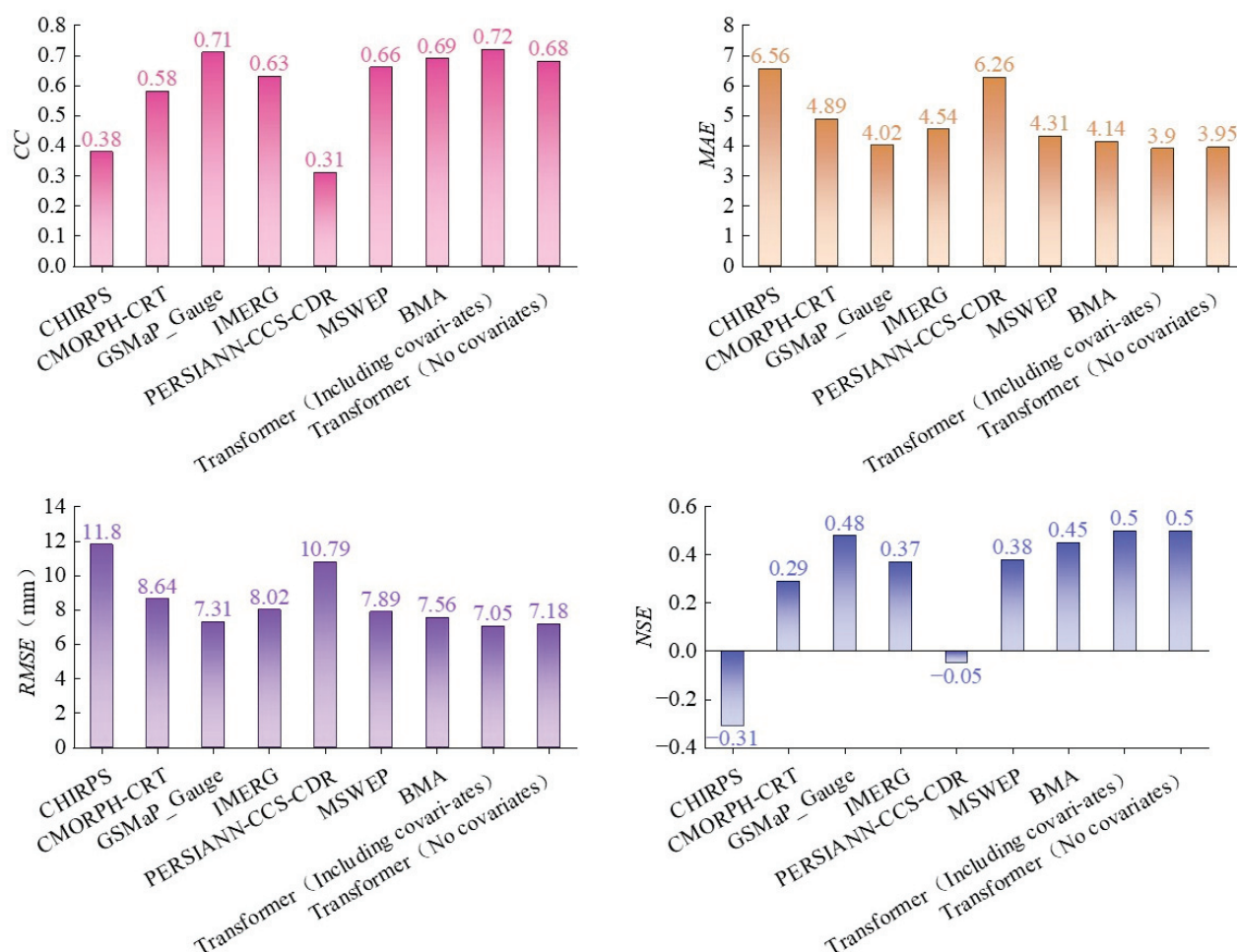


**Figure 9.** Accuracy indicators of fusion results.

Figure 10 comprehensively summarizes the accuracy improvement of the Transformer-based multi-source precipitation fusion algorithm on a daily scale for satellite remote sensing precipitation products. Figure 10 combines the results of the 17 stations into a single metric, without separating into training and validation sets. It compares and analyzes the changes in accuracy indicators of daily precipitation estimation before and after data fusion of satellite remote sensing precipitation products. From the evaluation results, it can be observed that the Transformer-based multi-source precipitation fusion algorithm had a robust improvement effect on precipitation estimation accuracy. Generally, after fusion using this algorithm, the MAE of daily precipitation estimation can be reduced to within 4.0, the RMSE can be reduced by more than 3% compared to single-source satellite remote sensing precipitation products, and the correlation coefficient can reach around 0.7. Furthermore, the Transformer model with covariates showed improvements in various indicators compared to the Transformer model without covariates, making it closer to actual precipitation. In order to further compare the fusion effect of the Transformer model, the Bayesian Model Averaging (BMA) [40] ensemble method was selected here as a control experiment to calculate the fusion effect evaluation indicators of six satellite remote sensing precipitation products. The principle of the BMA algorithm was to integrate the probability of multiple sources of rainfall results, thus generating unbiased integrated rainfall forecast results [41,42]. It was a representative machine learning model for rainfall fusion. From Figure 10, it can be seen that the fusion prediction accuracy of the Transformer model (without covariates) was slightly worse than that of the BMA model. However, the fusion accuracy of the Transformer model (with covariates) was better than that of the BMA model.

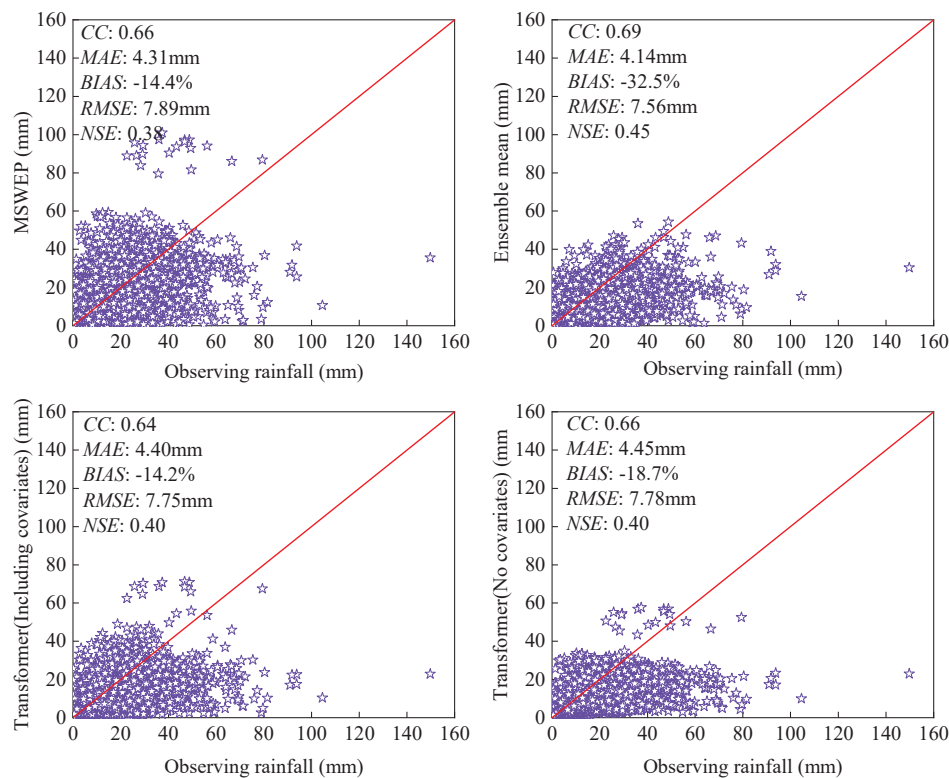


The fusion CC and NSE of the Transformer model (with covariates) were higher than those of the BMA model by 0.03 and 0.05, respectively, while the MAE and RMSE were lower by 0.24 and 0.51, respectively.



**Figure 10.** Statistical analysis of accuracy indicators at the daily scale for original satellite precipitation and fused precipitation.

Figure 11 shows scatter plots of satellite remote sensing precipitation products before and after fusion with the observed precipitation from ground rain gauge stations (as the ground truth). The results demonstrated that the fusion of multiple satellite remote sensing precipitation products greatly improved the accuracy of precipitation estimation. It successfully eliminated estimation biases, reduced the root mean square error, and significantly enhanced the correlation between precipitation estimates and observed precipitation from ground rain gauge stations. The results also indicated the presence of a “smoothing effect” in precipitation fusion estimates. For daily precipitation exceeding 50 mm, the fusion estimates tended to systematically underestimate the ground truth. This may be related to the non-normal distribution characteristics of extreme daily precipitation. Further discussion and analysis are needed in the future. Firstly, in this study, there were differences in the quality of multi-source satellite data, including differences in resolution, accuracy, and spatiotemporal coverage, which may affect the uncertainty of fusion results [7,34]. Secondly, the quantification of uncertainty in fusion results remains a challenge. Currently, there are various uncertainty quantification methods, such as probability-based methods and model output-based methods, but further research and improvement are still needed [11,37].



**Figure 11.** Comparison of daily precipitation estimation before and after fusion of satellite remote sensing precipitation products.

## 5. Conclusions

This study conducted an accuracy evaluation of multi-source satellite remote sensing precipitation products from both a daily scale and precipitation magnitude perspective. It also discussed the spatiotemporal variation characteristics of the inversion errors of satellite remote sensing precipitation products. Based on this, a Transformer-based multi-source satellite remote sensing precipitation data fusion method was proposed, which utilized ground rain gauge station precipitation observations, satellite remote sensing precipitation products, and relevant environmental factor data. The effectiveness of this method in fusing daily scale precipitation data in the Jingle River Basin was preliminarily analyzed. The main conclusions of this study were as follows:

(1) Compared with the observed precipitation data, the GSMap\_Gauge satellite remote sensing precipitation product showed the closest agreement with the observations, ranking first in all indicators except the Probability of Detection (POD). The MSWEP satellite remote sensing precipitation product followed in performance, while the CHIRPS satellite product performed the poorest. The spatial distribution characteristics of precipitation estimation errors for different products showed consistent changes with environmental factors, such as topography, vegetation coverage, and temperature.

(2) MSWEP exhibited a closer frequency of occurrence to station precipitation in different precipitation intervals. However, all satellite precipitation products showed a phenomenon of overestimation for low values and underestimation for high values. Moreover, as the precipitation magnitude increased, the accuracy of satellite detection decreased. Satellite remote sensing precipitation products had limited ability to characterize the spatial distribution of precipitation and exhibited similar error spatial distribution patterns. Mainly, they had insufficient capability to estimate precipitation extremes and

performed poorly in estimating precipitation in mountainous areas, resulting in low *POD* for high precipitation events and high *FAR* in low precipitation areas.

(3) The evaluation of daily precipitation data fusion showed that the Transformer-based multi-source precipitation data fusion algorithm could integrate multiple precipitation data sources effectively. This algorithm could reduce the relative bias of satellite observed precipitation estimates and improve their correlation with ground station measurements. The fusion scheme considering environmental factors such as daily temperature and NDVI performed better than the fusion scheme considering only precipitation characteristics, with higher precipitation estimation accuracy and fusion stability.

**Author Contributions:** Q.L.: Conceptualization, Methodology, Software, Reviewing and Editing. D.W.: Conceptualization, Methodology, Software, Reviewing and Editing. L.L.: Methodology, Software, Reviewing and Editing. C.H.: Methodology, Software and Editing. C.L.: Reviewing and Validation. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by Ministry of Science and Technology of the People's Republic of China (2023YFC3209303); National Natural Science Foundation of China, grant number (U2443215).

**Data Availability Statement:** The datasets analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

**Conflicts of Interest:** The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References

1. Zhao, N. A Novel Merging Method for Generating High-Quality Spatial Precipitation Information over Mainland China. *Water* **2023**, *15*, 575. [CrossRef]
2. Zhu, S.; Wei, J.; Zhang, H.; Xu, Y.; Qin, H. Spatiotemporal deep learning rainfall-runoff forecasting combined with remote sensing precipitation products in large scale basins. *J. Hydrol.* **2023**, *616*, 128727. [CrossRef]
3. Lyu, J.; Yin, S.; Sun, Y.; Wang, K.; Luo, P.; Meng, X. Flood Runoff Simulation under Changing Environment, Based on Multiple Satellite Data in the Jinghe River Basin of the Loess Plateau, China. *Remote Sens.* **2023**, *15*, 550. [CrossRef]
4. Goudenhoofdt, E.; Delobbe, L. Evaluation of radar-gauge merging methods for quantitative precipitation estimates. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 195–203. [CrossRef]
5. Chen, H.; Huo, R.; Zeng, Q.; Yang, W.; Chen, J.; Guo, S.; Xu, C. Comparative study on the influence of rain-gauge network on the uncertainty of hydrological modeling. *Adv. Water Sci.* **2019**, *30*, 34–44.
6. Sorooshian, S.; Hsu, K.L.; Gao, X.; Gupta, H.V.; Imam, B.; Braithwaite, D. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Am. Meteorol. Soc.* **2000**, *81*, 2035–2046. [CrossRef]
7. Joyce, R.J.; Janowiak, J.E.; Arkin, P.A.; Xie, P.P. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.* **2004**, *5*, 487–503. [CrossRef]
8. Huffman, G.J.; Adler, R.F.; Bolvin, D.T.; Gu, G.; Nelkin, E.J.; Bowman, K.P.; Hong, Y.; Stocker, E.F.; Wolff, D.B. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* **2007**, *8*, 38–55. [CrossRef]
9. Helmi, A.M.M.; Abdelhamed, M.S.S. Evaluation of CMORPH, PERSIANN-CDR, CHIRPS V2.0, TMPA 3B42 V7, and GPM IMERG V6 Satellite Precipitation Datasets in Arabian Arid Regions. *Water* **2023**, *15*, 92. [CrossRef]
10. Wei, Z.; Long, X.; Gai, Y.; Yang, Z.; Sui, X.; Chen, X.; Kan, G.; Fan, W.; Cui, Y. Evaluating Remote Sensing Precipitation Products Using Double Instrumental Variable Method. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
11. Zhan, C.; Chen, Y.; Yang, K.; Zhu, L.; Zhou, X.; Jiang, Y.; Ling, X.; Tian, J.; Wang, Y.; Li, X.; et al. First evaluation of GPM-Era satellite precipitation products with new observations on the western Tibetan Plateau. *Atmos. Res.* **2023**, *283*, 106559. [CrossRef]
12. Sharifi, E.; Steinacker, R.; Saghafian, B. Assessment of GPM-IMERG and Other Precipitation Products against Gauge Data under Different Topographic and Climatic Conditions in Iran: Preliminary Results. *Remote Sens.* **2016**, *8*, 135. [CrossRef]
13. Junfeng, L.; Rensheng, C.; Chuntan, H.; Chunping, T. Evaluating TRMM Mult-satellite Precipitation Analysis Using Gauge Precipitation and MODIS snow-cover Products. *Adv. Water Sci.* **2010**, *21*, 343–348.

14. Jiali, G.; Guangxu, D.; Xu, Y.; Zhengyang, T.; Hairong, Z.; Renwei, P.; Lingquan, D. Comparison of temporal and spatial accuracy of precipitation products from multi-source satellites in the Yangtze River basin. *South North Water Transf. Water Sci. Technol.* **2022**, *20*, 230–242+252.
15. Sreeparvathy, V.; Srinivas, V.V. A Bayesian Fuzzy Clustering Approach for Design of Precipitation Gauge Network Using Merged Remote Sensing and Ground-Based Precipitation Products. *Water Resour. Res.* **2022**, *58*, e2021WR030612. [CrossRef]
16. Yujia, S.; Zhongjing, W.; Ying, S. Evaluation of Haihe River basin precipitation resources based on multisource data fusion. *Adv. Water Sci.* **2022**, *33*, 602–613.
17. Lingjie, L.; Yintang, W.; Guoqiang, T.; Xuan, G.; Leizhi, W.; Qingfang, H. An innovative multi-source precipitation merging method with the identification of rain and no rain. *Adv. Water Sci.* **2022**, *33*, 780–793.
18. Sideris, I.V.; Gabella, M.; Erdin, R.; Germann, U. Real-time radar-rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland. *Q. J. R. Meteorol. Soc.* **2014**, *140*, 1097–1111. [CrossRef]
19. Ma, Y.; Hong, Y.; Chen, Y.; Yang, Y.; Tang, G.; Yao, Y.; Long, D.; Li, C.; Han, Z.; Liu, R. Performance of Optimally Merged Multisatellite Precipitation Products Using the Dynamic Bayesian Model Averaging Scheme Over the Tibetan Plateau. *J. Geophys. Res. Atmos.* **2018**, *123*, 814–834. [CrossRef]
20. Pan, M.; Sahoo, A.K.; Troy, T.J.; Vinukollu, R.K.; Sheffield, J.; Wood, E.F. Multisource Estimation of Long-Term Terrestrial Water Budget for Major Global River Basins. *J. Clim.* **2012**, *25*, 3191–3206. [CrossRef]
21. Shen, Y.; Zhao, P.; Pan, Y.; Yu, J. A high spatiotemporal gauge-satellite merged precipitation analysis over China. *J. Geophys. Res. Atmos.* **2014**, *119*, 3063–3075. [CrossRef]
22. Qingfang, H. Rainfall Spatial Estimation Using Multi-Source Information and Its Hydrological Application. Ph.D. Thesis, Tsinghua University, Beijing, China, 2013.
23. Bowler, N.E.; Pierce, C.E.; Seed, A.W. STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Q. J. R. Meteorol. Soc.* **2006**, *132*, 2127–2155. [CrossRef]
24. Liu, Z.; Yang, Q.; Shao, J.; Wang, G.; Liu, H.; Tang, X.; Xue, Y.; Bai, L. Improving daily precipitation estimation in the data scarce area by merging rain gauge and TRMM data with a transfer learning framework. *J. Hydrol.* **2022**, *613*, 128455. [CrossRef]
25. Shen, J.; Liu, P.; Xia, J.; Zhao, Y.; Dong, Y. Merging Multisatellite and Gauge Precipitation Based on Geographically Weighted Regression and Long Short-Term Memory Network. *Remote Sens.* **2022**, *14*, 3939. [CrossRef]
26. Jinglin, D.; Yayun, L.; Yanan, Y.; Weilan, Y. A Prediction of Precipitation Data Based on Support Vector Machine and Particle Swarm Optimization (PSO-SVM) Algorithms. *Algorithms* **2017**, *10*, 57. [CrossRef]
27. Bhuiyan, M.A.E.; Nikolopoulos, E.I.; Anagnostou, E.N.; Quintana-Segui, P.; Barella-Ortiz, A. A nonparametric statistical technique for combining global precipitation datasets: Development and hydrological evaluation over the Iberian Peninsula. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 1371–1389. [CrossRef]
28. Bhuiyan, M.A.E.; Nikolopoulos, E.I.; Anagnostou, E.N. Machine Learning-Based Blending of Satellite and Reanalysis Precipitation Datasets: A Multiregional Tropical Complex Terrain Evaluation. *J. Hydrometeorol.* **2019**, *20*, 2147–2161. [CrossRef]
29. Binh Thai, P.; Lu Minh, L.; Tien-Thinh, L.; Kien-Trinh Thi, B.; Vuong Minh, L.; Hai-Bang, L.; Prakash, I. Development of advanced artificial intelligence models for daily rainfall prediction. *Atmos. Res.* **2020**, *237*, 104845.
30. Xu, Y.; Lin, K.; Hu, C.; Wang, S.; Wu, Q.; Zhang, L.; Ran, G. Deep transfer learning based on transformer for flood forecasting in data-sparse basins. *J. Hydrol.* **2023**, *625*, 129956. [CrossRef]
31. Wenzhong, L.; Chengshuai, L.; Qiang, W.; Caihong, H.; Tianning, X.; Lu, T. Construction and Application of GRU-Transformer Flood Forecasting Model. *China Rural Water Hydropower* **2023**, *11*, 35–44. [CrossRef]
32. Yang, R.; Zheng, G.; Hu, P.; Liu, Y.; Xu, W.; Bao, A. Snowmelt Flood Susceptibility Assessment in Kunlun Mountains Based on the Swin Transformer Deep Learning Method. *Remote Sens.* **2022**, *14*, 6360. [CrossRef]
33. Li, Y.; Cheng, Z.; Wang, C.; Zhao, J.; Huang, L. RCCT-ASPPNet: Dual-Encoder Remote Image Segmentation Based on Transformer and ASPP. *Remote Sens.* **2023**, *15*, 379. [CrossRef]
34. Hong, Y.; Li, D.; Luo, S.; Chen, X.; Yang, Y.; Wang, M. An Improved End-to-End Multi-Target Tracking Method Based on Transformer Self-Attention. *Remote Sens.* **2022**, *14*, 6354. [CrossRef]
35. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. [CrossRef]
36. Hu, C.; Wu, Q.; Li, H.; Jian, S.; Li, N.; Lou, Z. Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation. *Water* **2018**, *10*, 1543. [CrossRef]
37. Jackson, E.K.; Roberts, W.; Nelsen, B.; Williams, G.P.; Nelson, E.J.; Ames, D.P. Introductory overview: Error metrics for hydrologic modelling—A review of common practices and an open source library to facilitate use and adoption. *Environ. Model. Softw.* **2019**, *119*, 32–48. [CrossRef]

38. Ran, G.; Jian, S.; Wu, Q.; Zhang, L.; Hu, C. Exploring the Dominant Runoff Processes in Two Typical Basins of the Yellow River, China. *Water* **2020**, *12*, 3055. [CrossRef]
39. GB/T 28592-2012; Grade of Precipitation. Standardization Administration of China: Beijing, China, 2012.
40. Rezaei, F.; Ghorbani, R.; Mahjouri, N. Improving Daily and Monthly River Discharge Forecasts using Geostatistical Ensemble Modeling. *Water Resour. Manag.* **2022**, *36*, 5063–5089. [CrossRef]
41. Sheikh, M.R.; Coulibaly, P. Review of Recent Developments in Hydrologic Forecast Merging Techniques. *Water* **2024**, *16*, 301. [CrossRef]
42. Wei, L.; Jiang, S.; Dong, J.; Ren, L.; Liu, Y.; Zhang, L.; Wang, M.; Duan, Z. Fusion of gauge-based, reanalysis, and satellite precipitation products using Bayesian model averaging approach: Determination of the influence of different input sources. *J. Hydrol.* **2023**, *618*, 129234. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



## Article

# Multi-Source Data Fusion and Heuristic-Optimized Machine Learning for Large-Scale River Water Quality Parameters Monitoring

Kehang Fang<sup>1,2</sup>, Feng Wu<sup>1,3,4</sup>, Xing Gao<sup>1,5</sup> and Zhihui Li<sup>1,3,4,\*</sup>

- <sup>1</sup> Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; 2252715@mail.dhu.edu.cn (K.F.); wufeng@igsnr.ac.cn (F.W.); gxing@igsnr.ac.cn (X.G.)  
<sup>2</sup> College of Mathematics and Statistics, Donghua University, Shanghai 200051, China  
<sup>3</sup> Key Laboratory of Land Surface Pattern and Simulation, Chinese Academy of Sciences, Beijing 100101, China  
<sup>4</sup> University of Chinese Academy of Sciences, Beijing 100049, China  
<sup>5</sup> State Key Laboratory of Resources and Environmental Information System, Chinese Academy of Sciences, Beijing 100101, China  
\* Correspondence: lizhihui@igsnr.ac.cn

## Highlights

### What are the main findings?

- Fused satellite, meteorology, land use, and landscape data can enhance the accuracy of water quality remote sensing inversion.
- Introduced WOA to auto-optimize ML models, enhancing inversion accuracy and stability.

### What are the implications of the main findings?

- Extended water quality inversion from urban to large-scale basins with diverse drivers.
- Integrated landscape pattern indices into models to reveal spatial pollution mechanisms.

## Abstract

Accurate and efficient surface water quality monitoring is essential for ecological protection and sustainable development. However, conventional monitoring methods, such as fixed-site observations, often suffer from spatial limitations and overlook crucial auxiliary variables. This study proposes an innovative modeling framework for large-scale river water quality inversion that integrates multi-source data—including Sentinel-2 imagery, meteorological conditions, land use classification, and landscape pattern indices. To improve predictive accuracy, three tree-based machine learning models (Random Forest, XGBoost, and LightGBM) were constructed and further optimized using the Whale Optimization Algorithm (WOA), a nature-inspired metaheuristic technique. Additionally, model interpretability was enhanced using SHAP (Shapley Additive Explanations), enabling a transparent understanding of each variable's contribution. The framework was applied to the Red River Basin (RRB) to predict six key water quality parameters: dissolved oxygen (DO), ammonia nitrogen (NH<sub>3</sub>-N), total phosphorus (TP), total nitrogen (TN), pH, and permanganate index (COD<sub>Mn</sub>). Results demonstrate that integrating landscape and meteorological variables significantly improves model performance compared to remote sensing alone. The best-performing models achieved R<sup>2</sup> values exceeding 0.45 for all parameters (DO: 0.70, NH<sub>3</sub>-N: 0.46, TP: 0.59, TN: 0.71, pH: 0.83, COD<sub>Mn</sub>: 0.57). Among them, WOA-optimized LightGBM consistently delivered superior performance. The study also confirms the feasibility of applying the models across the entire basin, offering a

transferable and interpretable approach to spatiotemporal water quality prediction in other large-scale or data-scarce regions.

**Keywords:** multi-source data fusion; water quality inversion; machine learning optimization; whale optimization algorithm; SHAP interpretability

## 1. Introduction

Surface water plays a vital role in sustaining ecological balance and supporting socio-economic development [1,2]. However, due to intensifying anthropogenic activities—such as rapid urban expansion and the discharge of untreated domestic wastewater—surface water quality has significantly deteriorated in recent years [3–5]. A substantial influx of pollutants into rivers, both directly and indirectly, has led to severe environmental issues, including water blackening, odor generation, and eutrophication. These deteriorations not only compromise the ecological integrity of river systems but also pose serious risks to water security for human consumption and environmental protection. Therefore, effective and timely monitoring of surface water quality is essential to ensure both environmental sustainability and public health [6–8].

Surface water quality is typically assessed using various physico-chemical parameters [9,10]. Conventional monitoring methods rely heavily on in situ data collection at fixed stations, which, while accurate, are spatially limited and often expensive to maintain [11,12]. In contrast, advances in remote sensing technologies have greatly improved access to high-resolution and high-frequency multispectral data, enabling broader and more efficient monitoring coverage [13]. Remote sensing-based water quality inversion methods, typically categorized as model-based or empirical approaches, have gained increasing attention [14–16]. In addition, machine learning methods have also been used recently due to their high prediction accuracy. Nevertheless, most current studies are concentrated in urban rivers or lakes, where water quality is relatively stable and less influenced by complex upstream interactions [17]. In contrast, large-scale water bodies—characterized by diverse hydrological conditions and heterogeneous human impacts—pose significant challenges for remote sensing inversion due to greater spatial variability.

To address these challenges, machine learning and deep learning techniques have increasingly been adopted for water quality assessment, given their capacity to capture nonlinear and complex relationships between input variables and water quality parameters [18–20]. However, most existing machine learning models primarily rely on spectral variables from remote sensing data, often neglecting crucial auxiliary variables such as meteorological conditions, land use, and watershed morphology [21,22]. Yet, environmental variables such as precipitation, temperature, and wind speed significantly influence pollutant transport and biogeochemical processes in river systems [23,24]. Similarly, land use changes—especially the expansion of impervious surfaces—directly alter runoff patterns and pollutant loads [25–27]. Moreover, landscape pattern indices, which capture spatial configuration, fragmentation, and connectivity, offer untapped potential for characterizing pollutant dispersion and retention at the watershed scale [28–30]. Despite growing evidence of their relevance, these indices have rarely been integrated into water quality model inversion.

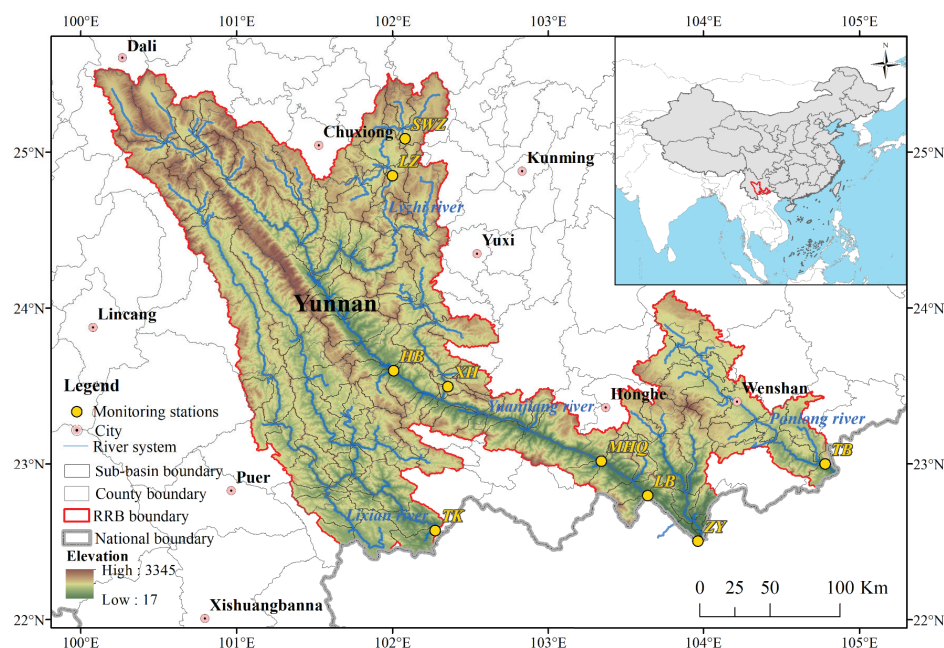
To fill this gap, this study proposes an innovative modeling framework that integrates multi-source data—including Sentinel-2 satellite imagery, meteorological variables, land use classification, and landscape pattern metrics—to predict key surface water quality parameters (DO, NH<sub>3</sub>-N, TP, TN, pH, and COD<sub>Mn</sub>) in the Red River Basin (RRB). We employ three

advanced tree-based machine learning algorithms—Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM)—and enhance their performance through hyperparameter tuning using the Whale Optimization Algorithm (WOA), a nature-inspired metaheuristic technique. Furthermore, we apply SHAP (Shapley Additive Explanations) to interpret the contribution of each input variable, enabling model transparency and ecological insight. The optimized models are validated using historical monitoring station data and applied to generate a spatio-temporal prediction of water quality conditions across the entire RRB in 2023. This research not only demonstrates the effectiveness of multi-source, interpretable machine learning models for large-scale water quality monitoring but also introduces a transferable methodological framework with strong potential for broader applications in hydrological and environmental management.

## 2. Methods and Materials

### 2.1. Study Area

This study is conducted in the RRB (100°04′–104°50′E, 22°23′–25°36′N), which spans an area of approximately 61,700 km<sup>2</sup> within the territory of China (Figure 1). The basin features a topography that descends from the highlands in the northwest to the lowlands in the southeast, characterized by significant elevation differences and deeply incised valleys. It experiences a typical subtropical monsoon climate, with distinct wet and dry seasons. The wet season extends from May to October, accounting for over 85% of the annual precipitation, while the dry season lasts from November to April, contributing less than 15% of the annual precipitation [31]. Variations in water quality within the RRB have substantial implications for the ecological integrity and socio-economic development of both upstream and downstream regions.



**Figure 1.** Geographical map of the RRB and the spatial distribution of water quality monitoring stations.

### 2.2. Multi-Source Data

#### 2.2.1. Water Quality Parameters

Water quality parameter data were collected from nine monitoring stations distributed across the RRB and its tributaries (Figure 1) during the period from 2021 to 2023. These data were obtained from the China National Environmental Monitoring Center ([www.cnemc.cn](http://www.cnemc.cn)),

which operates an automated surface water quality monitoring system established by the Ministry of Ecology and Environment of the People's Republic of China [32]. Measurements were recorded at four-hour intervals, providing high-frequency time series data. In accordance with the national Environmental Quality Standards for Surface Water (GB 3838-2002 [33]) and their relevance to water quality assessment, six key water quality parameters were selected for inversion: pH, dissolved oxygen (DO, mg/L), chemical oxygen demand ( $\text{COD}_{\text{Mn}}$ , mg/L), total phosphorus (TP, mg/L), total nitrogen (TN, mg/L), and ammonia nitrogen ( $\text{NH}_3\text{-N}$ , mg/L) [3,34]. To address data gaps, linear interpolation was applied to fill a small number of missing values in the in situ water-quality time series, which accounted for 2.6% of the total 60,061 observations. The interpolation process did not result in significant changes in the coefficient of variation for any of the water quality parameter time series ( $p$ -value of the two-sided paired  $t$ -test is equal to 0.55), thereby preserving their statistical integrity. For consistency with the temporal resolution of remote sensing data, all water quality parameters were aggregated to daily averages for subsequent modeling.

### 2.2.2. Remote Sensing Data and Preprocessing

Given its high spatial resolution (up to 10 m) and short revisit interval (5 days with the combined operation of two satellites), Sentinel-2 MSI imagery was selected for water quality inversion in this study. Sentinel-2, launched by the European Space Agency, consists of two optical satellites—Sentinel-2A and Sentinel-2B—that collectively provide data across 13 spectral bands. This study utilized Level-2A (L2A) products, which are freely available and have undergone geometric, radiometric, and atmospheric corrections. Due to the unavailability of Band 10 following atmospheric correction, only Bands 1–9, Band 8A, and Bands 11–12 were used for analysis. To minimize the influence of cloud contamination, only images with cloud cover below 10% from 2021 to 2023 were selected. Cloud-masked and filtered imagery was then subjected to further analysis and water quality model inversion.

### 2.2.3. Meteorological and Land Use Data

Meteorological data spanning from January 2021 to December 2023 were obtained from the China Meteorological Administration. The selected meteorological variables included air temperature (TE, °C), maximum air temperature (MaxTE, °C), minimum air temperature (MinTE, °C), accumulated precipitation (PREC, mm), relative humidity (RHU, %), atmospheric pressure (PRS, hPa), and wind speed (WinS, m/s). All meteorological variables were resampled to a uniform spatial resolution of 500 m using nine interpolation techniques: ordinary kriging, linear interpolation, nearest neighbor interpolation, cubic interpolation, global polynomial interpolation, radial basis function interpolation, local polynomial interpolation, synergistic kriging, and Thiessen (Tyson) polygon interpolation. Land use data for the years 2021–2023, with a spatial resolution of 10 m, were sourced from the Environmental Systems Research Institute (ESRI) Land Cover dataset (<https://livingatlas.arcgis.com/landcover/>, accessed on 16 September 2025). The dataset includes classification into nine land use categories: water, trees, flooded vegetation, crops, built area, bare ground, snow/ice, and rangeland [35]. These variables were used to assess the impact of land surface conditions on water quality dynamics within the study area.

### 2.2.4. Landscape Pattern Indices

The landscape pattern indices provide insights into the spatial distribution characteristics of different land cover types within the basin. Although these indices do not directly measure water quality, they are closely associated with hydrological processes, pollutant transport pathways, and biogeochemical regulation. Variations in landscape aggregation,

connectivity, shape complexity, and diversity can modify runoff generation and nutrient retention, thereby influencing the concentration of pollutants and micro-organism activities in adjacent water bodies. In this study, thirteen landscape pattern indices potentially influencing water quality were calculated using Fragstats 4.2 software [3,36–38]. The selected landscape pattern indices include number of patches (NP), contagion index (CONTAG), patch cohesion index (COHESION), patch density (PD), interspersed juxtaposition index (IJI), mean shape index (MSI), Shannon’s diversity index (SHDI), Shannon’s evenness index (SEI), percentage of landscape (PLAND), landscape division index (DIVISION), aggregation index (AI), landscape shape index (LSI), and largest patch index (LPI) [39,40]. For example, LPI describes the percentage of the largest patches of the same patch type within total landscape area. Depending on different dimensions of spatial structure of landscape pattern indices (e.g., area, aggregation and shape), these indices can be grouped into four categories, including Area and Edge metrics (PLAND, LPI), Shape metrics (MSI), Aggregation metrics (NP, CONTAG, COHESION, PD, IJI, DIVISION, AI, LSI), and Diversity metrics (SHDI, SEI). These categories provide a comprehensive view of landscape structure and its potential implications for water quality.

### 2.3. Methods

#### 2.3.1. Correlation Analysis

Correlation analysis is commonly employed to quantify the strength and direction of the relationship between two variables. In this study, the Pearson correlation coefficient (Equation (1)) was used to evaluate the interrelationships among various water quality parameters and to guide the selection of appropriate remote sensing bands and landscape pattern indices for model construction. The Pearson coefficient, denoted as  $r_{xy}$ , measures the linear correlation between two variables. A positive  $r_{xy}$  indicates a direct (positive) correlation, whereas a negative value suggests an inverse (negative) relationship. The closer the absolute value of  $r_{xy}$  is to 1, the stronger the correlation. Here,  $x_i$  and  $y_i$  represent the sample values of the two variables under consideration.

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

#### 2.3.2. Whale Optimization Algorithm

The WOA is inspired by the hunting behavior of humpback whales [41]. In this study, the algorithm is used to optimize the hyperparameters of the machine learning models. For example, hyperparameters include “max depth” and “eta” in XGBoost and “num leaves” and “learning rate” in LightGBM. The hyperparameters were tuned by minimizing the prediction error on the validation dataset. Specifically, the WOA employs a grid search procedure to select the optimal combination of hyperparameters that minimized the mean squared error (MSE) computed via validation dataset. The WOA operates through three main mechanisms: encircling prey, bubble-net attacking, and searching for prey. In the encircling prey phase, each individual iteratively approaches the current best solution along a straight path. The bubble-net attacking phase simulates the local search behavior, allowing individuals to either continue encircling or to spiral towards the optimal solution. To avoid becoming trapped in local optima, some individuals perform a global search by leaping out of the current range during the search for prey phase. By employing the WOA, this study not only avoids the manual tuning of hyperparameters but also enhances the predictive ability of machine learning models.



### 2.3.3. Machine Learning Methods

In this study, three tree-based machine learning models—RF, XGBoost, and LightGBM—were employed to develop water quality parameter prediction models. RF integrates multiple decision trees, each trained on independently sampled data, to make predictions [42]. XGBoost leverages gradient boosting, regularization, and feature parallelism to enhance predictive performance [43]. LightGBM further improves prediction accuracy by employing feature dimension reduction techniques and selecting large-gradient samples to calculate information gain, which reduces memory usage and increases computational efficiency [44]. All samples were divided into a training dataset (80%), a validation dataset (10%), and a test dataset (10%). The training dataset was used to train the models, the validation dataset to optimize hyperparameters, and the test dataset to evaluate model performance. The performance of the three machine learning models was comprehensively assessed, and the best-performing model was selected for each water quality parameter.

### 2.3.4. Interpretation and Evaluation of Input Variables and Models

In this study, SHAP was employed to quantify the extent to which input variables influence changes in water quality parameters. SHAP is a post hoc explanation method that draws on game-theoretic principles to interpret machine learning models by calculating the marginal contribution of each feature [45], thereby measuring the effect magnitude of each variable (Equation (2)):

$$g(x) = \phi_0 + \sum_{i=1}^M \phi_i \quad (2)$$

where  $\phi_0$  represents the baseline value (i.e., the mean of the predicted values across all samples),  $\phi_i$  denotes the Shapley value for the  $i_{th}$  feature, and  $M$  is the total number of features in the model. Essentially,  $\phi_i$  is an expectation value reflecting the change in model output when feature  $x_i$  is included or excluded in various feature combinations.

The performance of the machine learning models was evaluated using the coefficient of determination ( $R^2$ ) and mean squared error (MSE) as defined in Equations (3) and (4):

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2} \quad (3)$$

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (4)$$

where  $y_i$  is the observed value,  $\hat{y}_i$  is the predicted value,  $\bar{y}_i$  is the mean of the observed values, and  $n$  is the total number of samples.

## 3. Results and Analysis

### 3.1. Correlation Analysis of Water Quality Parameters

Pearson's correlation coefficient (Table 1) was calculated to analyze the degree of association between water quality parameters. The correlation coefficients between  $\text{COD}_{\text{Mn}}$  and TP,  $\text{NH}_3\text{-N}$  and DO were relatively high ( $r = 0.54, 0.46$ , and  $-0.43$ , respectively). Additionally, TP and TN showed a moderate correlation with a coefficient of  $0.49$ . pH showed no significant correlation with  $\text{NH}_3\text{-N}$  and  $\text{COD}_{\text{Mn}}$ . These findings suggest that changes in one water quality parameter may lead to changes in others, potentially affecting overall water quality.

**Table 1.** Pearson correlation coefficients among water quality parameters.

|                    | DO        | NH <sub>3</sub> -N | TP        | TN       | pH       | COD <sub>Mn</sub> |
|--------------------|-----------|--------------------|-----------|----------|----------|-------------------|
| DO                 | 1         | −0.25 ***          | −0.20 *** | −0.03    | 0.32 *** | −0.43 ***         |
| NH <sub>3</sub> -N | −0.25 *** | 1                  | 0.38 ***  | 0.37 *** | −0.06    | 0.46 ***          |
| TP                 | −0.20 *** | 0.38 ***           | 1         | 0.49 *** | 0.19 *** | 0.54 ***          |
| TN                 | −0.03     | 0.37 ***           | 0.49 ***  | 1        | −0.11 *  | 0.30 ***          |
| pH                 | 0.32 ***  | −0.06              | 0.19 ***  | −0.11 *  | 1        | 0.02              |
| COD <sub>Mn</sub>  | −0.43 *** | 0.46 ***           | 0.54 ***  | 0.30 *** | 0.02     | 1                 |

Note: \* and \*\*\* indicate significance at the 0.05 and 0.001 levels, respectively.

### 3.2. Input Variables Selection

The aim of this study is to develop suitable prediction models for key water quality parameters (DO, NH<sub>3</sub>-N, TP, TN, pH, COD<sub>Mn</sub>) to facilitate river water quality monitoring. The input variables considered in the modeling process included spectral reflectance, meteorological data, land use information, and landscape pattern indices. The Sentinel-2 satellite provides 12 bands, denoted B1 through B12 (excluding Band 10) in order of increasing wavelength. Since combinations of different bands can effectively enhance feature extraction [21,46], four combinations were constructed in this study (Equations (5)–(8)):

$$AI(B_i, B_j) = B_i + B_j \quad (5)$$

$$DI(B_i, B_j) = B_i - B_j \quad (6)$$

$$RI(B_i, B_j) = \frac{B_i}{B_j} \quad (7)$$

$$NI(B_i, B_j) = \frac{B_i - B_j}{B_i + B_j} \quad (8)$$

where  $B_i$  and  $B_j$  represent band  $i$  and band  $j$  respectively ( $i, j = 1, 2, 3, 4, 5, 6, 7, 8, 8A, 9, 11, 12$ ). For each water quality parameter, the ten bands or band combinations showing the highest Pearson correlation coefficients were selected as input variables. The selected band combinations for each water quality parameter are detailed in Table S1.

To avoid multicollinearity and to capture the impact of land use on water quality, built-up areas, trees, and rangeland were included as land use input variables. Built-up area affects the quality of surface water by increasing the impervious surface and changing the surface runoff. Trees and rangeland can modify hydrological and ecological processes by increasing vegetation interception and soil infiltration, which in turn influences nutrient transport and pollutant attenuation near water bodies. Additionally, landscape pattern indices are often correlated and may contain redundant information [47]. Therefore, after calculating the landscape pattern indices mentioned previously, pairwise Pearson correlation coefficients were calculated. Based on their relevance and their ability to represent different aspects of landscape composition and configuration, seven indices were included: CONTAG, COHESION, MSI, SHDI, SEI, LSI, and LPI. CONTAG and COHESION represent the aggregation and connectivity of the landscape, which affecting water bodies by altering surface runoff. LSI and MSI represent the shape and edge of the landscape. Different shapes of the landscape will change the contact area between water bodies and land. SHDI, SEI and LPI describe the composition of the landscape. The presence of variety of landscape types within an area can enhance the interception and dilution of pollutants, and reduce the concentration of pollutants entering water bodies. The selected indices cover different structural dimensions of the landscape and have impacts on hydrological process and pollutants transportation.

By matching all input variables and water quality parameters in both the temporal and spatial dimensions, the dataset for this study was obtained. Select the date of the band combinations data as the time reference. The water quality parameters data are based on the daily average values for the day of collection. The values of the meteorological variables are based on the daily average values of the grid (500 m of spatial resolution) where the monitoring stations are located. The land use and landscape pattern indices used the values of the sub-watersheds where the monitoring stations are located.

Integrating the above components, each water quality parameter model in this study utilized a total of 27 input variables: the ten bands or band combinations with the highest correlation, seven meteorological variables, three land use variables, and seven landscape pattern indices. Each dataset of water quality parameters consists of 729 pieces of data.

### 3.3. Model Construction and Performance Evaluation

In this study, three machine learning methods—RF, XGBoost, and LightGBM—were employed to construct the prediction models for key water quality parameters. A significant methodological innovation of this study lies in the use of the WOA to fine-tune the hyperparameters of XGBoost and LightGBM, leading to the WOA-XGBoost and WOA-LightGBM models, respectively. This approach reduces the need for manual hyperparameter tuning and enhances model generalization performance. The optimal hyperparameter values for the two optimized models are presented in Table S2. The six water quality parameters (DO,  $\text{NH}_3\text{-N}$ , TP, TN, pH,  $\text{COD}_{\text{Mn}}$ ) were inverted using a combination of remote sensing image data, meteorological data, land use data, and landscape pattern indices as input variables. The training results of the three models for each water quality parameter are summarized in Table 2. The results demonstrate that all three machine learning models are capable of inverting water quality parameters with reasonable accuracy. Among them, the RF model achieved the highest performance for pH ( $R^2 = 0.8309$ ,  $\text{MSE} = 0.0182$ ) and also performed best for DO. For the other four water quality parameters ( $\text{NH}_3\text{-N}$ , TP, TN, and  $\text{COD}_{\text{Mn}}$ ), the WOA-LightGBM model consistently outperformed the others, highlighting the strength of combining LightGBM with the WOA approach for water quality prediction tasks. Compared with the LightGBM using default hyperparameters, WOA-LightGBM significantly improved the prediction accuracy of most water quality parameters. This innovative integration of an evolutionary optimization algorithm with advanced ensemble learning models shows promising potential for improving water quality monitoring and management.

The fitted scatter plots obtained from the predictions of the best model for each water quality parameter on the validation dataset are presented in Figure 2. The prediction models for DO, TN, and pH demonstrated strong predictive performance, with  $R^2$  values exceeding 0.7 on the test dataset. Specifically, the DO model achieved an  $R^2$  of 0.7025 with an MSE of 0.4048 mg/L, the TN model reached an  $R^2$  of 0.7115 with an MSE of 0.1970 mg/L, and the pH model yielded an  $R^2$  of 0.8309 with an MSE of 0.0182. Although the best model for TP exhibited a relatively modest  $R^2$  of 0.5880, its MSE was only 0.0006 mg/L, indicating that the model still had considerable explanatory power for TP. For  $\text{COD}_{\text{Mn}}$ , the best prediction model achieved an  $R^2$  of 0.5710 and an MSE of 0.6224 mg/L on the test dataset. The  $\text{NH}_3\text{-N}$  model performed less favorably compared to the other water quality parameters, but its  $R^2$  exceeded 0.4, indicating that the model retained a moderate level of predictive ability. An analysis of the fitted scatter plot for  $\text{NH}_3\text{-N}$  revealed that a large proportion of observations were concentrated at 0.025 mg/L, suggesting a potential detection limit in  $\text{NH}_3\text{-N}$ . From a modeling perspective, such value clustering reduces the effective variability of the water quality parameter and weakens the statistical relationship between input variables and  $\text{NH}_3\text{-N}$ . This limited signal-to-noise ratio constrains the learning ability of data-driven

models and leads to degraded predictive performance. The poor performance of the  $\text{NH}_3\text{-N}$  model is not solely attributable to monitoring accuracy, but also reflects inherent methodological limitations in capturing low-variability. In the future, the accuracy of the model can be improved by using other models or by adding variables that may affect  $\text{NH}_3\text{-N}$ . Overall, while all the prediction models tended to overestimate low values and underestimate high values, the model fits were generally satisfactory, demonstrating their potential applicability in river water quality monitoring.

**Table 2.** The accuracy of the water quality parameters prediction models.

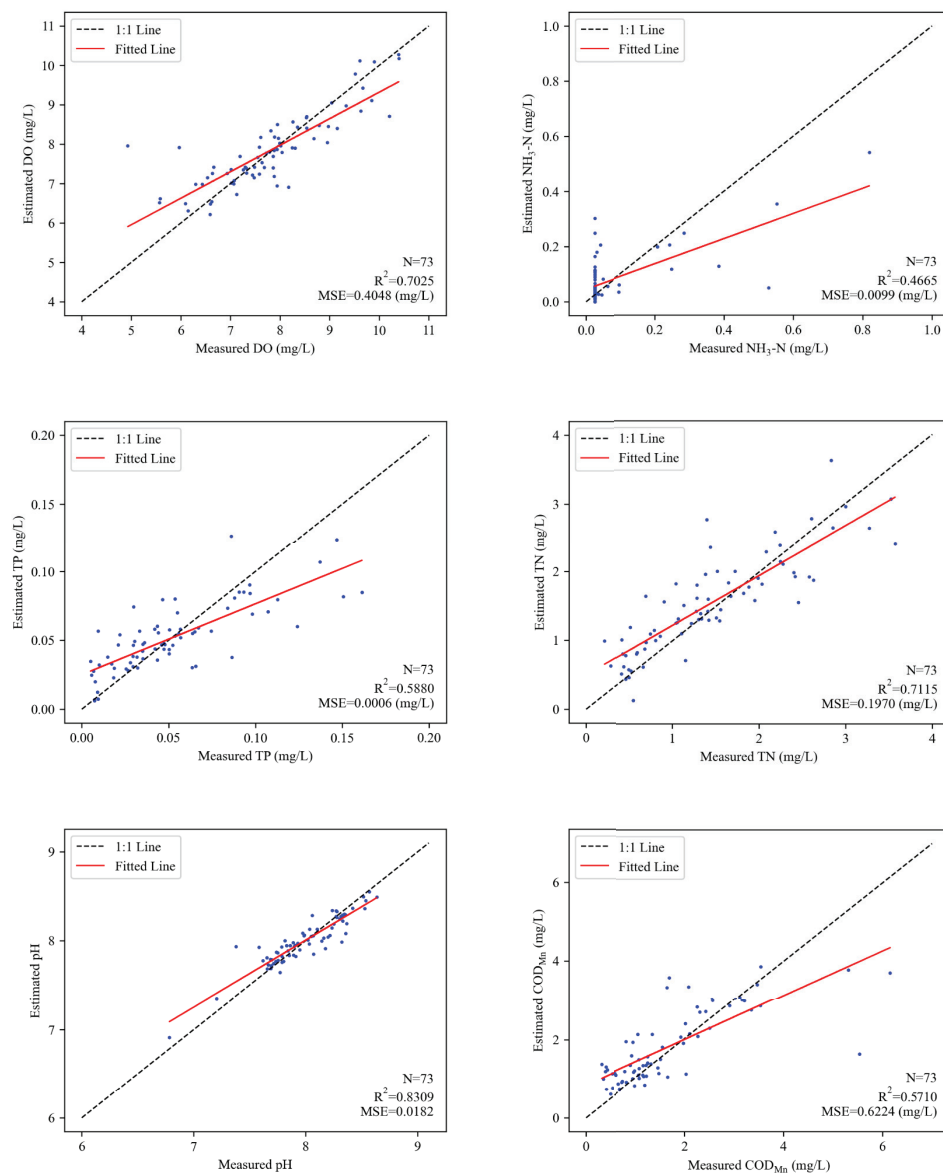
| Water Quality Parameter  | Model        | Testing Dataset |               |
|--------------------------|--------------|-----------------|---------------|
|                          |              | $R^2$           | MSE           |
| DO                       | RF           | <b>0.7025</b>   | <b>0.4048</b> |
|                          | WOA-XGBoost  | 0.6870          | 0.4259        |
|                          | LightGBM     | 0.6607          | 0.4616        |
|                          | WOA-LightGBM | 0.6866          | 0.4264        |
| $\text{NH}_3\text{-N}$   | RF           | 0.2384          | 0.0141        |
|                          | WOA-XGBoost  | 0.4146          | 0.0108        |
|                          | LightGBM     | 0.1749          | 0.0153        |
|                          | WOA-LightGBM | <b>0.4665</b>   | <b>0.0099</b> |
| TP                       | RF           | 0.4467          | 0.0007        |
|                          | WOA-XGBoost  | 0.4276          | 0.0008        |
|                          | LightGBM     | 0.1624          | 0.0011        |
|                          | WOA-LightGBM | <b>0.5880</b>   | <b>0.0006</b> |
| TN                       | RF           | 0.6781          | 0.2136        |
|                          | WOA-XGBoost  | 0.6547          | 0.2358        |
|                          | LightGBM     | 0.5368          | 0.3162        |
|                          | WOA-LightGBM | <b>0.7115</b>   | <b>0.1970</b> |
| pH                       | RF           | <b>0.8309</b>   | <b>0.0182</b> |
|                          | WOA-XGBoost  | 0.8162          | 0.0198        |
|                          | LightGBM     | 0.7618          | 0.0256        |
|                          | WOA-LightGBM | 0.7425          | 0.0277        |
| $\text{COD}_{\text{Mn}}$ | RF           | 0.5184          | 0.6988        |
|                          | WOA-XGBoost  | 0.5399          | 0.6676        |
|                          | LightGBM     | 0.4093          | 0.8570        |
|                          | WOA-LightGBM | <b>0.5710</b>   | <b>0.6224</b> |

### 3.4. Application of the Water Quality Parameters Prediction Model

This study applied the optimized prediction models of six key water quality parameters across the entire RRB, demonstrating the innovative integration of machine learning with remote sensing for large-scale water quality monitoring. The water bodies were extracted using Google Earth Engine (GEE), leveraging its powerful cloud computing platform to process all Sentinel-2 images from 2023 with less than 10% cloud cover in the RRB. This approach ensured the acquisition of high-quality remote sensing data for subsequent water quality inversion. The inversion results for each water quality parameter are presented in Figures 3–8.

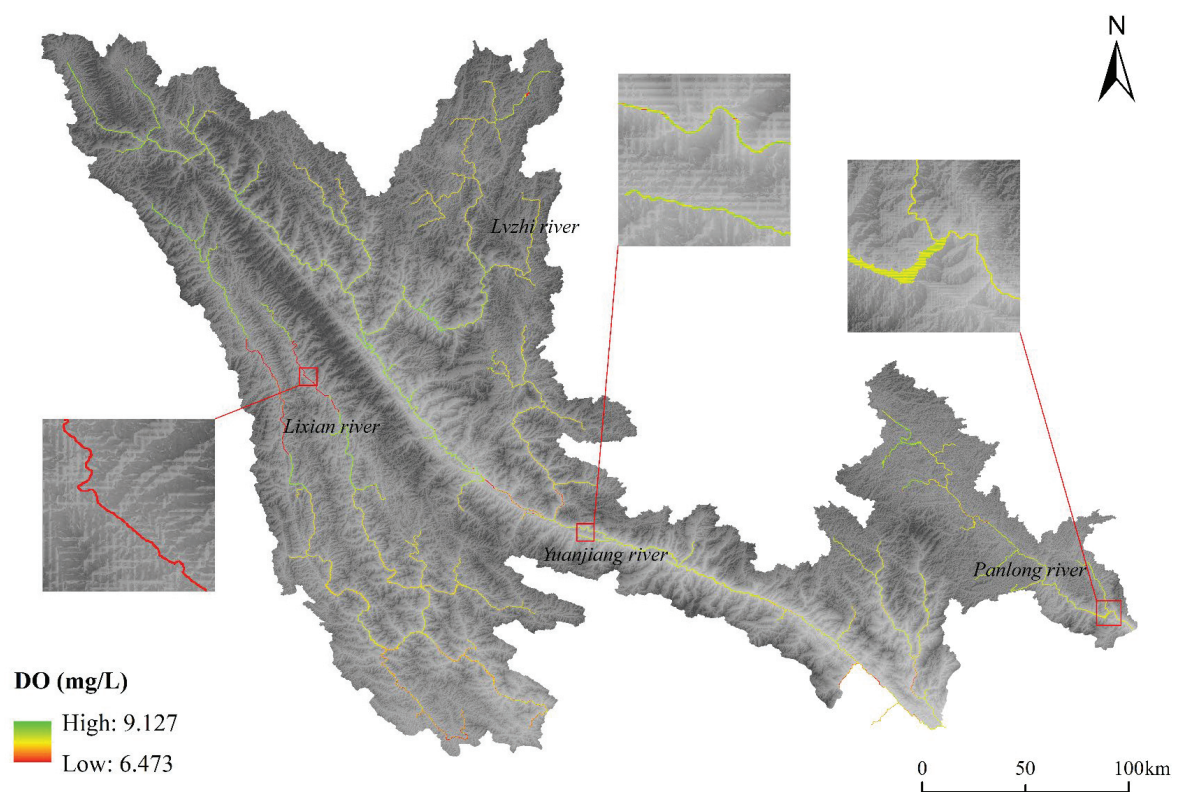
Figure 3 shows the distribution of DO concentrations, ranging from 6.473 to 9.127 mg/L. According to the Environmental Quality Standards for Surface Water (GB 3838-2002), all DO values in the basin meet at least Class II standards. Although DO levels in the midstream were slightly lower, they remained satisfactory overall. For  $\text{NH}_3\text{-N}$  (Figure 4), the model predicted a maximum concentration of 0.677 mg/L, placing it within Class III water quality. The downstream regions showed higher  $\text{NH}_3\text{-N}$  concentrations than upstream areas, likely due to more intensive human activities leading to nitrogen

pollution [48]. The spatial distribution of TP (0.013–0.112 mg/L) in Figure 5 closely mirrors that of  $\text{NH}_3\text{-N}$ , with significantly lower concentrations in the upstream areas compared to the middle and lower reaches. Overall, TP levels across the basin meet acceptable standards. Figure 6 displays TN concentrations, which ranged from 0.258 to 3.281 mg/L. Notably, the lowest TN concentrations still did not meet Class I standards, and concentrations in most middle and lower reaches exceeded 1.5 mg/L (Class IV–V standards), indicating serious nitrogen pollution. This finding underscores the need for improved regional production and living management and for effective control of urban nitrogen discharge to mitigate nitrogen pollution in the watershed. The inversion results for pH (Figure 7) ranged from 7.252 to 8.337, indicating that all values were within a suitable range without excessive acidity or alkalinity.  $\text{COD}_{\text{Mn}}$  concentrations (Figure 8) varied from 1.090 to 3.223 mg/L, meeting at least the Class II standard. High  $\text{COD}_{\text{Mn}}$  concentrations were mainly concentrated downstream, with additional hotspots in the Lvzhi River and the upper reaches of the Panlong River, suggesting potential natural sources of concern.

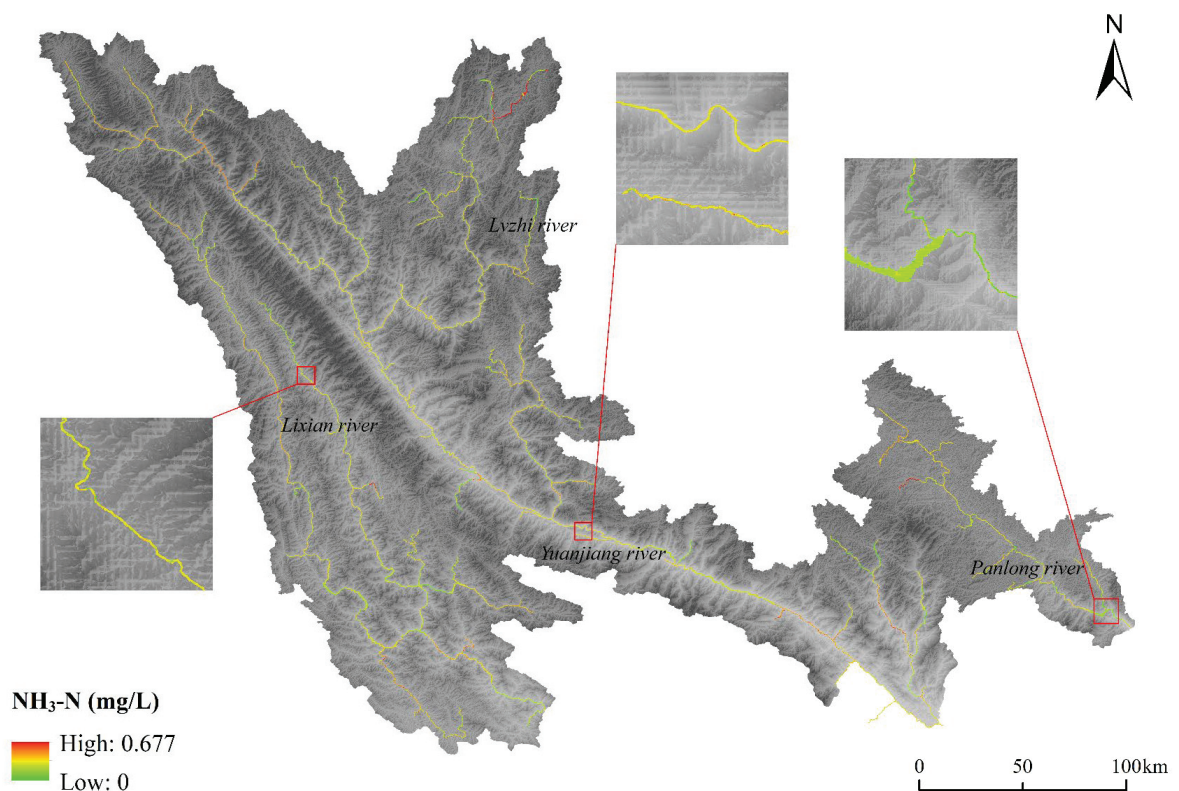


**Figure 2.** The fitting scatter plots of best models for water quality parameters.

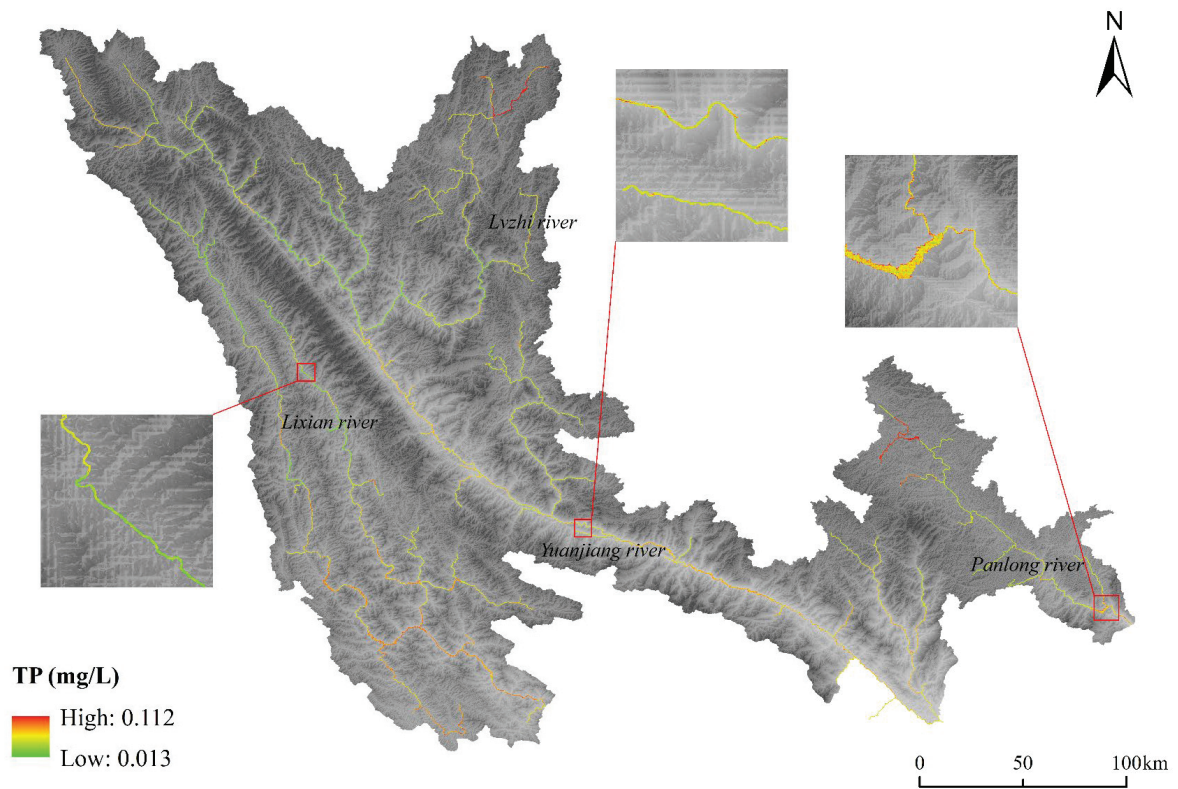




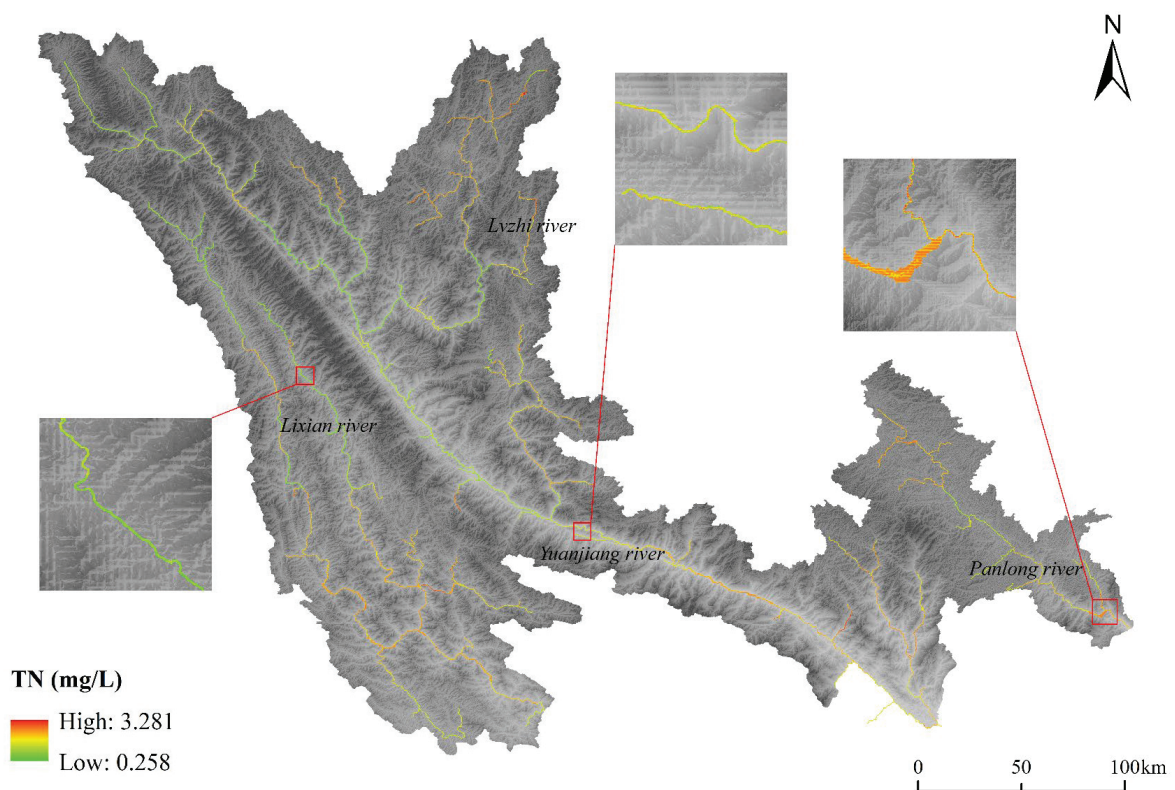
**Figure 3.** Inversion results of DO concentration in the RRB in 2023.



**Figure 4.** Inversion results of NH<sub>3</sub>-N concentration in the RRB in 2023.

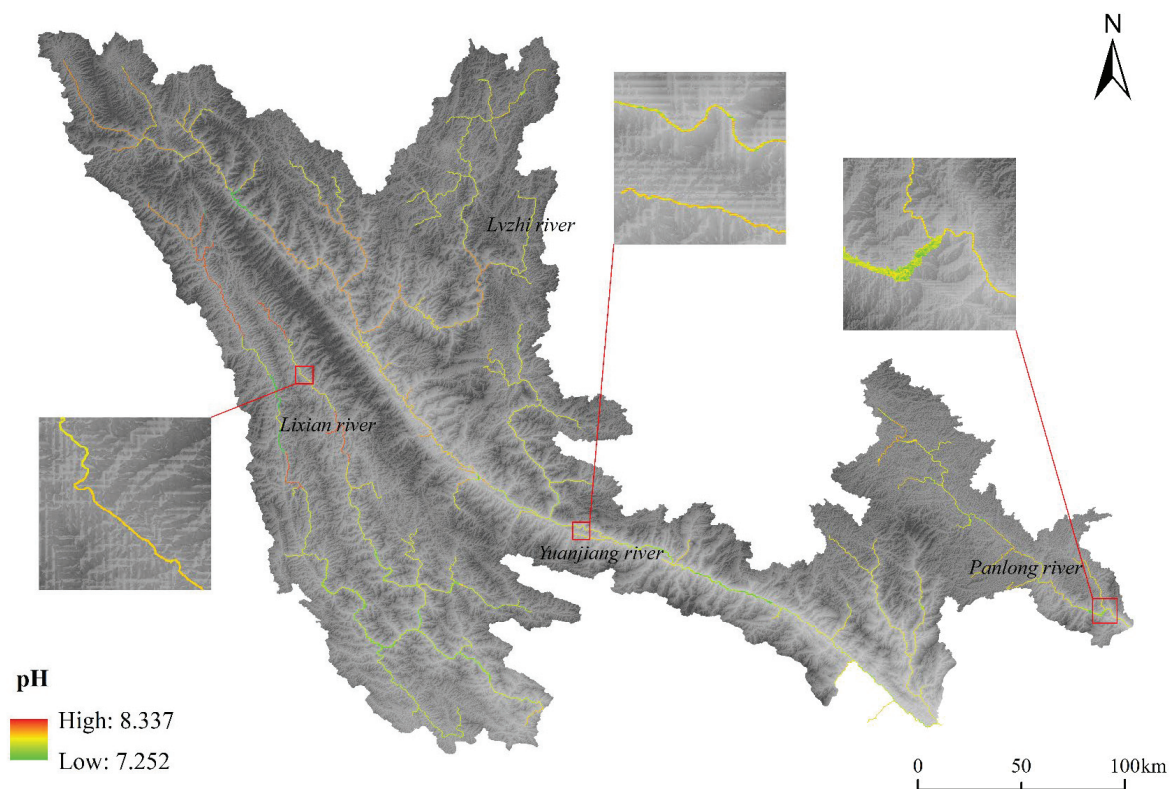


**Figure 5.** Inversion results of TP concentration in the RRB in 2023.

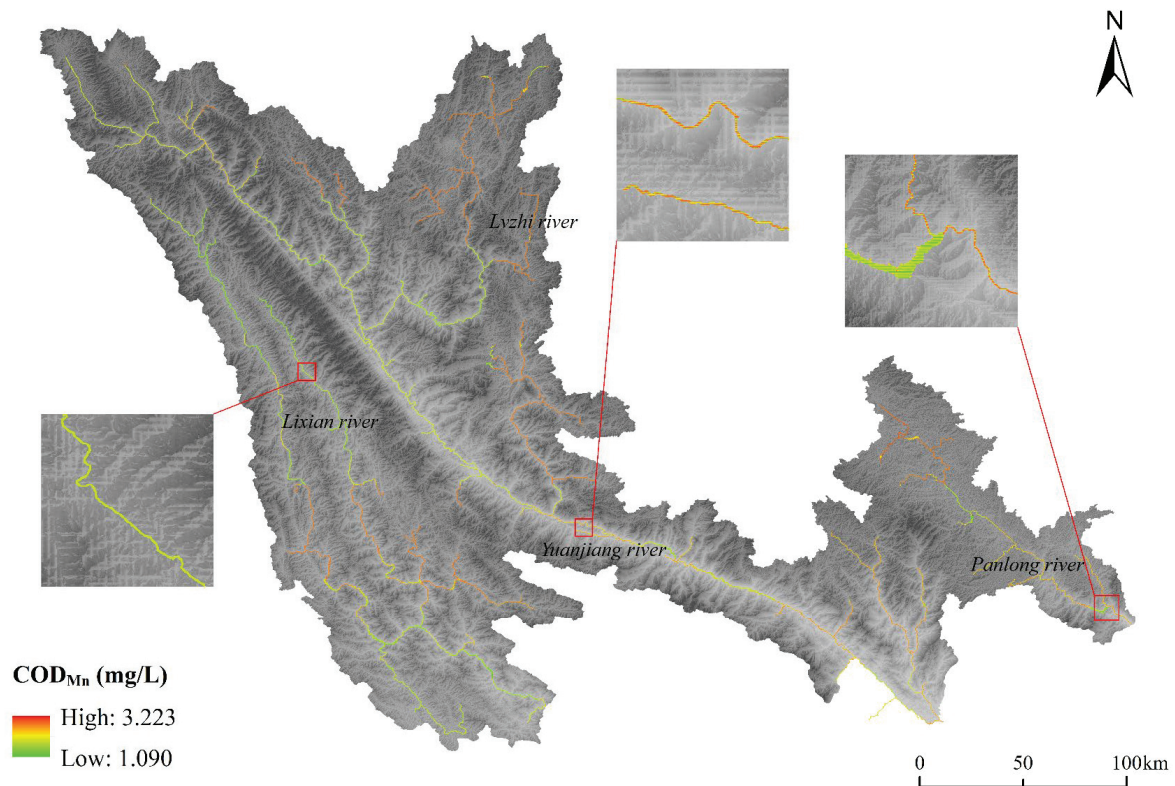


**Figure 6.** Inversion results of TN concentration in the RRB in 2023.





**Figure 7.** Inversion results of pH concentration in the RRB in 2023.



**Figure 8.** Inversion results of COD<sub>Mn</sub> concentration in the RRB in 2023.

This study demonstrates the methodological advantages of integrating machine learning models optimized by WOA with multi-source data and GEE-based feature extraction.

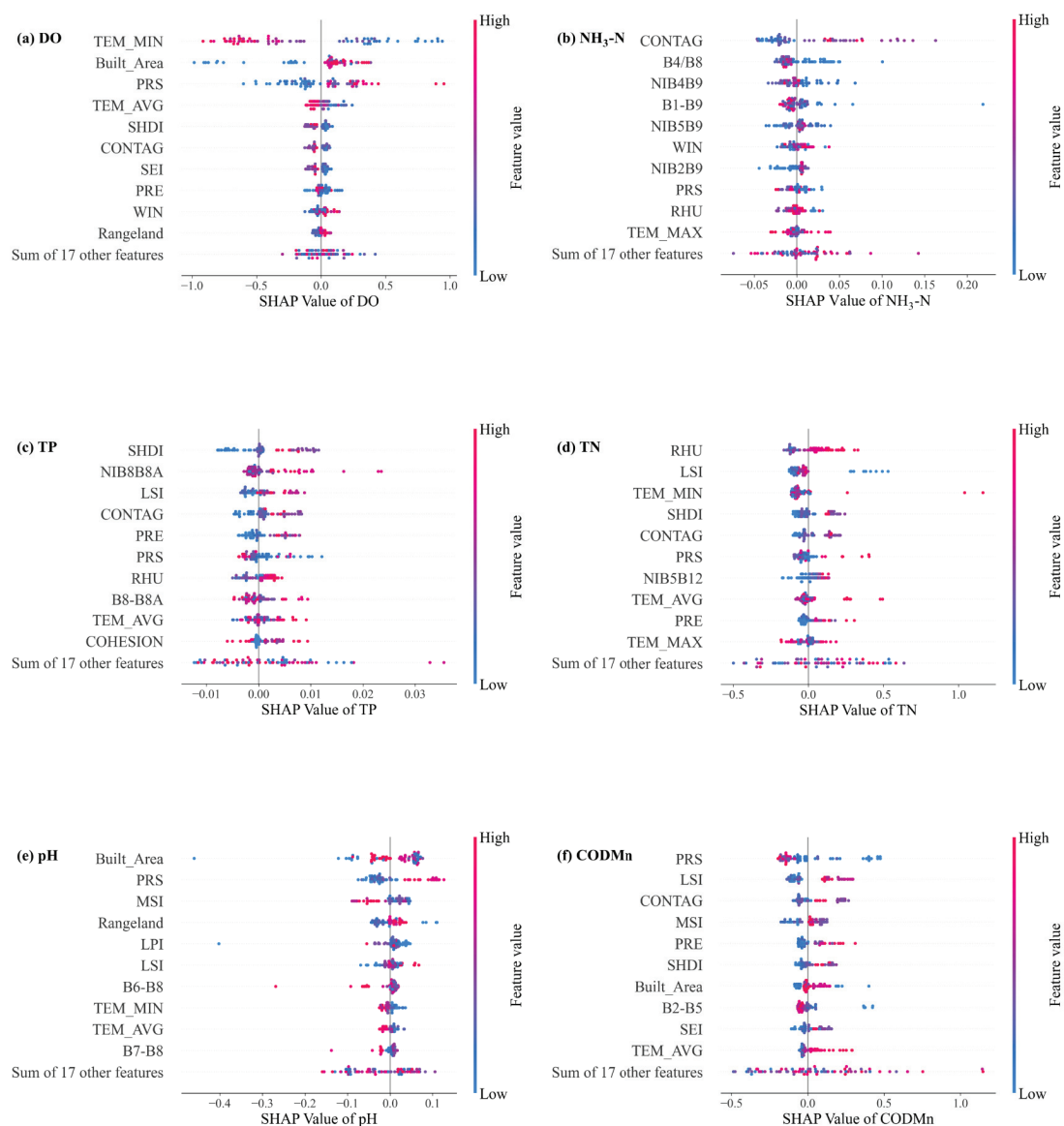
The results indicate that the proposed framework effectively captures spatial variability in water quality indicators, thereby validating the model's capability to learn complex spatial patterns from heterogeneous environmental data. The proposed framework can effectively capture the spatial variation characteristics of water quality parameters, verifying the model's ability to learn spatial patterns from different environmental data. The application of the model throughout the Red River Basin indicates that the design of this method is scalable and transferable, and is applicable to other river basins with similar data availability. The application of this method within the basin a new idea for regional water resource management and pollution control.

## 4. Discussion

### 4.1. SHAP Analysis of the Interpretability of Variables

Due to the inherent lack of interpretability in many machine learning models, this study employs the SHAP model to analyze the importance of input variables for each water quality parameter. Figure 9 presents the global variable importance map generated by SHAP for each water quality parameter. For each sub-figure, the variables are ranked from top to bottom according to their decreasing importance. Each point in the figure represents a sample: red indicates higher variable values, while blue indicates lower variable values. The horizontal axis displays the SHAP value for each variable, reflecting its influence on the model's output. Points to the left of the central axis indicate a negative contribution to the predicted value, while those to the right indicate a positive contribution.

The results show that remote sensing imagery, meteorological data, land use, and landscape pattern indices all impact water quality parameters, although the specific sensitivities vary among them. For DO, meteorological variables such as TEM\_MIN and PRS play a substantial role, with lower minimum temperatures corresponding to higher SHAP values (indicating a stronger positive effect on DO). The proportion of built-up area also shows a significant impact, suggesting that DO is sensitive to human activities within the basin. For NH<sub>3</sub>-N, SHAP reveals a strong dependence on remote sensing bands (B4/B8, NIB4B9, B1–B9). Notably, the CONTAG ranks highest in importance, highlighting the influence of water body aggregation and dispersion on NH<sub>3</sub>-N concentrations. TP and COD<sub>Mn</sub> are both strongly influenced by landscape pattern indices, indicating that the spatial distribution of water bodies within the basin interacts closely with phosphorus and nitrogen dynamics. Specifically, SHDI, LSI, and CONTAG have prominent effects on TP, while the band combination NIB8B8A also contributes. It reflects the effects of land-use heterogeneity, landscape fragmentation, and spatial aggregation on phosphorus transport and retention processes. Higher SHDI and lower CONTAG usually imply a more heterogeneous and fragmented landscape, which enhances surface runoff pathways and soil erosion, thereby promoting particulate phosphorus transport into water bodies. COD<sub>Mn</sub> is most affected by LSI, CONTAG, and MSI, with atmospheric pressure also playing a significant role. Complex landscape shapes (high LSI and MSI) increase the contact area between land and water bodies, facilitating the input and transformation of organic matter and consequently elevating COD<sub>Mn</sub>. For TN, SHAP interpretability results suggest that both landscape pattern indices (LSI, SHDI) and meteorological variables (RHU, TEM\_MIN) jointly influence the model's output. Finally, pH is influenced by all categories of input variables, demonstrating the diversity and complexity of variables affecting pH. These findings underscore the complexity of river water quality parameters, indicating that remote sensing imagery alone is insufficient to fully explain water quality variability.



**Figure 9.** Global interpretability of SHAP for water quality parameters.

An overall inspection of Figure 9 reveals that while the importance of specific variables varies by water quality parameter, certain variables consistently exhibit significant influence, such as CONTAG, Built\_Area, and TEM\_MIN. This suggests that although different water quality parameters are driven by different sets of variables, there are also important interconnections among them. Interestingly, Built\_Area impacts both DO and pH but in opposite directions, further illustrating the nuanced influence of input variables on water quality. Overall, the SHAP interpretability model provides valuable insights into both the direction and magnitude of variable influences, greatly aiding the analysis of water quality dynamics and supporting informed water quality management.

#### 4.2. Impacts of Multiple Variables on Water Quality Prediction

Most prediction models for water quality parameters have historically relied solely on remote sensing data, while few studies have integrated meteorological and land use variables. In this study, the integration of multi-source data substantially improved the model's ability to explain the observed spatial and temporal variability of water quality



parameters, particularly for non-optically active parameters. The results suggest that meteorological, land use, and landscape patterns can affect water quality parameters by altering the generation, transmission, and accumulation of pollutants. Based on the SHAP interpretability analysis of water quality parameters, we discuss the influence path of input variables with large SHAP values on water quality.

Key meteorological elements—temperature, precipitation, and humidity—affect water quality both directly and indirectly. Environmental temperature can alter the physical properties and biological activities of water bodies. An increase in water temperature reduces the solubility of oxygen and accelerates the growth of algae. This not only reduces the DO concentration but also has an impact on TN (algae activities are often accompanied by the process of absorbing and releasing inorganic nitrogen). Precipitation has a significant impact on DO, pH and COD<sub>Mn</sub>. More precipitation increases the total volume of surface runoff, allowing more organic matter to enter water bodies through surface runoff and causing the concentration of COD<sub>Mn</sub> to rise. At the same time, the increase in water volume will dilute the original water body and accelerate the flow rate of the river, indirectly affecting the DO concentration and pH value. High humidity will increase the concentration of TN in water bodies. This might be an indirect transmission path: when the humidity is high, the moisture content of soil will increase, causing the residual nitrogen elements in the farmland to flow into water bodies, which leads to an increase in TN concentration.

Built-up areas near water bodies have an impact on river water quality. Built-up areas are hubs of human activity, where treated or untreated domestic sewage and industrial wastewater often enter the environment. Improper wastewater management can severely degrade water quality. Additionally, the expansion of impervious surfaces in built-up areas alters surface runoff patterns, leading to increased non-point source pollution.

Among the landscape pattern indices, CONTAG, SHDI and LSI have the greatest impact on water quality parameters. CONTAG indicates the degree of aggregation and connectivity of similar landscape. Given that the Red River Basin is dominated by the primary industry, high CONTAG might be due to contiguous farmland. This leads to the spatial concentration of fertilizer application, thereby increasing the concentration of NH<sub>3</sub>-N in water bodies. SHDI describes the diversity of landscape types, which has the greatest impact on TP. High SHDI indicates that the landscape types in the area are more diverse, resulting in a more varied source of TP in the water body. LSI represents the complexity of the landscape boundary. It has an impact on three water quality parameters (TP, TN and COD<sub>Mn</sub>) for measuring pollutants. This might be because the boundary areas are more prone to human disturbance and natural erosion. The expansion of the boundary will increase soil erosion and contact area between farmland and water bodies, allowing phosphorus and organic matter to enter water bodies. The influence of LSI on TN is indirect. LSI may affect TN concentration by altering surface runoff. The above analysis is consistent with the SHAP values of LSI. LSI shows a significantly positive correlation with TP and COD<sub>Mn</sub>, while it seems to have a nonlinear relationship with TN.

Based on the above analysis, it can be found that the machine learning model based on multi-source data proposed in this study can well explain the variables that affect water quality parameters. It provides a new analytical direction for the future management of water resources in river basins.

#### 4.3. Potential of the Model Application in Broader Regions

Most previous studies have focused on urban rivers or lakes with relatively small spatial scales and stable water quality, while few have addressed large-scale watersheds. On the one hand, it is impractical to use UAVs or other equipment for large-scale water

quality inversion; on the other hand, significant variations in water quality within large watersheds pose challenges for model construction. In this study, the model was developed using data from monitoring stations located in different segments of the RRB, enabling its application to the entire watershed. The model performance results indicate that it effectively predicts water quality parameters and is well-suited for diverse water bodies. Applying prediction models to fast-flowing and spatially extensive rivers—particularly across entire watersheds—remains a significant challenge. This study represents a valuable exploration of large-scale water quality inversion, offering technical support and a methodological framework for future research.

## 5. Conclusions

This study presents an innovative and interpretable modeling framework for large-scale river water quality inversion by integrating multi-source environmental data and heuristic-optimized machine learning techniques. Targeting six key water quality parameters (DO,  $\text{NH}_3\text{-N}$ , TP, TN, pH, and  $\text{COD}_{\text{Mn}}$ ), the proposed framework incorporates remote sensing imagery, meteorological variables, land use classification, and landscape pattern indices as input variables. A core methodological contribution lies in the use of the WOA to fine-tune hyperparameters of tree-based machine learning models (XGBoost and LightGBM), significantly enhancing model performance and generalization. In addition, the application of SHAP enables transparent interpretation of variable contributions, addressing the black-box nature of most machine learning approaches.

The optimized models achieved high predictive accuracy, with  $R^2$  values of 0.70 for DO, 0.46 for  $\text{NH}_3\text{-N}$ , 0.59 for TP, 0.71 for TN, 0.83 for pH, and 0.57 for  $\text{COD}_{\text{Mn}}$ , confirming their effectiveness across diverse water quality parameters. Notably, the integration of landscape metrics and meteorological elements proved crucial in improving the inversion of non-optically active water quality parameters—an area traditionally limited by the reliance on remote sensing reflectance alone. Furthermore, the framework was successfully applied at the scale of the entire RRB ( $\sim 61,700 \text{ km}^2$ ), demonstrating its scalability, robustness, and practical feasibility in large, heterogeneous river systems. The spatial prediction maps generated using 2023 satellite imagery provide fine-scale, continuous estimates of water quality that can support basin-wide environmental decision-making and pollution control strategies. In conclusion, this study offers both methodological innovation and practical utility. By coupling multi-source data fusion with heuristic-optimized, interpretable machine learning models, it advances the capabilities of remote sensing-based water quality monitoring. The framework is transferable to other basins and provides a valuable decision-support tool for hydrological management, especially in regions facing data scarcity or complex watershed dynamics.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs18020320/s1>, Table S1: The selected spectral bands or their combinations for the inversion of each water quality parameter.; Table S2: The hyperparameters of XGBoost and LightGBM optimized by the WOA.

**Author Contributions:** Conceptualization, Z.L. and F.W.; methodology, Z.L. and K.F.; formal analysis, Z.L. and K.F.; investigation, X.G.; data curation, K.F.; writing—original draft preparation, K.F.; writing—review and editing, Z.L., F.W. and X.G.; supervision, Z.L., F.W. and X.G.; funding acquisition, X.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China [Grant no. U2102208] and Major Project of the National Social Science Fund of China (Grant No. 23&ZD104).

**Data Availability Statement:** The data presented in this study are available upon request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

|                    |                                 |
|--------------------|---------------------------------|
| RF                 | Random Forest                   |
| XGBoost            | eXtreme Gradient Boosting       |
| LightGBM           | Light Gradient Boosting Machine |
| WOA                | Whale Optimization Algorithm    |
| SHAP               | Shapley Additive Explanations   |
| RRB                | Red River Basin                 |
| DO                 | dissolved oxygen                |
| COD <sub>Mn</sub>  | chemical oxygen demand          |
| TP                 | total phosphorus                |
| TN                 | total nitrogen                  |
| NH <sub>3</sub> -N | ammonia nitrogen                |

## References

- Guo, J.; Bian, R.; Guan, A.; Cao, X.; Peng, J.; Wu, X.; Wang, D.; Qi, W.; Liu, H.; Qu, J. Assessment and identification of primary factors controlling Yangtze River water quality. *ACS EST Water* **2023**, *3*, 1329–1340. [CrossRef]
- Yan, C.; Zhang, W.; Zhang, Z.; Liu, Y.; Deng, C.; Nie, N. Assessment of water quality and identification of polluted risky regions based on field observations & GIS in the Honghe River Watershed, China. *PLoS ONE* **2015**, *10*, e0119130.
- Feng, Z.; Zhang, Z.; Zuo, Y.; Wan, X.; Wang, L.; Chen, H.; Xiong, G.; Liu, Y.; Tang, Q.; Liang, T. Analysis of long term water quality variations driven by multiple factors in a typical basin of Beijing-Tianjin-Hebei region combined with neural networks. *J. Clean. Prod.* **2023**, *382*, 135367. [CrossRef]
- Ge, Y.; Wu, Z.; Chen, Y.; Guo, P.; Wu, A.; Liu, H.; Yuan, G.; Li, Y.; Fu, H.; Jeppesen, E. Cascading effects of human activities and ENSO on the water quality of Poyang Lake in China. *Catena* **2024**, *246*, 108380. [CrossRef]
- Ren, X.; Yang, C.; Zhao, B.; Xiao, J.; Gao, D.; Zhang, H. Water quality assessment and pollution source apportionment using multivariate statistical and PMF receptor modeling techniques in a sub-watershed of the upper Yangtze River, Southwest China. *Environ. Geochem. Health* **2023**, *45*, 6869–6887. [CrossRef]
- Bangira, T.; Matongera, T.N.; Mabhaudhi, T.; Mutanga, O. Remote sensing-based water quality monitoring in African reservoirs, potential and limitations of sensors and algorithms: A systematic review. *Phys. Chem. Earth Parts A/B/C* **2024**, *134*, 103536. [CrossRef]
- Kouadri, S.; Elbeltagi, A.; Islam, A.R.M.T.; Kateb, S. Performance of machine learning methods in predicting water quality index based on irregular data set: Application on Illizi region (Algerian southeast). *Appl. Water Sci.* **2021**, *11*, 190. [CrossRef]
- Wu, Z.; Lai, X.; Li, K. Water quality assessment of rivers in Lake Chaohu Basin (China) using water quality index. *Ecol. Indic.* **2021**, *121*, 107021. [CrossRef]
- Chen, P.; Li, W.; He, K. Impacts of different types of El Niño events on water quality over the Corn Belt, United States. *Hydrol Earth Syst Sc.* **2022**, *26*, 4875–4892. [CrossRef]
- Shahid, M.S.B.; Rifat, H.R.; Uddin, M.A.; Islam, M.M.; Mahmud, M.Z.; Sakib, M.K.H.; Roy, A. Hypertuning-based ensemble machine learning approach for real-time water quality monitoring and prediction. *Appl. Sci.* **2024**, *14*, 8622. [CrossRef]
- Lai, Y.; Zhang, J.; Li, W.; Song, Y. Water quality monitoring of large reservoirs in China based on water color change from 1999 to 2021. *J. Hydrol.* **2024**, *633*, 130988. [CrossRef]
- Sun, X.; Zhang, Y.; Shi, K.; Zhang, Y.; Li, N.; Wang, W.; Huang, X.; Qin, B. Monitoring water quality using proximal remote sensing technology. *Sci. Total Environ.* **2022**, *803*, 149805. [CrossRef] [PubMed]
- Chen, J.; Wang, K.; Li, X.; Zhao, X.; Cheng, X.; Liu, Z.; Zhang, J.; Zhang, S. Remote sensing inversion of water quality parameters (TSM, Chl-a, and CDOM) in subtidal seaweed beds and surrounding waters. *Ecol. Indic.* **2024**, *167*, 112716. [CrossRef]
- Cao, Q.; Yu, G.; Qiao, Z. Application and recent progress of inland water monitoring using remote sensing techniques. *Environ. Monit. Assess.* **2023**, *195*, 125. [CrossRef] [PubMed]

15. Chen, J.; Chen, S.; Fu, R.; Wang, C.; Li, D.; Peng, Y.; Wang, L.; Jiang, H.; Zheng, Q. Remote sensing estimation of chlorophyll-A in case-II waters of coastal areas: Three-band model versus genetic algorithm-artificial neural networks model. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2021**, *14*, 3640–3658. [CrossRef]
16. Thopanaiah, C.K.; CN, G.B.; Gurani, V.; Rajani, T.; Pavani, A.; Muthukumaran, D.; Deivasigamani, S. Advanced Remote Sensing and Generative Models for Comprehensive Water Quality Management in a Changing Climate. *Remote Sens. Earth Syst. Sci.* **2024**, *7*, 596–611. [CrossRef]
17. Jiang, Y.; Kong, J.; Zhong, Y.; Zhang, J.; Zheng, Z.; Wang, L.; Liu, D. The optimal method for water quality parameters retrieval of urban river based on machine learning algorithms using remote sensing images. *Int. J. Remote Sens.* **2024**, *45*, 7297–7317. [CrossRef]
18. Song, W.; Yinglan, A.; Wang, Y.; Fang, Q.; Tang, R. Study on remote sensing inversion and temporal-spatial variation of Hulun lake water quality based on machine learning. *J. Contam. Hydrol.* **2024**, *260*, 104282. [CrossRef]
19. Tian, D.; Zhao, X.; Gao, L.; Liang, Z.; Yang, Z.; Zhang, P.; Wu, Q.; Ren, K.; Li, R.; Yang, C. Estimation of water quality variables based on machine learning model and cluster analysis-based empirical model using multi-source remote sensing data in inland reservoirs, South China. *Environ. Pollut.* **2024**, *342*, 123104. [CrossRef]
20. Zhang, Y.; Li, W.; Wen, W.; Zhuang, F.; Yu, T.; Zhang, L.; Zhuang, Y. Universal high-frequency monitoring methods of river water quality in China based on machine learning. *Sci. Total Environ.* **2024**, *947*, 174641. [CrossRef]
21. Harkort, L.; Duan, Z. Estimation of dissolved organic carbon from inland waters at a large scale using satellite data and machine learning methods. *Water Res.* **2023**, *229*, 119478. [CrossRef] [PubMed]
22. Liang, Y.; Ding, F.; Liu, L.; Yin, F.; Hao, M.; Kang, T.; Zhao, C.; Wang, Z.; Jiang, D. Monitoring water quality parameters in urban rivers using multi-source data and machine learning approach. *J. Hydrol.* **2025**, *648*, 132394. [CrossRef]
23. Bullard, J.E.; Ockelford, A.; O'Brien, P.; Neuman, C.M. Preferential transport of microplastics by wind. *Atmos. Environ.* **2021**, *245*, 118038. [CrossRef]
24. Tiyyasha, T.; Tung, T.M.; Bhagat, S.K.; Tan, M.L.; Jawad, A.H.; Mohtar, W.H.M.W.; Yaseen, Z.M. Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: Development of hybrid tree-based artificial intelligence models. *Mar. Pollut. Bull.* **2021**, *170*, 112639. [CrossRef]
25. Murphy, S.F.; Runkel, R.L.; Stets, E.G.; Nolan, A.J.; Repert, D.A. Urbanization and water management control stream water quality along a mountain to plains transition. *Water Resour. Res.* **2024**, *60*, e2023WR035633. [CrossRef]
26. Wei, W.; Gao, Y.; Huang, J.; Gao, J. Exploring the effect of basin land degradation on lake and reservoir water quality in China. *J. Clean. Prod.* **2020**, *268*, 122249. [CrossRef]
27. Xu, J.; Bai, Y.; You, H.; Wang, X.; Ma, Z.; Zhang, H. Water quality assessment and the influence of landscape metrics at multiple scales in Poyang Lake basin. *Ecol. Indic.* **2022**, *141*, 109096. [CrossRef]
28. Shen, Z.; Hou, X.; Li, W.; Aini, G. Relating landscape characteristics to non-point source pollution in a typical urbanized watershed in the municipality of Beijing. *Landsc. Urban Plan.* **2014**, *123*, 96–107. [CrossRef]
29. Tanaka, T.; Sato, T.; Watanabe, K.; Wang, Y.; Yang, D.; Inoue, H.; Li, K.; Inamura, T. Irrigation system and land use effect on surface water quality in river, at lake Dianchi, Yunnan, China. *J. Environ. Sci.* **2013**, *25*, 1107–1116. [CrossRef]
30. Wang, J.; Yin, W.; Wang, L.; Li, N.; Hao, R.; Xu, J.; Xiao, H.; Zhong, Z.; Zhang, Y.; Shi, Z. Unraveling the effects of hydrological connectivity and landscape characteristics on reservoir water quality. *J. Hydrol.* **2022**, *613*, 128410. [CrossRef]
31. Chen, J.; Fan, Y.; Zhang, Y.; Peng, J.; Zhang, J.; Cao, C. Comprehensive propagation characteristics between paired meteorological and hydrological drought events: Insights from various underlying surfaces. *Atmos. Res.* **2024**, *299*, 107193. [CrossRef]
32. Deng, C.; Liu, L.; Li, H.; Peng, D.; Wu, Y.; Xia, H.; Zhang, Z.; Zhu, Q. A data-driven framework for spatiotemporal characteristics, complexity dynamics, and environmental risk evaluation of river water quality. *Sci. Total Environ.* **2021**, *785*, 147134. [CrossRef] [PubMed]
33. GB 3838-2002; Environmental quality standard for surface water. Ministry of Ecology and Environment, PRC: Beijing, China, 2002.
34. Wu, J.; Cheng, S.; He, L.; Wang, Y.; Yue, Y.; Zeng, H.; Xu, N. Assessing water quality in the Pearl River for the last decade based on clustering: Characteristic, evolution and policy implications. *Water Res.* **2023**, *244*, 120492. [CrossRef] [PubMed]
35. Peng, K.; Si, B.; Jiang, W.; Ma, M.; Wang, X. Annual river dataset in China: A new product with a 10 m spatial resolution from 2016 to 2023. *Earth Syst. Sci. Data Discuss* **2024**, *2024*, 1–26.
36. McGrane, S.J. Impacts of urbanisation on hydrological and water quality dynamics, and urban water management: A review. *Hydrol. Sci. J.* **2016**, *61*, 2295–2311. [CrossRef]
37. Peng, S.; Li, S. Scale relationship between landscape pattern and water quality in different pollution source areas: A case study of the Fuxian Lake watershed, China. *Ecol. Indic.* **2021**, *121*, 107136. [CrossRef]

38. Zhang, F.; Chen, Y.; Wang, W.; Jim, C.Y.; Zhang, Z.; Tan, M.L.; Liu, C.; Chan, N.W.; Wang, D.; Wang, Z. Impact of land-use/land-cover and landscape pattern on seasonal in-stream water quality in small watersheds. *J. Clean. Prod.* **2022**, *357*, 131907. [CrossRef]
39. McGarigal, K. *FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure*; US Department of Agriculture, Forest Service, Pacific Northwest Research Station: Portland, OR, USA, 1995.
40. Teixidó, N.; Garrabou, J.; Arntz, W.E. Spatial pattern quantification of Antarctic benthic communities using landscape indices. *Mar. Ecol. Prog. Ser.* **2002**, *242*, 1–14. [CrossRef]
41. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Software* **2016**, *95*, 51–67. [CrossRef]
42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
43. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
44. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T. Lightgbm: A Highly Efficient Gradient Boosting Decision Tree. *NeurIPS* **2017**, 3149–3157. Available online: [https://papers.nips.cc/paper\\_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html](https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html) (accessed on 14 January 2026).
45. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4774.
46. Wang, W.; Yang, P.; Xia, J.; Huang, H.; Li, J. Impact of land use on water quality in buffer zones at different scales in the Poyang Lake, middle reaches of the Yangtze River basin. *Sci. Total Environ.* **2023**, *896*, 165161. [CrossRef]
47. Xiao, R.; Wang, G.; Zhang, Q.; Zhang, Z. Multi-scale analysis of relationship between landscape pattern and urban river water quality in different seasons. *Sci. Rep.* **2016**, *6*, 25250. [CrossRef]
48. Zhao, B.; Zhang, A.; Wang, H.; Pang, J.; Hou, Y.; Ma, P.; Zhao, B. Water function zone: A method to improve the accuracy of remote sensing retrieval of water bodies. *Ecol. Indic.* **2024**, *164*, 112105. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



## Review

# Advances in Laser-Induced Acoustic Technology for Underwater Detection

Jin Zhao <sup>1</sup>, Kexin Yu <sup>1</sup>, Shuaiqi Xu <sup>1</sup>, Maorong Wang <sup>1,\*</sup>, Yiguang Yang <sup>2,\*</sup>, Degang Xu <sup>3</sup>, Jianquan Yao <sup>3</sup> and Xia Wang <sup>1,4</sup>

<sup>1</sup> College of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao 266061, China; orangeade95@163.com (J.Z.)

<sup>2</sup> Laoshan Laboratory, Qingdao 266061, China

<sup>3</sup> College of Precision Instrument and Optoelectronic Engineering, Tianjin University, Tianjin 300072, China; jqyao@tju.edu.cn (J.Y.)

<sup>4</sup> College of Physical Science and Technology, Jinan University, Jinan 250024, China

\* Correspondence: qust@vip.126.com (M.W.); yangyiguang1981@163.com (Y.Y.); Tel.: +86-0532-88958628 (M.W.)

**Abstract:** Laser-induced acoustic (LIA) underwater detection, as a next-generation sensing paradigm, combines high spatial resolution, rapid temporal response, and cross-medium detection capability, positioning it as a strategically significant technology in marine resource exploration, military security, and ocean environmental monitoring. The fundamental principles underlying LIA technology are systematically examined, together with recent advances in representative experimental systems and critical enabling techniques. The characteristics of the laser–acoustic transmission channel are comprehensively investigated, and the mechanisms through which laser parameters modulate the properties of acoustic signals are rigorously elucidated. Moreover, several challenges hindering practical applications are underscored, including laser energy attenuation, interference arising from complex underwater environments, and the comparatively high cost of equipment. Finally, future research directions are outlined, encompassing the development of high-efficiency laser sources, multimodal integrated sensing strategies, intelligent signal processing algorithms, and improved environmental adaptability. These efforts are intended to provide theoretical underpinnings for the continued advancement and broader application of LIA-based underwater detection technologies.

**Keywords:** Laser-Induced Acoustic (LIA); underwater detection; underwater acoustics; cross-medium propagation; marine exploration

## 1. Introduction

Conventional underwater detection technologies rely on electromagnetic waves and mechanical sonar systems [1]. However, electromagnetic waves suffer from strong attenuation in seawater with an attenuation coefficient of approximately  $10^3$  dB/km [2], while sonar systems face limitations such as low transduction efficiency, narrow bandwidth, and susceptibility to mechanical disturbances. LIA technology generates broadband acoustic waves through the interaction between high-energy pulsed lasers and water media. By leveraging an optical-to-acoustic energy conversion mechanism, LIA enables non-contact acoustic source excitation and circumvents the immersion-related disturbances inherent to traditional sonar, offering an innovative solution for deep-sea exploration [3–5]. A comparison of the advantages and disadvantages, as well as the detection parameters, of various

underwater detection technologies is presented in Table 1. Overall, different underwater sensing technologies present distinct trade-offs across range, resolution, environmental adaptability, and system cost, among which LIA demonstrates clear advantages in handling complex sea states, achieving high resolution, and enabling long-range detection. The core advantages of this technology are reflected in three aspects. First, the integration of air-optical and water-acoustic dual channels utilizes the low-loss transmission of lasers in the atmosphere and the low attenuation of acoustics in water, substantially extending long-range detection. Second, the broadband acoustic source supports high-resolution target recognition, enabling the detection of millimeter-scale underwater structures. Third, by tuning laser parameters such as single-pulse energy and pulse duration, the spectral characteristics of the generated acoustic signals can be tailored to adapt to complex marine environments. With an underwater transmission range exceeding one hundred meters, LIA provides technological support for submarine communication, mineral exploration, and ecological environment monitoring. Moreover, the continued development of this technology is expected to advance the intelligence of marine equipment and diversify detection modalities, carrying profound implications for safeguarding maritime interests, strengthening national defense security, and fostering marine economic growth.

**Table 1.** Comparison of Underwater Detection Technology.

| Method                         | Typical Platforms                                                                                                                 | Strengths                                                                                                        | Core Constraints                                                                                                    | Representative Performance                                                |                                                                                                |
|--------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|
|                                |                                                                                                                                   |                                                                                                                  |                                                                                                                     | Detection Depth                                                           | Resolution                                                                                     |
| Active Sonar                   | Shore-based; surface ships; submarines; airborne;                                                                                 | Active emission; long-range; high-resolution [7]                                                                 | Operational exposure; high energy demand; sensitivity to environmental conditions [8]                               | Over 10 km given favorable sound-channel conditions [9,10]                | At 220–260 kHz, Synthetic aperture sonar (SAS) can deliver 3 cm × 2 cm imaging resolution [11] |
| Passive Sonar                  | unmanned surface Vehicle (USV) and unmanned underwater Vehicle (UUV) [6]                                                          | Covert operation, low-power, and persistent surveillance capability [12,13]                                      | Limited detection range; incomplete information; and constrained sensitivity to low-acoustic-signature targets [14] | In general, the propagation depth is within 200 m [15,16]                 | /                                                                                              |
| Airborne Blue-Green LiDAR      | Fixed wing; helicopters; unmanned aerial vehicles (UAVs), satellites [17,18]                                                      | Highly maneuverable; high-resolution; with strong penetration through water [19]                                 | Constrained by water turbidity; strongly affected by environmental conditions; high cost [20]                       | Achieved a detection depth of 94 m [21]                                   | The spatial resolution is generally between 10 cm and 12 cm ( $\pm 0.2$ m) [22,23]             |
| Laser Doppler vibrometer (LDV) | Single-point LDV; Scanning-LDV; Continuous-scan LDV (CSLDV); Full-field / 3-D LDV; Fiber-coaxial LDV; Mobile-platform LDV [24–28] | High accuracy, good linearity, fast dynamic response, wide measurement range, non-contact measurement [24,29–36] | High equipment cost; Sensitivity to environmental disturbances; Laser phase noise [36]                              | Drone-mounted LDV can achieve a depth of field of approximately 10 m [30] | LDVs can achieve femtometer-scale ( $10^{-15}$ m) displacement resolution.                     |

Table 1. Cont.

| Method                  | Typical Platforms                         | Strengths                                                                                                                | Core Constraints                                                                                                                 | Representative Performance                                                                                                                                                                 |                                                                                                                                    |
|-------------------------|-------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|
|                         |                                           |                                                                                                                          |                                                                                                                                  | Detection Depth                                                                                                                                                                            | Resolution                                                                                                                         |
| Fiber-Optic Hydrophones | MEDUSA multi-cloud demonstrator [31]      | Immunity to electromagnetic interference; Small size & light weight<br>Broad bandwidth & multiplexing capability [32,33] | Limited low-frequency response; Bandwidth ceiling; Complex optical interrogation [31,33,34]                                      | 3000 m ( $\approx 30$ MPa hydrostatic pressure) [35,36]                                                                                                                                    |                                                                                                                                    |
| Laser-Induced Acoustics | Shore-based; surface ships; airborne [37] | Non-contact; high resolution; adaptable to complex marine environments [38]                                              | Laser-generated acoustic sources cannot propagate over long distances; the optical-to-acoustic conversion efficiency is low [39] | Finite-element simulations indicate that when the pulse reaches a power density of $2.8 \times 10^{10}$ W/cm <sup>2</sup> , the acoustic wave can propagate to a depth of up to 400 m [40] | A laser system (wavelength of 10.6 $\mu$ m and a pulse width of 20 ns, delivering 500 W peak power) can image a 1.3 cm sphere [41] |

Research on LIA technology began in 1962, when White and Prokhorov first reported the phenomenon of acoustic wave generation by pulsed lasers in condensed media. In 1973, they experimentally revealed the spectral and spatial distribution characteristics of laser-generated acoustic signals [42]. From 1980 onward, research in this field entered a stage focused on technical validation and system development. The Hickman team in the United States made a series of pioneering contributions: in 1980, they achieved underwater detection at a depth of 20 m using a CO<sub>2</sub> laser, providing the first experimental verification of the feasibility of laser-induced underwater detection [43]; in 1981, they developed the Experimental Laser / Acoustic System (ELAS), which enabled depth sounding in turbid shallow waters using a CO<sub>2</sub> laser, and extended the system's applications to submarine communication and thermocline detection [44]; in 1987, they demonstrated 1 m ice thickness measurement using a 6.5 J laser and proposed a linear relationship model between ice thickness and laser energy [45]; and by 1990, they replaced the CO<sub>2</sub> laser with a Nd:YAG laser to enhance the sensitivity of ice-layer detection [44].

With the advancement of technology, increasing attention has been devoted to multiphysics coupling research in LIA since 2015, accelerating its transition toward practical applications. In 2016, Bidin et al. established a quantitative relationship between acoustic velocity/amplitude and hydrocarbon concentration in water based on laser-induced breakdown (LIB) technology, introducing a novel approach to pollutant monitoring [46]. In the same year, Jukna reported that broadband acoustic signals in the 0.1–10 MHz range generated by terawatt laser filamentation exhibited spatial distributions constrained by the laser propagation direction [47]. In 2020, the Fitzpatrick team developed a laser-CMUT (capacitive micromachined ultrasonic transducers) detection system, achieving millimeter-scale resolution in underwater imaging [41]. In parallel, Baimler et al. revealed the acousto-chemical coupling mechanism during optical breakdown in a nickel nanoparticle solution, confirming a strong correlation between the generated acoustic signals and the formation of molecular gases [48]. In 2021, Yellaiah et al. systematically investigated the time–frequency characteristics of shock waves induced by nanosecond (ns) pulses. Finite-element simulations demonstrated that interactions between a shock wave and a gas interface cause

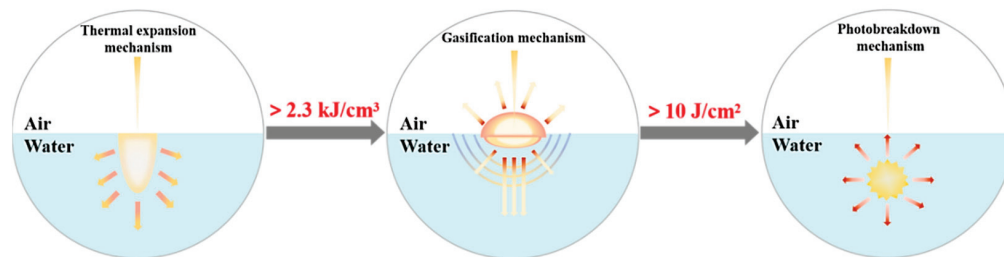
forward attenuation in acoustic signals and showed that the spectral features can serve as indicators of energy transfer efficiency at the interface [49]. From early engineering validation dominated by CO<sub>2</sub> lasers to recent studies involving ultrafast lasers and multi-physics coupling, the effective frequency domain of LIA has expanded from the hertz to the megahertz range, while its application scope has evolved from simple depth sounding to encompass chemical sensing and high-resolution imaging.

Research on LIA technology in China began in 1986, when Li, R. and Cui, G.'s team at the 706 Research Institute in Dalian took the lead in exploring laser–acoustic remote sensing. In 1990, the team successfully achieved ocean depth detection of 17 m and underwater target identification using a single microphone [50,51]. In 1997, they developed China's first independently designed laser–acoustic remote sensing system, which enabled underwater target detection at a depth of 30 m and bathymetric measurement down to 58 m, thereby validating the feasibility of laser–acoustic coupling for underwater detection [52–54]. Recent key technological breakthroughs in China have focused on acoustic source optimization, communication applications, and efficient processing of underwater acoustic signals. In 2009, Zhu, H. et al. experimentally demonstrated a nonlinear relationship between the amplitude of acoustic signals in saline water and both salinity and laser energy using pulsed CO<sub>2</sub> lasers. In 2010, Li, Q. et al. developed an all-fiber laser–acoustic source system, enabling covert and high-speed underwater acoustic communication; their optical–acoustic coupling model laid the foundation for engineering applications [55]. In 2019, Wang, W. et al. proposed a frequency-domain energy detection method to quantitatively analyze the attenuation characteristics of acoustic signals in the 0.1–10 MHz range [56]. To address the low bandwidth utilization of chaotic signals, domestic researchers have proposed adaptive modulation algorithms to improve the transmission efficiency of underwater acoustic channels. In addition, significant progress has been made in understanding the laser-to-acoustic energy conversion mechanisms and optimizing the underwater propagation of acoustic signals [57]. The research team led by Wang, J., designed a high-sensitivity laser–acoustic transducer, achieving a 40% increase in acoustic signal amplitude. Wang Y. et al. conducted cavitation dynamics experiments and established a quantitative mapping model linking laser energy, cavitation oscillations, and acoustic pressure characteristics [58]. Wang, Y. et al. identified the optimal repetition frequency for LIAs to be in the range of 10–100 Hz through frequency-domain analysis [59]. Li, S. et al. demonstrated that low-frequency acoustic waves (10 kHz) experience lower transmission loss in complex liquid media. Zong, S. et al. systematically elucidated the relationship between optical breakdown thresholds and the spectral characteristics of cavitation oscillations [60, 61]. Current research in China is increasingly focused on enhancing optoacoustic coupling efficiency, characterizing acoustic signal propagation in complex media, and developing interference-resistant techniques for both high- and low-frequency regimes, thereby providing theoretical foundations for deep-sea exploration and covert communication.

## 2. Mechanisms of LIAs

When a high-intensity pulsed laser is focused at the gas–liquid interface, strong absorption of the liquid medium at specific wavelengths results in significant energy deposition. Within a timescale ranging from ns to microseconds ( $\mu$ s), the deposited laser energy sequentially triggers three distinct mechanisms for acoustic wave generation: thermoelastic expansion, phase-change vaporization, and LIB plasma. Each of these mechanisms is associated with a specific laser energy threshold. As shown in Figure 1, the linear photoacoustic conversion mechanism dominated by thermoelastic expansion differs significantly from the nonlinear mechanism governed by phase-change vaporization. While the latter exhibits higher energy conversion efficiency, its controllability is considerably lower due to the strongly un-

steady nature of bubble dynamics during the vaporization process. Overcoming the challenge of controllable modulation of nonlinear acoustic sources has become a key scientific problem that urgently needs to be addressed in the field of photoacoustic conversion.

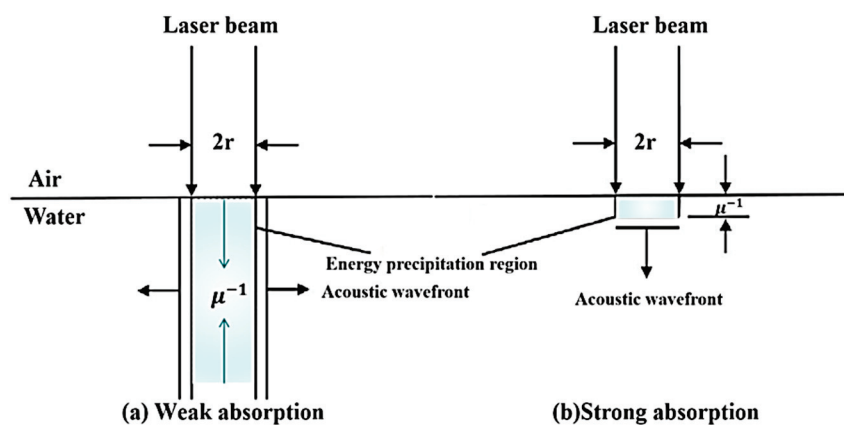


**Figure 1.** A schematic diagram of three mechanisms of LIA.

### 2.1. Thermoelastic Mechanism

When the laser energy density is below the vaporization threshold of the liquid, thermoelastic photoacoustic waves are generated. Upon laser irradiation of the liquid surface or within the liquid, the medium rapidly absorbs the incident energy, producing non-uniform heating in the irradiated region and leading to an inhomogeneous temperature field. Under thermal coupling, local thermal expansion induces thermoelastic pressure, which in turn generates acoustic waves propagating outward from the heat-absorbing region. These waves are referred to as thermoelastic photoacoustic waves [62]. The thermoelastic mechanism produces acoustic waves with high repeatability and avoids phase transitions in the liquid, thereby preserving its intrinsic physical properties. Consequently, it is widely employed in nondestructive testing applications. However, the photoacoustic conversion efficiency of thermoelastic waves is typically below 0.1%, resulting in relatively weak acoustic amplitudes. This limitation imposes stringent sensitivity requirements on acoustic signal receivers [63].

Assuming a laser beam with a characteristic radius  $r$  and an absorption coefficient  $\mu$  for water, the optical extinction depth  $\mu^{-1}$  is defined as the distance at which the laser energy attenuates to  $1/e$  of its initial value. As illustrated in Figure 2, the shaded region represents the energy deposition zone within the liquid medium, which can be classified into two regimes: strong absorption and weak absorption. When  $r\mu \ll 1$  the condition corresponds to weak absorption; conversely,  $r\mu \gg 1$  represents strong absorption. In water, laser wavelengths with high absorption coefficients, combined with large beam divergence angles, are favorable for generating stronger LIA signals. On the other hand, if the laser wavelength falls outside the absorption band, the beam can propagate longer distances through water with minimal attenuation [64].



**Figure 2.** The energy distribution after irradiation of water by a laser.



Table 2 summarizes the relationship between the optical extinction length in water and commonly used laser wavelengths, where  $\lambda$  denotes the laser wavelength and  $\mu$  represents the absorption coefficient. The extinction lengths corresponding to CO<sub>2</sub> and Nd:YAG lasers differ significantly, as water exhibits much stronger absorption at the CO<sub>2</sub> laser wavelength compared to that of the Nd:YAG laser. Therefore, laser sources with appropriate wavelengths can be selected according to specific experimental objectives [65].

**Table 2.** Absorption coefficient of water for different wavelengths of laser [65].

| Laser Source          | $\lambda$ ( $\mu\text{m}$ ) | $\mu$ ( $\text{cm}^{-1}$ ) |
|-----------------------|-----------------------------|----------------------------|
| Ruby laser            | 0.6943                      | $5.3 \times 10^{-3}$       |
| Nd:YAG laser          | 1.06                        | $1.4 \times 10^{-3}$       |
| CO <sub>2</sub> laser | 10.6                        | $8.7 \times 10^{-3}$       |
| Er: YAG laser         | 2.94                        | $1.32 \times 10^{-3}$      |

## 2.2. Phase-Change Vaporization Mechanism

In the photoacoustic energy conversion process within liquid media, vaporization-induced recoil effects can generate mechanical forces, thereby producing characteristic acoustic pulse signals. The laser energy density threshold required for vaporization-induced acoustic generation is approximately  $2.3 \text{ J/cm}^3$  [66]. Two typical vaporization modes are generally observed: A steady-state surface vaporization process occurring at the gas–liquid interface, and a transient explosive boiling process occurring in the subsurface region. The transition between these dominant modes is primarily governed by thermophysical parameters of the liquid, such as specific heat capacity and latent heat of vaporization. The photoacoustic conversion efficiency of vaporization processes can reach up to 1%, with the transient explosive boiling mode exhibiting superior acoustic generation efficiency compared to the steady-state surface vaporization mode [67,68].

Fletcher Blackmon et al. performed experiments under laboratory freshwater conditions using a pulsed laser with a 1 kHz repetition rate. In addition to inducing surface vaporization of the water medium, the results revealed the possibility of nonlinear interaction mechanisms. This nonlinear effect originates from the expansion and collapse of cavitation bubbles within the laser focal volume, giving rise to explosive boiling phenomena. The experiments demonstrated that the underwater sound pressure level (SPL) could reach up to 185 dB, and, for the first time, introduced the concept of nonlinear photoacoustic conversion [5].

## 2.3. Laser-Induced Optical Breakdown Mechanism

When the laser energy density exceeds the optical breakdown threshold of the liquid ( $10^7 \text{ W/cm}^2$ ) [69], plasma is generated within the medium. This plasma continues to absorb optical energy, triggering an explosive expansion that produces outward-propagating shock waves. Approximately 65–85% of the shock wave energy is dissipated in water and eventually attenuates into acoustic waves. Compared to thermoelastic and vaporization mechanisms, the optical breakdown mechanism offers significantly higher photoacoustic conversion efficiency, ranging from 7 to 30% and thus serves as the dominant mechanism for high-power pulsed LIA generation. Table 3 presents the reported thresholds for optical breakdown.

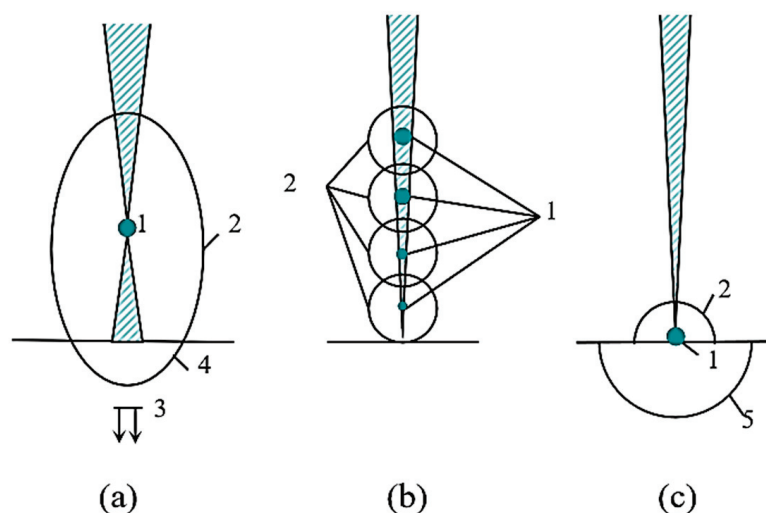
**Table 3.** Optical breakdown threshold [70].

| Pulse Duration<br>/(ns) | Wavelength<br>/( $\mu\text{m}$ ) | Convergence<br>Angle/ $^\circ$ | Focused Spot<br>Diameter/( $\mu\text{m}$ ) | Threshold<br>Power/(1010 W·<br>$\text{cm}^2$ ) | Threshold Energy<br>/(J· $\text{cm}^{-2}$ ) |
|-------------------------|----------------------------------|--------------------------------|--------------------------------------------|------------------------------------------------|---------------------------------------------|
| 76                      | 0.75                             | 20                             | 20.1                                       | 2.3                                            | 1760                                        |
| 6                       | 1.06                             | 33                             | 5.5                                        | 6.6                                            | 399                                         |
| 6                       | 1.06                             | 21                             | 7.7                                        | 4.7                                            | 285                                         |
| 6                       | 1.06                             | 5.4                            | 14.5                                       | 11                                             | 650                                         |
| 0.03                    | 1.06                             | 27                             | 4.6                                        | 46                                             | 13.7                                        |
| 0.03                    | 1.06                             | 8.5                            | 9.7                                        | 45                                             | 13.5                                        |
| 0.03                    | 1.06                             | 4                              | 19.6                                       | 37                                             | 11.1                                        |
| 0.06                    | 0.53                             | 13                             | 5.5                                        | 28                                             | 16.9                                        |
| $3 \times 10^{-3}$      | 0.58                             | 16                             | 5.0                                        | 85                                             | 2.6                                         |
| $3 \times 10^{-4}$      | 0.58                             | 16                             | 5.1                                        | 476                                            | 1.4                                         |
| $10^{-4}$               | 0.58                             | 16                             | 4.4                                        | 1111                                           | 1.1                                         |

In cross-medium LIA processes, the spatial distribution of the optical breakdown threshold is influenced by multiple factors, including laser focal spot size, aerosol concentration, and beam incidence angle. These factors lead to spatial variability in the breakdown location. For example, using a CO<sub>2</sub> laser with a wavelength of 10.6  $\mu\text{m}$  and a focal spot diameter of 0.3 mm, three typical optical breakdown modes at the air–water interface have been experimentally observed. When a high-energy laser beam (15 J/cm<sup>2</sup>) is incident, optical breakdown occurs preferentially in the air region above the interface, as illustrated in Figure 3a. Due to plasma shielding effects, the trailing edge of the laser pulse is truncated, resulting in the generation of bipolar acoustic pulses with thermoelastic characteristics and peak pressures on the order of megapascals. The accompanying air shock wave, upon crossing the air–liquid interface, experiences a pressure amplitude jump caused by the abrupt acoustic impedance mismatch between the two media. This phenomenon follows the shock wave transmission behavior described by the Rankine–Hugoniot equations. Under low-divergence laser irradiation, a multi-focal breakdown structure is formed near the interface, as shown in Figure 3b. The shock waves generated by independently evolving plasma clusters exhibit nonlinear superposition during propagation, leading to a complex transient pressure field distribution. When the energy density decreases to approximately 10–15 J/cm<sup>2</sup>, optical breakdown occurs only at the surface layer of the water, as illustrated in Figure 3c. At this stage, a prominent surface flash is observed, accompanied by the simultaneous generation of shock waves in both the air and liquid phases. Notably, in laser acoustic remote sensing for underwater detection, the air shock wave and the acoustic pulse in water jointly form a composite acoustic source. The coupling characteristics of their propagation play a critical role in determining the detection sensitivity [69].

The dynamic behavior of cavitation bubbles and their acoustic radiation characteristics constitute an important research direction in LIAs. The generation and collapse of cavitation bubbles can induce high-intensity acoustic pulses, with their dynamics strongly influenced by boundary conditions and medium properties. In 2004, Chen, X. et al. [71] conducted foundational theoretical and experimental studies on bubble dynamics, characterizing bubble wall expansion, maximum and minimum diameters, and mechanical shock features during late-stage collapse. In 2007, Zhao, R. et al. [72] improved the experimental system to investigate the influencing factors of bubble generation and collapse and analyzed the acoustic signatures of cavitation near solid boundaries. In 2010, Zong, S. et al. [73] focused on cavitation erosion effects triggered by near-wall bubble collapse. In complex media, Li, S. et al. [74] in 2012 found that the dissolved gas content in liquids significantly modulates the frequency spectrum of cavitation-induced acoustic signals.

Subsequent studies [75] revealed that double-bubble systems produce higher acoustic pressure amplitudes than single-bubble systems, while maintaining similar dominant frequency characteristics. In 2015, Li. further demonstrated through controlled pressure field experiments that pressure gradients can induce asymmetric bubble collapse, resulting in characteristic waveform distortions in the emitted acoustic pulses [76]. These findings provide a theoretical foundation for optimizing underwater acoustic sources and mitigating cavitation-induced damage.



**Figure 3.** Dielectric breakdown caused by different laser beam structures: (a) air breakdown above the interface; (b) multiple air breakdowns above the interface; (c) water breakdown on the water surface. 1—Laser spark; 2—Shockwave in air; 3—Thermoelastic photoacoustic waves; 4—Shockwave entering water from air; 5—Shockwave in water.

#### 2.4. Cross-Medium Propagation Characteristics of LIA Waves

In the atmosphere, laser propagation is influenced by the composition of the air. Table 4 lists the major atmospheric components, with those of higher concentration having a greater impact on laser transmission. A large number of gas molecules and aerosol particles of various sizes in the air cause scattering of the laser beam, primarily through Rayleigh scattering and Mie scattering mechanisms [64]. These scattering processes lead to attenuation of the laser beam during propagation. The extent of scattering depends on the laser wavelength, particle size distribution, and the concentration of atmospheric constituents. These factors collectively affect the transmission efficiency of laser systems under different atmospheric conditions.

**Table 4.** Low atmospheric composition [64].

| Constant Component |                               | Varying Component |                          |
|--------------------|-------------------------------|-------------------|--------------------------|
| Name               | Volume Percentage             | Name              | Volume Percentage        |
| N                  | 78.084                        | H                 | $0.5 \times 10^{-4}$     |
| O                  | 20.948                        | N <sub>2</sub> O  | $0.27 \times 10^{-4}$    |
| Ar                 | 0.934                         | CO                | $0.19 \times 10^{-4}$    |
| water vapor        | 0~0.04                        | Xe                | $0.089 \times 10^{-4}$   |
| CO                 | 0.033                         | NH <sub>3</sub>   | $0.004 \times 10^{-4}$   |
| Ne                 | $18.18 \times 10^{-4}$        | SO <sub>2</sub>   | $0.001 \times 10^{-4}$   |
| O <sub>3</sub>     | $0\text{--}13 \times 10^{-4}$ | NO <sub>2</sub>   | $0.001 \times 10^{-4}$   |
| He                 | $5.24 \times 10^{-4}$         | NO                | $0.0005 \times 10^{-4}$  |
| CH <sub>4</sub>    | $1.5 \times 10^{-4}$          | H <sub>2</sub> S  | $0.00005 \times 10^{-4}$ |

Secondly, the directivity of the acoustic source has a notable influence on the SPL. Let  $r$  denote the distance from the measurement point to the acoustic source, and  $\theta$  represent the angle between the normal direction of the acoustic source and the line connecting the source to the detector.  $K = \omega/c_1$  denotes the wavenumber, where  $c_1$  is the speed of sound in water, and  $A$  is the amplitude constant of the acoustic wave with angular frequency. The acoustic pressure with an angular frequency  $\omega$  at a propagation distance  $r$ , generated by the laser-induced acoustic wave in water, can be expressed as

$$p = \frac{A \cos \theta}{r} e^{i(kr - \omega t)} \quad (1)$$

Due to the expansion of the acoustic wavefront, its effect on the SPL can be described by Equation (2):

$$L_{tw} = 20 \lg r - 20 \lg (\cos \theta) \quad (2)$$

Based on the quantitative relationship between acoustic source directivity and SPL, the acoustic field distribution can be divided into two characteristic regions. For angles  $\theta < 50^\circ$ , the SPL exhibits low sensitivity to the radiation direction, which is primarily attributed to the coherent superposition effect along the axial acoustic field. In contrast, for  $\theta \geq 50^\circ$ , the SPL shows significant directional dependence, mainly due to the coupling between edge diffraction effects and acoustic wave interference [69].

In atmospheric channels, the wavelength of  $1.06 \mu\text{m}$  is a typical “atmospheric window,” characterized by high transmittance and low propagation attenuation. In practical LIA experiments, most of the laser energy at this wavelength can be effectively utilized. Under low-visibility conditions, the laser is capable of maintaining short-range propagation, whereas under high-visibility conditions, it exhibits significantly higher transmittance [64].

When a laser crosses the air–water interface, energy loss occurs due to reflection at the boundary, and the corresponding loss in SPL is dependent on the laser incidence angle at the water surface. The acoustic energy loss across the air–water interface exhibits a nonlinear increase with the incident angle. For incidence angles between  $0^\circ$  and  $80^\circ$ , the interface-induced loss remains below 45 dB and increases gradually. However, when the incidence angle exceeds  $80^\circ$ , the loss rises exponentially, reaching values as high as 110 dB. Even under normal incidence, each laser crossing of the interface results in a SPL loss of approximately 65 dB [69].

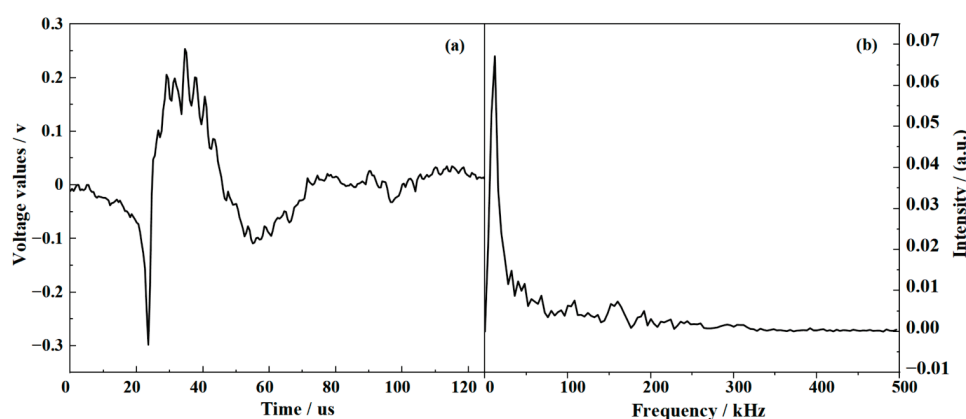
The acoustic waves generated by laser incidence propagate in water through energy transfer mechanisms that rely on vibrational coupling between adjacent water molecules. During propagation, part of the acoustic energy is gradually converted into thermal energy, leading to attenuation of the acoustic signal. This energy loss is primarily attributed to viscous absorption and molecular relaxation processes. W.H. Thorp provided an empirical formula for calculating the acoustic attenuation coefficient in seawater. This formula is applicable only to acoustic waves with frequencies greater than 5 kHz [77]:

$$\alpha \approx 3 \times 10^{-4} f^2 + \frac{44 f^2}{4100 + f^2} + \frac{0.11 f^2}{1 + f^2} \quad (3)$$

In the equation,  $\alpha$  denotes the acoustic attenuation coefficient in seawater (dB/km), and  $f$  represents the acoustic frequency (kHz).  $f^2$  represents viscous absorption in water;  $44f^2/4100 + f^2$  represents relaxation absorption due to borate ions;  $0.11f^2/(1 + f^2)$  represents relaxation effects caused by magnesium sulfate ions below 100 kHz. The acoustic attenuation coefficients in freshwater and seawater vary with frequency, and at the same acoustic frequency, the attenuation in seawater is consistently higher than that in freshwa-

ter. In freshwater, the absorption coefficient increases approximately linearly with rising acoustic frequency, whereas in seawater, the increase follows a nonlinear trend [64]. This indicates that the low-frequency components of LIA signals experience less attenuation in water, particularly in cross-air–water propagation scenarios, making the low-frequency band more favorable for underwater information transmission.

The time–frequency and propagation characteristics of typical laser-induced acoustic signals were analyzed by our team. A laser system with a wavelength of 1064 nm, a pulse repetition rate of 1–10 Hz, a pulse width of 6–8 ns, and a single-pulse energy of 260 mJ was employed. The time-frequency characteristics of the acoustic signal generated at a depth of 20 cm below the lake surface are presented in Figure 4. The laser-induced acoustic waveform exhibits a characteristic bipolar profile with a pulse duration of approximately 125  $\mu$ s and a frequency range spanning 0–500 kHz. The signal energy is primarily distributed within the 0–50 kHz band, with a dominant peak at 24 kHz. The variations in SPL and signal-to-noise ratio (SNR) with depth were investigated. Both SPL and SNR decrease progressively with increasing depth, and a nonlinear fitting yielded an excellent correlation ( $R^2 = 0.99968$ ), indicating that the model provides an accurate representation of the observed trend. This finding is consistent with the theoretical framework of acoustic wave propagation in stratified media, suggesting that depth induces a nonlinear attenuation effect on the acoustic signal. The result provides a theoretical basis for evaluating SPL behavior at varying depths [37].



**Figure 4.** Laser acoustic source signal and its spectrogram. (a) Time-domain plot; (b) spectral plot [37].

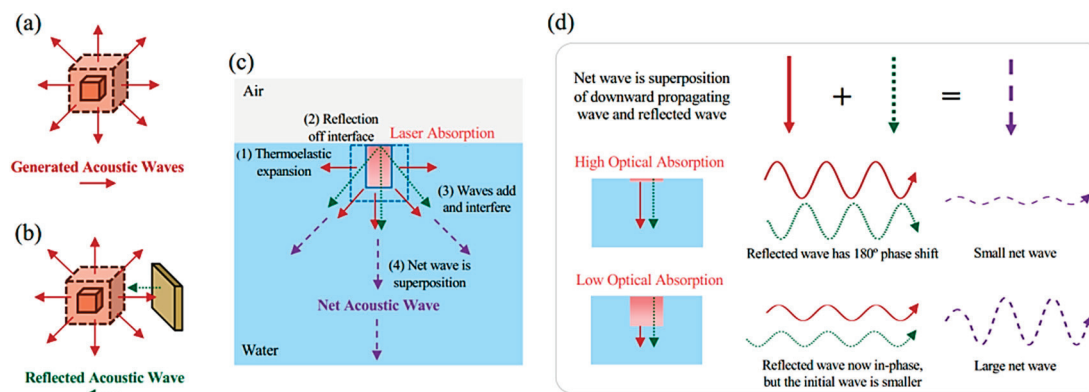
### 3. Influence of Laser Parameters on Acoustic Signal Characteristics

#### 3.1. Selection of Laser Wavelength

In underwater acoustic detection systems, impedance mismatches at medium interfaces can induce complex wave field coupling phenomena. As illustrated in Figure 5, the abrupt mechanical property transition at the air–water interface causes significant reflection of incident acoustic waves. These reflected waves vectorially superimpose with the direct acoustic waves generated by the laser source, and the resulting interference pattern is influenced by the selected laser wavelength.

When a laser with a strongly absorbing wavelength is used, the reflected wave tends to destructively interfere with the original downward-propagating wave, leading to a pronounced cancellation effect. Conversely, for weakly absorbing wavelengths, the reflected wave can remain in phase with the downward acoustic wave, resulting in constructive interference and an enhanced net acoustic pressure field [78,79]. However, weakly absorbing wavelengths exhibit lower optical-to-acoustic energy conversion efficiency, which limits the amplitude of the initial acoustic signal [41].





**Figure 5.** The influence of the air–water interface on the incident sound wave: (a) photoacoustic-generated waves due to thermal volumetric expansion; (b) reflected acoustic wave off a mechanical discontinuity; (c) net acoustic wave which is the coherent sum of the generated and reflected acoustic waves; (d) role of optical absorption in the superposition of the reflected wave with the downward propagating wave [41].

To achieve effective excitation of high-amplitude acoustic waves underwater, it is essential to systematically consider the coupling mechanisms between laser parameters and medium properties. In the process of selecting the laser wavelength, three key factors must be jointly optimized: destructive interference effects, photoacoustic conversion efficiency, and compatibility with commercially available laser sources. The extinction lengths of CO<sub>2</sub> and Nd:YAG lasers in water differ by several orders of magnitude, which directly reflects the fact that water molecules exhibit absorption coefficients in the mid-infrared range that are approximately three orders of magnitude higher than those in the near-infrared range. Based on the aforementioned differences in extinction characteristics, engineering applications must optimize laser parameters according to specific operational scenarios. For near-field acoustic pressure enhancement, strongly absorbing CO<sub>2</sub> lasers are preferred due to their high transient thermal deposition efficiency. In contrast, for long-range acoustic propagation, Nd:YAG lasers are more suitable owing to their lower attenuation in the medium, which helps maintain acoustic intensity over extended distances. This wavelength selection strategy enables effective control over spatial energy deposition during laser–fluid interactions, thereby optimizing the spatiotemporal evolution of the underwater acoustic field.

### 3.2. Influence of Laser Pulse Duration and Single-Pulse Energy

Laser pulse duration and single-pulse energy are decisive factors in determining laser intensity and are also critical to the effectiveness of cross-medium acoustic generation. As shown in Table 5, ns lasers generally exhibit higher energy conversion efficiency than picosecond (ps) pulsed lasers. With increasing incident laser energy, the energy conversion efficiency of ns pulses tends to decrease, whereas that of ps pulses remains relatively stable. In terms of photoacoustic energy conversion performance, ns laser outperforms ps laser [49].

The quantitative mapping model between the acoustic signal characteristics and the laser energy parameters is systematically analyzed. With increasing input laser energy, the energy deposition and photon absorption efficiency are enhanced, leading to a prolonged acoustic pulse duration. Simultaneously, the peak-to-peak overpressure of the direct, reflected, and transmitted signals increases, while the arrival times of the reflected and transmitted signals are shortened. The acoustic pulse energy exhibits a linear increase with rising laser input energy, and the energy conversion efficiencies of both reflected and transmitted waves at the air–water interface also increase approximately linearly. In addition, the peak frequencies of the direct and reflected signals shift toward lower

frequencies, which is attributed to the extended duration of optical radiation and thermal emission effects. The low-frequency components of the signals experience less attenuation in air and can propagate over longer distances. Furthermore, the spectral bandwidths of the direct and reflected signals are reduced, whereas the bandwidth of the transmitted signal remains unchanged [49].

**Table 5.** Under different laser pulses, the incident laser energy, the laser energy absorbed by water, the sound energy generated, and the efficiency of laser energy conversion into sound energy [49].

| ns                  |                     |                     |                                 | ps                  |                     |                     |                                 |
|---------------------|---------------------|---------------------|---------------------------------|---------------------|---------------------|---------------------|---------------------------------|
| Incident Energy /mJ | Absorbed Energy /mJ | Acoustic Energy /uJ | Energy Conversion Efficiency /% | Incident Energy /mJ | Absorbed Energy /mJ | Acoustic Energy /uJ | Energy Conversion Efficiency /% |
| 25                  | 20                  | 128                 | 0.64                            | 10                  | 6                   | 1.7                 | 0.03                            |
| 50                  | 41                  | 140                 | 0.34                            | 15                  | 9.5                 | 1.8                 | 0.02                            |
| 75                  | 64                  | 143                 | 0.22                            | 20                  | 13                  | 3                   | 0.023                           |
| 100                 | 85                  | 148                 | 0.17                            | 25                  | 15                  | 4.6                 | 0.03                            |

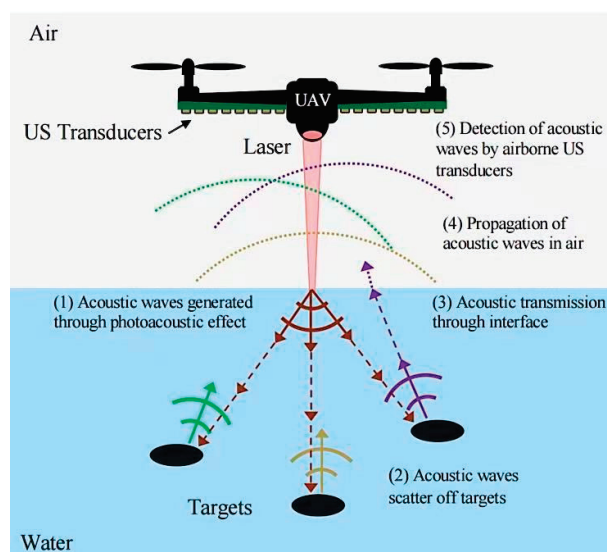
#### 4. LIA Detection Systems

In 1981, the Hickman team developed a modular laser–acoustic underwater detection system [44]. The system consists of four core modules: A Lumonics TEA-103-2 small-aperture CO<sub>2</sub> laser capable of delivering adjustable pulse energies ranging from 1 to 15 J at a wavelength of 10.6 µm; an acoustic receiving array based on a parabolic reflector; a multi-channel signal acquisition and analysis unit; and a mobile deployment platform. Laboratory calibration and field testing confirmed the system’s adaptability across multiple application scenarios. It has been successfully employed in shallow and turbid water depth sounding, underwater target detection, and soil moisture measurement. Since the system is not hardwired to any specific transmitter or sensor, it can also be adapted for navigation and communication purposes.

In 2016, the Bidin team developed a Laser-Induced Acoustic spectroscopy (LIAS) based system for detecting waterborne pollutants [46]. The pulsed excitation module employs a Q-switched Nd:YAG laser with a wavelength of 1064 nm and a pulse energy of 40 mJ to induce laser breakdown in the solution. The optical detection module includes a wavelength-tunable dye laser used as the light source for schlieren imaging. The synchronization control module utilizes a photodiode to monitor the laser pulse in real time, with timing coordination achieved via a delay generator. The high-speed imaging module incorporates a CCD camera to capture the dynamics of acoustic wave propagation. Experimental results revealed a correlation between the acoustic characteristics of LIA in water and the concentration of hydrocarbons. Specifically, the propagation velocity of the acoustic signal was found to be proportional to the hydrocarbon concentration, and the signal amplitude increased linearly with rising hydrocarbon levels. It has been confirmed that LIAS technology offers significant advantages for real-time, in situ monitoring, turbid-water detection, and pollution source localization. These findings provide crucial technical support for the development of next-generation water environment monitoring systems.

In 2020, the Fitzpatrick research team developed an airborne acoustic imaging system based on CMUT [41]. Its innovative architecture is illustrated in Figure 6. A key distinguishing feature of this system lies in its use of air-coupled CMUT, which addresses the issues of acoustic impedance mismatch and transmission loss associated with airborne echoes [80–82]. Additionally, due to the limited bandwidth of CMUT, a sufficient number

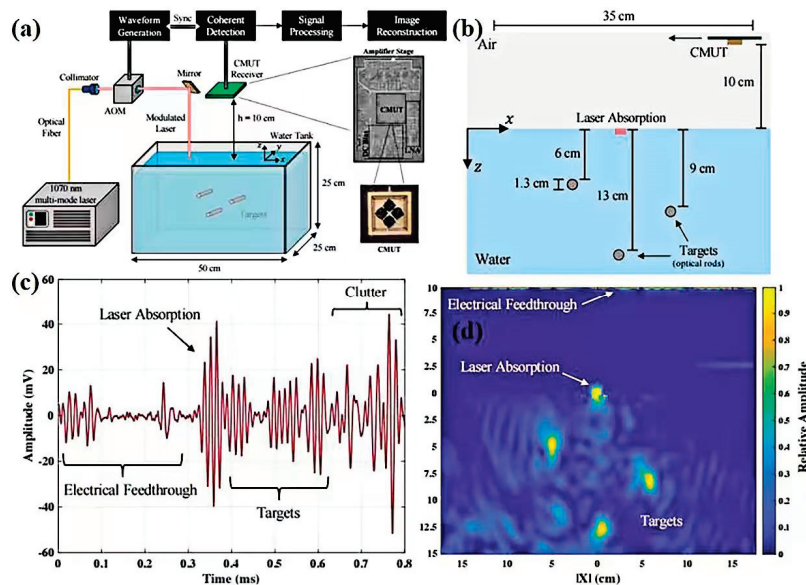
of oscillation cycles at the resonant frequency is required to achieve the high sensitivity that sets them apart from other non-contact sensing devices [83].



**Figure 6.** Fitzpatrick's conceptual experimental system diagram [41].

As illustrated in Figure 7a, the imaging mechanism of this system involves three key physical processes: LIA waves propagate through the underwater medium and are reflected by the target; the reflected acoustic waves cross the air–water interface, with a phase compensation layer enhancing transmission; the CMUT array receives the acoustic signals on the air side, and time–frequency analysis is used to reconstruct the target [84]. Theoretical modeling indicates that conventional 10.6  $\mu\text{m}$  lasers suffer from destructive interference caused by surface reflections, leading to a SPL attenuation of up to 46 dB. To address this, the system employs an optimized multi-peak laser source with a wavelength of 1070 nm, a peak power of 500 W, and a pulse width of 20 ns, resulting in a significant improvement in acoustic energy transmission efficiency. The experimental setup is shown in Figure 7b. Imaging tests were conducted on standard targets placed at water depths of 6 cm, 9 cm, and 13 cm, with a detection height of 10 cm. The corresponding acoustic feature analysis is presented in Figure 7c. While the test height was set to 10 cm, it can be proportionally scaled up in real-world system demonstrations, with operational altitudes reaching several meters. Figure 7d shows the laser–acoustic imaging result obtained using this system. The feasibility of airborne laser–acoustic remote sensing imaging has been demonstrated.

In 2021, the Yellaiah team established an integrated experimental-simulation platform to investigate the acoustic coupling characteristics at the water–air interface [40]. Based on experimental observations, finite element analysis was performed to successfully visualize the propagation and interaction of underwater acoustic pulses at the interface. The analysis revealed that shock waves attenuate in the forward direction during propagation and transform into acoustic pulses, whose signal characteristics can be used to characterize both the shock wave propagation behavior and interface interactions. The experiments employed a Nd:YAG laser with a wavelength of 532 nm, a pulse duration of 10 ns, and an energy range of 40–100 mJ. Underwater acoustic signals generated via LIB were captured using a needle-type hydrophone, while a polarized microphone placed above the water surface was used to record the transmitted and reflected acoustic signals in air. The attenuation and frequency shift in acoustic waves during propagation from water into air were analyzed, providing important insights into acoustic interactions across different media.



**Figure 7.** (a) Schematic diagram of Fitzpatrick proof-of-concept system; (b) experimental design drawings; (c) time domain analysis diagram of acoustic features; (d) laser-acoustic imaging [41].

In 2008, Zong, S. et al. conducted a systematic study on the acoustic radiation mechanisms of laser-induced cavitation dynamics [85]. The experiments employed a Q-switched Nd:YAG laser with a wavelength of 1064 nm, a pulse duration of 8 ns, and an adjustable energy range of 100–800 mJ. Particular emphasis was placed on evaluating the acoustic characteristics under single-pulse energy of 500 mJ and a repetition rate of 5 Hz. The hydrophone was positioned 80 mm from the optical breakdown point, and the expansion and collapse phases of the cavitation bubble were analyzed based on the peak position and amplitude of the received acoustic signals. Experimental results showed that, under the optical breakdown regime, the acoustic signal generated during bubble collapse exhibited higher intensity than that generated during bubble expansion [86,87]. When the hydrophone was placed 1 m from the breakdown point, the acoustic signal reached a SPL of 200.6 dB, meeting the requirements for underwater detection [88]. It was found that the SPL of the underwater acoustic signal increased proportionally with the incident laser energy. However, when the laser energy reached 500 mJ, the SPL exhibited nonlinear behavior.

In 2009, Zhao, A. et al. made a breakthrough by developing a full-condition laser-acoustic detection system [89]. Under non-anechoic water tank conditions, the acoustic waves received by the array elements of the system's base array were concentrated within a conical region with a half-apex angle of approximately 130°. To effectively capture the acoustic signals, the array needed to be positioned at a considerable distance from the laser-water interaction point. Given that the speed of sound in water is significantly higher than that in air, this configuration allows for effective temporal and spatial separation between the LIA echoes and ambient noise, thereby improving signal-to-noise discrimination.

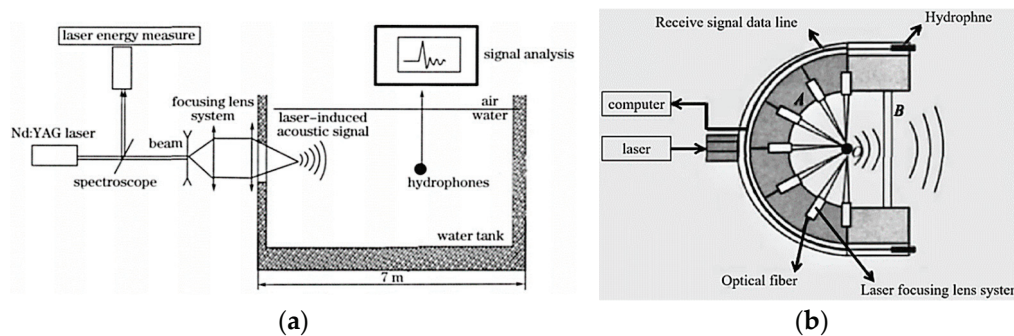
In 2010, Zong, S. et al. designed a laser-acoustic radiation measurement system [90], with its core innovation being the construction of a quasi-free-field testing environment. By installing ATF-1 sulfur-free rubber sound-absorbing materials on the walls, bottom, and surface of the water tank, the setup effectively simulated free-field acoustic propagation conditions similar to those in the ocean. This configuration enabled accurate measurement of both the direct acoustic wave generated by the laser and the echo signals reflected from underwater targets. Experimental results showed that the LIA signal had a peak pressure pulse width of 15  $\mu$ s, with dominant frequencies at 15 kHz, 69 kHz, 166 kHz, and 365 kHz.

The research team derived a far-field acoustic radiation model for laser-induced sound, as shown in Equations (4) and (5). Assuming the cavitation bubble is spherical,  $r$  represents the radius of the plasma cavity,  $t$  is time, and  $p(r, t)$  denotes the sound pressure at position  $r$  and time  $t$ .  $|A|/r$  is the pressure amplitude,  $k = \omega/c$  is the acoustic wave number, and  $u_0$  is the vibration amplitude at the surface of the spherical source. It can be seen that the sound pressure amplitude depends not only on the vibration amplitude of the spherical source but also on the radius of the cavitation bubble.

$$p(r, t) = \frac{|A|}{r} \exp[j(\omega t - kr + \theta)] \quad (4)$$

$$|A| = \left| \frac{\rho_0 c_0 k r_0 u_0}{1 + (k r_0)^2} \right| \quad (5)$$

In 2012, Liu, T. et al. developed a laser–acoustic directional detection system based on wavefront reconstruction [91], with its innovative architecture illustrated in Figure 8b. Experimental results demonstrated that a strong pulsed laser focused into a liquid medium can generate an explosive spherical acoustic source, and the detector is capable of converting the spherical acoustic wave into a highly directional plane wave signal. In the system, wall A guides the laser pulse through a beam-expansion and focusing module, focusing it onto point O. Simultaneously, multiple laser pulses delivered via optical fibers are also focused at point O, where they induce acoustic waves through water surface breakdown. These acoustic waves are reflected by the acoustic reflector wall A and emitted into the water through the acoustic vibration surface B. Upon encountering a target, the reflected acoustic echoes are received by a hydrophone and processed by a signal processing unit to complete the underwater detection task.



**Figure 8.** (a) The experimental system used by Zong, S. et al. in 2010 [90]; (b) Structure diagram of laser acoustic underwater target detector [92].

In 2015, Ye, J. et al. proposed a method for underwater target estimation using LIAs. By measuring the echo signal attenuation of different materials and calculating the frequency spacing between extrema in the reflected comb-like spectrum, the target size can be estimated [92]. Under the boundary conditions of continuity of normal particle velocity and acoustic pressure at the two interfaces in front of and behind the target, the reflection coefficient  $R$  for underwater acoustic–solid coupling is given by

$$R = \frac{Z_2(Z_3 - Z_1) \cos(k_2 L \cos \theta) + j(Z_2^2 - Z_1 Z_3) \sin(k_2 L \cos \theta)}{Z_2(Z_3 + Z_1) \cos(k_2 L \cos \theta) + j(Z_2^2 + Z_1 Z_3) \sin(k_2 L \cos \theta)} \quad (6)$$

$Z_1$  is the acoustic impedance of the water medium, and  $Z_1 = Z_3$  is the acoustic impedance of the underwater target material,  $L$  denotes the thickness of the reflecting

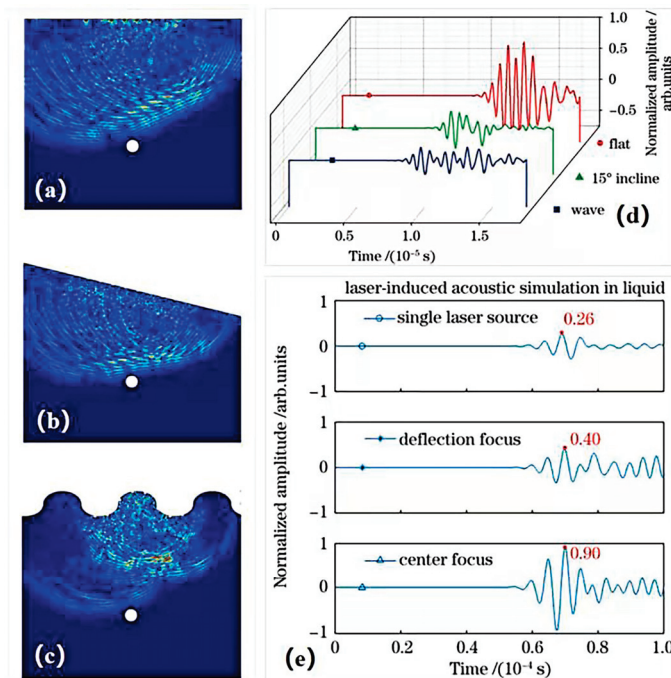


medium,  $\theta$  is the incident angle of the acoustic signal, and  $k_2 = f/c$  is the acoustic wave number in the target material. It can be seen that the reflection coefficient  $R$  is a function of frequency. By taking the logarithm of the reflection coefficient and analyzing the frequency spacing between adjacent extrema in the comb-like spectrum of  $\log(R)$ , the thickness  $d$  of the reflecting medium can be estimated.

$$d = \frac{2\pi c}{\Delta f} \quad (7)$$

Here,  $c$  represents the speed of sound in water for the LIA signal, and  $\Delta f$  is the frequency spacing between adjacent extrema in the comb-like spectrum. Using this method, the size of the underwater target can be effectively estimated [92].

In 2022, Zhao, Y. et al. applied the LIA Synthetic Aperture Focusing Technique (SAFT) to enhance target detection. By constructing a laser source array to concentrate optical energy, the system achieved focused and intensified acoustic energy delivery to underwater targets, thereby improving lateral resolution and overall imaging accuracy [93]. Finite element simulations were conducted to analyze the impact of surface water disturbances on LIA signals, as illustrated in Figure 9a–d. The results show that surface perturbations significantly affect the signal: greater disturbance amplitudes lead to lower acoustic signal amplitudes and reduced signal-to-noise ratios (SNR). Figure 9e compares acoustic signals under different laser focusing configurations. The detection effectiveness of a central-focusing eight-source laser array was approximately 3.46 times greater than that of a single-source array, demonstrating that integrating LIAs with SAFT is beneficial for improving underwater target detection performance.



**Figure 9.** Finite element numerical simulation results: (a) the distribution of sound pressure on the ocean surface; (b) sound pressure distribution on a 15° inclined plane; (c) wave surface sound pressure distribution; (d) comparison of sound pressure signals at the same position under different water surfaces; (e) comparison of Time-Domain Signal of Acoustic Pressure after SAFT Scanning with Single Laser and Multi-Source Lasers [93].

## 5. Discussion

As an emerging approach for underwater detection, LIA technology still encounters several critical technical challenges in practical applications. The primary difficulties involve achieving efficient conversion of laser energy into acoustic energy, maintaining long-range propagation of acoustic waves in underwater environments, and enhancing the anti-interference capability of optical focusing systems under real sea-surface wave conditions [94,95].

Currently, the energy conversion efficiency of LIA technology is constrained by seawater absorption and scattering, as well as by laser parameters such as wavelength and pulse width. Future efforts should focus on optimizing laser wavelength selection, pulse energy, and focusing strategies to enhance laser-to-acoustic conversion efficiency. Report the conversion efficiency from energy to SPL and SNR under specified turbidity and dissolved-gas conditions, including laser fluence, focal spot, laser-to-acoustic efficiency bandwidth, and the stability of peak sound pressure. Explore pulse-train shaping to raise thermoelastic yield without damaging breakdown. Regarding underwater acoustic propagation, viscous absorption, molecular relaxation, and scattering significantly degrade signal strength and fidelity [96]. To extend detection range and improve resolution, advanced compensation algorithms for acoustic attenuation are required. Moreover, multi-frequency signal fusion techniques hold considerable potential for suppressing ambient noise and improving the SNR. Given the inherently weak acoustic signals generated by LIA, high-sensitivity acoustic sensors are indispensable [97]. Among them, fiber-optic hydrophones and piezoelectric composite sensors have emerged as key research directions owing to their high sensitivity and strong anti-interference capabilities [98]. Future development should aim to broaden their frequency response and expand their dynamic range to better adapt to complex underwater environments. Additionally, integrating LIA systems with optical imaging and magnetic sensing promises to enable the development of multimodal detection platforms. Leveraging deep learning algorithms to fuse and interpret multi-source data can further enhance target recognition accuracy and improve environmental perception. Finally, breakthroughs in real-time data processing and transmission will be critical for achieving the operational viability of LIA-based underwater sensing systems.

## 6. Conclusions

With the miniaturization of lasers, acoustic sensors, and signal processing modules, LIA systems are expected to evolve toward portable and highly integrated configurations, enabling deployment on autonomous underwater vehicles and unmanned platforms. By optimizing laser parameters and acoustic signal processing algorithms, future LIA technologies are poised to achieve higher spatial resolution and longer-range underwater target detection, providing strong support for ocean resource exploration and military applications. To address the challenges posed by turbid waters, high-pressure deep-sea environments, and other complex marine conditions, the development of real-time environmental calibration techniques and highly stable LIA systems with enhanced adaptability becomes a critical priority. Coupled with interface-aware focusing, uncertainty-calibrated multimodal fusion, and rigorous safety/compliance reporting, these measurable milestones provide a practical path to transition LIA from promise to practice in complex marine environments. In summary, LIA technology demonstrates broad application prospects in underwater sensing. With continuous breakthroughs in key enabling technologies, LIA systems are expected to play a pivotal role in defense, marine environmental monitoring, and disaster early-warning systems.

**Author Contributions:** Conceptualization, J.Z. and M.W.; methodology, Y.Y. and K.Y.; validation, D.X., J.Y. and S.X.; formal analysis, J.Z. and S.X.; investigation, J.Z.; resources, D.X., J.Y. and M.W.; writing—original draft preparation, J.Z. and M.W.; writing—review and editing, Y.Y. and K.Y.; project administration, X.W.; funding acquisition, J.Y. and D.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (Nos.12174211, 11874232, 12174212) and the Laoshan Laboratory Science and Technology Innovation Project, grant number No. LSKJ202200801. We give our thanks for the support of the Qingchuang team.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Aussel, J.D.; Le Brun, A.; Baboux, J.C. Generating acoustic waves by laser: Theoretical and experimental study of the emission source. *Ultrasonics* **1988**, *26*, 245–255. [CrossRef]
2. Kaushal, H.; Kaddoum, G. Underwater Optical Wireless Communication. *IEEE Access* **2016**, *4*, 1518–1547. [CrossRef]
3. Zemlyanov, A.A.; Alexander, A.; Bochkarev, N.N.; Nikolay, N.; Kabanov, A.M.; Andrey, M.; Pogodaev, V.A.; Vitalii, A. Generation of acoustic pulses on the natural centers of absorption with propagation CO<sub>2</sub> laser radiation on atmospheric paths. *Proc. SPIE Int. Soc. Opt. Eng.* **2001**, *4341*, 202–209.
4. Anisimov, S.I. Vaporization Of Metal Absorbing Laser Radiation. In *30 Years of the Landau Institute—Selected Papers*; World Scientific Series in 20th Century Physics; World Scientific: Singapore, 1996; pp. 14–15.
5. Antonelli, L.; Blackmon, F. Experimental demonstration of remote, passive acousto-optic sensing. *J. Acoust. Soc. Am.* **2004**, *116*, 3393–3403. [CrossRef]
6. Ma, Y.L.; Yang, Y.X.; He, Z.Y.; Yang, K.D.; Sun, C.; Wang, Y.M. Theoretical and practical solutions for high-order superdirectivity of circular sensor arrays. *IEEE Trans. Ind. Electron.* **2013**, *60*, 203–209. [CrossRef]
7. Yang, Y.X.; Zhou, J.B.; L, M.Y.; Hou, X.H.; Yang, L.; Xie, L.; Wang, L.; Shen, H.; Liu, X.H.; Lei, B.; et al. Underwater acoustics: Current status and future trends. *Tech. Acoust.* **2025**, *44*, 309–332.
8. Du, Z.P.; Chen, G.; Wang, D. Foreign sonar technology development research summary. *Ship Sci. Technol.* **2019**, *41*, 145–151.
9. Zhang, G.; Zhang, L.; Wang, Y.; Ma, Y.; Zhou, X.; Yu, Y. Spatiotemporal Analysis of Sonar Detection Range in Luzon Strait. *J. Mar. Sci. Eng.* **2024**, *12*, 1191. [CrossRef]
10. Jia, Y.J.; Wang, P.B.; Wei, H.K.; Chen, Q.; Hu, Y.H. Research on high-resolution processing of frequency modulation combined signal for active sonar detection. *Front. Phys.* **2025**, *13*, 1542037. [CrossRef]
11. Wang, X.; Zhang, Z.; Shang, X. Research on Improved YOLO11 for Detecting Small Targets in Sonar Images Based on Data Enhancement. *Appl. Sci.* **2025**, *15*, 6919. [CrossRef]
12. Xie, L.; Sun, C.; Kong, D.Z. Matched-phase weighting beamformer to improve the gain of a long linear array in the range-dependent ocean waveguide. *IEEE Access* **2020**, *8*, 177437–177446. [CrossRef]
13. Lu, J.Y.; Yang, Y.X.; Yang, L. An efficient off-grid direction-of-arrival estimation method based on inversefree sparse Bayesian learning. *Appl. Acoust.* **2023**, *211*, 109521. [CrossRef]
14. Fillingier, L.; Hunter, A.J.; Zampolli, M.; Clarijs, M.C. Passive Acoustic Detection of Closed-Circuit Underwater Breathing Apparatus in an Operational Port Environment. *J. Acoust. Soc. Am.* **2012**, *132*, EL310–EL316. [CrossRef]
15. Sun, Y.C.; Chen, W.Y.; Wang, P.B. Review of research on the characteristics of underwater acoustic signal of frogman. *Tech. Acoust.* **2022**, *41*, 479–488.
16. Dong, L.; Li, N.; Xie, X.; Bao, C.; Li, X.; Li, D. A Fast Analysis Method for Blue-Green Laser Transmission through the Sea Surface. *Sensors* **2020**, *20*, 1758. [CrossRef] [PubMed]
17. Collin, A.; Archambault, P.; Long, B. Mapping the Shallow Water Seabed Habitat with the SHOALS. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2947–2955. [CrossRef]
18. Tuldahl, H.M.; Philipson, P.; Kautsky, H.; Wikstrom, S.A. Sea Floor Classification with Satellite Data and Airborne Lidar Bathymetry. In *Ocean Sensing and Monitoring V, Proceeding of the SPIE, the International Society for Optics and Photonics, Baltimore*,

- MD, USA, 30 April–1 May 2013; Hou, W.W., Arnone, R.A., Eds.; SPIE-Int Soc Optical Engineering: Bellingham, WA, USA, 2013; Volume 8724, p. UNSP 87240B, ISBN 978-0-8194-9515-0.
19. Li, X.; Miao, X.; Qi, X.; Han, X. Research on the propagation characteristics of blue-green laser through sea surface with foams. *Optik* **2018**, *170*, 265–271. [CrossRef]
20. Birkebak, M.; Eren, F.; Pe'eri, S.; Weston, N. The Effect of Surface Waves on Airborne Lidar Bathymetry (ALB) Measurement Uncertainties. *Remote Sens.* **2018**, *10*, 453. [CrossRef]
21. Ma, J.; Lu, T.; He, Y.; Jiang, Z.; Hou, C.; Li, K.; Liu, F.; Zhu, X.; Chen, W. Compact Dual-Wavelength Blue-Green Laser for Airborne Ocean Detection Lidar. *Appl. Opt.* **2020**, *59*, C87–C91. [CrossRef]
22. Hostetler, C.A.; Behrenfeld, M.J.; Hu, Y.; Hair, J.W.; Schullien, J.A. Spaceborne Lidar in the Study of Marine Systems. *Annu. Rev. Mar. Sci.* **2018**, *10*, 121–147. [CrossRef]
23. Wu, Z.; Li, H.; Yang, F.; Li, S.; Gao, J.; Zhang, W.; Jin, S.; Huo, G.; Shang, J.; Yang, K.; et al. *High-Resolution Seafloor Topography and Morphology—Detection, Processing Theory and Techniques*; Science Press: Beijing, China, 2017; ISBN 978-7-03-052904-6.
24. Ng, R.; Sachat, A.; Cespedes, F.; Poblet, M.; Madiot, G.; Jaramillo-Fernandez, J.; Xiao, P.; Florez, O.; Sledzinska, M.; Sotomayor-Torres, C.; et al. Excitation and Detection of Acoustic Phonons in Nanoscale Systems. *Nanoscale* **2022**, *14*, 13428–13451. [CrossRef]
25. Ikei, A.; Sampath, K. Remote Operation of a Single-Point LDV System to Acquire 2D Measurements. *arXiv* **2021**, arXiv:2102.13031. [CrossRef]
26. Jin, Y.; Dollevoet, R.; Li, Z. Removing Speckle Noise from the Signals of a Laser Doppler Vibrometer on Moving Platforms (LDVom) by Ensemble Empirical Mode Decomposition. *Meas. Sci. Technol.* **2022**, *33*, 125205. [CrossRef]
27. Yuan, K.; Zhu, W.D. A Novel General-Purpose Three-Dimensional Continuously Scanning Laser Doppler Vibrometer System for Full-Field Vibration Measurement of a Structure with a Curved Surface. *J. Sound Vib.* **2022**, *540*, 117274. [CrossRef]
28. Tsubouchi, K.; Yoshida, S.; Satoh, T.; Ikeuchi, K. A Fiber LDV System for Measurement of Atomized Droplets Away from the Trailing Edge of a Fixed Blade in Two-Phase Flow. *JSME Int. J. Ser. 2 Fluids Eng. Heat Transf. Power Combust. Thermophys. Prop.* **1990**, *33*, 511–516. [CrossRef]
29. Chen, H. Research on Signal Processing Methods of Laser Doppler Velocimetry System. Master's Thesis, Qingdao University of Science and Technology, Qingdao, China, 2023.
30. Chen, L.J.; Xi, C.B.; Zhou, J.; Nie, X.M.; Jin, S.L.; Luo, H. Research and Flight Test on Airborne Laser Doppler Velocimeter for Unmanned Aerial Vehicles. *Chin. J. Opt.* **2023**, *43*, 65–74.
31. Guardato, S.; Riccio, R.; Janneh, M.; Bruno, F.A.; Pisco, M.; Cusano, A.; Iannaccone, G. An Innovative Fiber Optic Hydrophone for Seismology: Testing Detection Capacity for Very Low-Energy Earthquakes. *Sensors* **2023**, *23*, 3374. [CrossRef]
32. Li, M.; Yu, Y.; Lu, Y.; Hu, X.; Wang, Y.; Qin, S.; Lu, J.; Yang, J.; Zhang, Z. Optical Microfiber All-Optical Phase Modulator for Fiber Optic Hydrophone. *Nanomaterials* **2021**, *11*, 2215. [CrossRef]
33. Zhang, Y.; Meng, Z.; Wang, J.; Chen, M.; Liang, Y.; Hu, X. The Two-Stage Suspension System of the Fiber Optic Vector Hydrophone for Isolating the Vibration from the Mooring Rope. *Sensors* **2022**, *22*, 9261. [CrossRef]
34. Yu, Z.; Cai, Y.; Mo, D. Comparative Study on Noise Reduction Effect of Fiber Optic Hydrophone Based on LMS and NLMS Algorithm. *Sensors* **2020**, *20*, 301. [CrossRef]
35. Sun, X.Y.; Guang, D.; Deng, C.Y.; Yu, B.L. Design and test of fiber optic hydrophone element for deep-sea application. *Opt. Precis. Eng.* **2025**, *33*, 1550–1557.
36. Hu, Y.; Yan, Y.; Wang, L.; Qian, X. Non-Contact Vibration Monitoring of Power Transmission Belts Through Electrostatic Sensing. *IEEE Sens. J.* **2016**, *16*, 3541–3550. [CrossRef]
37. Zhao, J.; Wang, M.R.; Xu, D.G.; Li, J.X.; Wang, X.; Yang, Y.G. Underwater Transmission of Pulsed Laser-acoustic Signals and Acoustic Scattering Characteristics of Metallic Targets. *Adv. Lasers Optoelectron.* **2025**, *62*, 06.
38. Du, Y.C.; Cao, X.L.; Yang, Y.G.; Zhang, T.; Yuan, J.Q.; Cui, T.Y.; Yao, J.Q. Acoustic Signal Reconstruction Across Water–Air Interface Through Millimeter-Wave Radar Micro-Vibration Detection. *J. Mar. Sci. Eng.* **2024**, *12*, 1989. [CrossRef]
39. Zhang, T.C.; Cao, X.L.; Du, Y.C.; Yao, J.Q. Study on time-frequency domain characteristics and cross-media communication of laser acoustic signals based on optical breakdown mechanism. *Phys. Scr.* **2025**, *100*, 035402. [CrossRef]
40. Li, P.; Zhao, Y.; Zhou, Z.Q.; Zhang, P.H.; Bai, X.; Ma, J. Research on laser induced acoustic detection of trans-media aerial-underwaterInfrared and Laser Engineering. *Infrared Laser Eng.* **2021**, *50*, 20200310.
41. Fitzpatrick, A.; Singhvi, A.; Arbabian, A. An Airborne Sonar System for Underwater Remote Sensing and Imaging. *IEEE Access* **2020**, *8*, 189945–189959. [CrossRef]
42. Egerev, S.V. In search of a noncontact underwater acoustic source. *Acoust. Phys.* **2003**, *49*, 51–61. [CrossRef]
43. Hickman, G.D.; Edmonds, J.A. An Experimental Facility for Laser/Acoustic Applications. *NASA STI/Recon Tech. Rep. N.* **1981**, *82*, 26657.



44. Hickman, G.D.; Edmonds, J.A. Laser-acoustic measurements for remotely determining bathymetry in shallow turbid waters. *J. Acoust. Soc. Am.* **1983**, *73*, 840–843. [CrossRef]
45. Blackmon, F.A.; Antonelli, L.T. Experimental detection and reception performance for uplink underwater acoustic communication using a remote, in-air, acousto-optic sensor. *IEEE J. Ocean. Eng.* **2006**, *31*, 179–187. [CrossRef]
46. Bidin, N.; Hossenian, R.; Duralim, M.; Krishnan, G.; Marsin, F.; Nughro, W.; Zainal, J. Determination of hydrocarbon levels in water via LIAs wave. *Opt. Lasers Eng.* **2016**, *79*, 61–67. [CrossRef]
47. Jukna, V.; Jarnac, A.; Milián, C.; Brelet, Y.; Carbonnel, J.; André, Y.-B.; Guillermin, R.; Sessarego, J.-P.; Fattaccioli, D.; Mysyrowicz, A.; et al. Underwater acoustic wave generation by filamentation of terawatt ultrashort laser pulses. *Phys. Rev.* **2016**, *93*, 063106. [CrossRef]
48. Baimler, I.V.; Lisitsyn, A.B.; Gudkov, S.V. Water Decomposition Occurring During Laser Breakdown of Aqueous Solutions Containing Individual Gold, Zirconium, Molybdenum, Iron or Nickel Nanoparticles. *Front. Phys.* **2020**, *8*, 620938. [CrossRef]
49. Yellaiah, J.; Kiran, P.P. Input pulse duration effect on laser-induced underwater acoustic signals. *Appl. Opt.* **2021**, *60*, 4582–4590. [CrossRef]
50. Cui, G.; Li, R.; Tian, Z.; Sang, G. A Laser Remote Sensing Study for Detection of Under Water Acoustic Signals. *Ship Sci. Technol.* **2002**, *24*, 46–50.
51. Cui, G.; Li, R.; Li, J. Research on LIA Technology and Its Application Prospects. *Acoust. Electron. Eng.* **1989**, *2*, 17–22.
52. Li, R.; Cui, G.; Sang, G. Characteristics of Sound Impulse Induced by CO<sub>2</sub> Laser in Water and its Transmission Lose. *Ship Sci. Technol.* **2002**, *24*, 41–45.
53. Li, R.; Cui, G.; Yun, T.; Sang, G. Laser/acoustic remote sensing for underwater objects. *Prog. Nat. Sci.* **1996**, *6*, S755–S759.
54. Li, R.; Cui, H. Application of Laser/Acoustic Technology in Anti-Stealth Submarine Detection. *Ship Sci. Technol.* **1995**, *2*, 11–14.
55. Li, Q. Fundamental Research on Novel All-Fiber Photoacoustic Source for Underwater Communication. Ph.D. Thesis, National University of Defense Technology, Changsha, China, 2010.
56. Wang, W.; Yang, Y.; Zhang, W. Pulsed-Laser-Sound-Signal Detection Based on Frequency-Domain Energy Dtector. *Chin. J. Lasers* **2019**, *46*, 53–61.
57. Liu, Z. Research on Frequency Band Efficient Utilization of Underwateracoustic Communication Based on Chaos. Master's Thesis, North University of China, Taiyuan, China, 2023.
58. Wang, Y.; Wang, J.; Wu, R. Influences of liquids properties and absorbed energy on laser induced expanding sound. *High Power Laser Part Beams* **2009**, *21*, 998–1002.
59. Wang, Y.; Wang, J.; Zong, S. Coincidence relation between laser pulse energy and hydroacoustic speciality. *Laser Infrared* **2008**, *38*, 981–984.
60. Zong, S.; Wang, J.; Wang, Y. Experiment Investigation of Water Breakdown Characteristic by High Power Laser. *Opto-Electron. Eng.* **2008**, *35*, 68–72.
61. Li, S.; Wu, R.; Wang, X.; Wang, J.; Zong, S. Properties of LIA signals transmission in liquid. *Infrared Laser Eng.* **2017**, *46*, 54.
62. Sigrist, M.W. Laser generation of acoustic waves in liquids and gases. *J. Appl. Phys.* **1986**, *60*, R83–R122. [CrossRef]
63. Zong, S.; Wang, J.; Wang, Y.; Miao, H. Laser-acoustic source characteristic and its application in ocean. *Ying Yong Guang Xue* **2008**, *29*, 408–411.
64. Li, P. Research on the Method and Key Technology of Trans-media Detection Based on Laser Induced Acoustic Mechanism. Master's Thesis, Qilu University of Technology, Shandong, China, 2021.
65. Sigrist, M.W.; Kneubühl, F.K. Laser-generated stress waves in liquids. *J. Acoust. Soc. Am.* **1978**, *64*, 1652–1663. [CrossRef]
66. Farrant, D.; Burke, J.; Dickinson, L.; Fairman, P.; Wendoloski, J. Opto-acoustic underwater remote sensing (OAURS)—An optical sonar? In Proceedings of the OCEANS 2010 IEEE, Sydney, Australia, 24–27 May 2010; pp. 1–7.
67. Kiselev, V.D.; Maksimov, A.O. Dispersion deformation of optoacoustic pulses in the surface layer of the ocean. *J. Appl. Mech. Tech. Phys.* **1993**, *34*, 796–801. [CrossRef]
68. Teslenko, V.S. Electrical, laser and acoustic breakdown of a liquid, resemblance and distinctions. In Proceedings of the 2002 IEEE 14th International Conference on Dielectric Liquids, ICDL 2002, Graz, Austria, 7–12 July 2002; pp. 163–166.
69. Li, R.; Cui, G.; Tian, Z. *Laser-Acoustic Remote Sensing Technology*; National Defense Industry Press: Beijing, China, 2003.
70. He, Q.; Zong, S. Analysis and thinking about the development of air and water crossed laser acoustic technology. *Laser Infrared* **2019**, *49*, 3–8.
71. Chen, X. Studies on Mechanisms of the Interaction Between High-Power Laser and Matter in Water. Ph.D. Thesis, Nanjing University of Science and Technology, Nanjing, China, 2004.
72. Zhao, R. Studies on Characteristic of Laser-induced Plasma Shock Wave Evolution and Cavitation Bubble Dynamics. Ph.D. Thesis, Nanjing University of Science and Technology, Nanjing, China, 2007.



73. Zong, S.; Wang, J.; Ma, Z.; Wang, Y. Erosion characteristics of laser bubble collapse in the vicinity of a solid boundary. *Chin. J. Opt.* **2010**, *30*, 885–892.
74. Li, S.; Liu, X.; Wang, J.; Zong, S.; Shen, Z.; Ni, X. Influence of bubble content on spectrum properties of laser-induced cavitation bubble collapse sound waves in liquid. *Qiang Ji Guang Yu Li Zi Shu* **2012**, *24*, 1067–1070.
75. Li, S.Y.; Liu, T.; Wang, J.A.; Zong, S.G. Properties of laser-induced two-bubble sound waves. *Guang Dian Zi Ji Guang* **2012**, *23*, 2440–2445.
76. Li, S.; Wang, X.; Wang, J.; Zong, S.; Liu, T. Experimental investigation of influence of ambient pressure on properties of laser-induced cavitation bubble collapse sound waves. *Infrared Laser Eng.* **2015**, *44*, 879–883.
77. Thorp, W.H. Analytic Description of the Low-Frequency Attenuation Coefficient. *J. Acoust. Soc. Am.* **1967**, *42*, 270. [CrossRef]
78. Wang, L.V. *Photoacoustic Imaging and Spectroscopy*; CRC Press eBooks; CRC Press: Boca Raton, FL, USA, 2017; 536p. [CrossRef]
79. Hoelen, C.G.A.; de Mul, F.F.M. A new theoretical approach to photoacoustic signal generation. *J. Acoust. Soc. Am.* **1999**, *106*, 695–706. [CrossRef]
80. Ma, B.; Firouzi, K.; Brenner, K.; Khuri-Yakub, B.T. Wide Bandwidth and Low Driving Voltage Vented CMUTs for Airborne Applications. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2019**, *66*, 1777–1785. [CrossRef]
81. Park, K.; Khuri-Yakub, B.T. 3-D airborne ultrasound synthetic aperture imaging based on CMUT. *Ultrasonics* **2013**, *53*, 1355–1362. [CrossRef]
82. Nan, H.; Boyle, K.C.; Apte, N.; Aliroth, M.S.; Bhuyan, A.; Nikoozadeh, A.; Khuri-Yakub, B.T.; Arbabian, A. Non-contact thermoacoustic detection of embedded targets using airborne-CMUT. *Appl. Phys. Lett.* **2015**, *106*, 084101. [CrossRef]
83. Boyle, K.C.; Nan, H.; Khuri-Yakub, B.T.; Arbabian, A. Noncontact thermoacoustic detection of targets embedded in dispersive media. In Proceedings of the SPIE Security + Defence, Edinburgh, UK, 26–29 September 2016; p. 99880H.
84. Kirk, J.C. A Discussion of Digital Processing in Synthetic Aperture Radar. *IEEE Trans. Aerosp. Electron. Syst.* **1975**, *AES-11*, 326–337. [CrossRef]
85. Blackmon, F.; Antonelli, L.; Carapezza, E. Remote, aerial, opto-acoustic communications and sonar. *Proc. SPIE Int. Soc. Opt. Eng.* **2005**, *5778*, 800–808.
86. He, D.M.; Scheps, R. High-power Nd: YAG-generated underwater sound source for air-submarine communication. *Proc. SPIE Int. Soc. Opt. Eng.* **1999**, *3613*, 83–92.
87. Blackmon, F.; Antonelli, L. Experimental demonstration of multiple pulse nonlinear optoacoustic signal generation and control. *Appl. Opt.* **2005**, *44*, 103–112. [CrossRef]
88. Zong, S.; Wang, J. Both wavelength laser fusion for detecting underwater target. *Opt. Commun. Technol.* **2008**, *32*, 53–56.
89. Zhao, A.; Dong, H.; Sun, G.; Hui, J.; Cai, P. A laser-acoustic system for underwater target detection. *Sheng Xue Ji Shu* **2009**, *28*, 14–16.
90. Zong, S.; Wang, J.; Ma, M.; Ma, Z. Investigation of LIAs signal for target detection. *Chin. J. Lasers* **2010**, *37*, 1332–1337. [CrossRef]
91. Liu, T.; Wang, J.; Zong, S.; Liang, S. Design of underwater target detector by laser-induced sound signal. *Infrared Laser Eng.* **2012**, *41*, 1767–1771.
92. Ye, J. Underwater Target Estimation Based on Wideband LIA Source. Master’s Thesis, Zhejiang University, Hangzhou, China, 2015.
93. Zhao, Y.; Li, P.; Zhang, P.; Zhou, Z.; Bai, X.; Ma, J. Underwater Small Target Detection Method of LIA SAFT Based on PSM Algorithm. *Chin. J. Lasers* **2022**, *49*, 0909001.
94. Spagnolo, G.S.; Cozzella, L.; Leccese, F. Underwater Optical Wireless Communications: Overview. *Sensors* **2020**, *20*, 2261. [CrossRef]
95. Mahmud, M.; Islam, S.; Ahmed, A.; Younis, M.; Choa, F.S. Cross-Medium Photoacoustic Communications: Challenges, and State of the Art. *Sensors* **2022**, *22*, 4224. [CrossRef]
96. Chen, Q.M. Research of Acoustics Induced by Pulsed CO<sub>2</sub> Laser in Water. Ph.D. Thesis, Huazhong University of Science and Technology, Wuhan, China, 2009.
97. Hu, X.; Chen, M.; Zhang, X.; Rao, J.; Li, S.; Song, X. MFCA-Transformer: Modulation Signal Recognition Based on Multidimensional Feature Fusion. *Sensors* **2025**, *25*, 5061. [CrossRef] [PubMed]
98. Cao, Y.; Hu, R.; Shi, W.; Zhou, R. Experimental Study on Near-Wall Laser-Induced Cavitation Bubble Micro-Dimple Formation on 7050 Aluminum Alloy. *Water* **2024**, *16*, 1410. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## Article

# Spatiotemporal Patterns of Climate-Vegetation Regulation of Soil Moisture with Phenological Feedback Effects Using Satellite Data

Hanmin Yin <sup>1,2,3,†</sup>, Xiaohan Liao <sup>1,3,4,†</sup>, Huping Ye <sup>1,3,4</sup>, Jie Bai <sup>5,6</sup>, Wentao Yu <sup>7</sup>, Yue Li <sup>8</sup>, Junbo Wei <sup>2</sup>, Jincheng Yuan <sup>2</sup> and Qiang Liu <sup>2,8,\*</sup>

<sup>1</sup> State Key Laboratory of Resources and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; yinhanmin18@mailsucas.ac.cn (H.Y.); liaoxh@igsrr.ac.cn (X.L.); yehp@igsrr.ac.cn (H.Y.)

<sup>2</sup> Pengcheng Laboratory, Shenzhen 518000, China; weijb@pcl.ac.cn (J.W.); yuanjch@pcl.ac.cn (J.Y.)

<sup>3</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>4</sup> Key Laboratory of Low Altitude Geographic Information and Air Route, Civil Aviation Administration of China, Beijing 100101, China

<sup>5</sup> State Key Laboratory of Ecological Safety and Sustainable Development in Arid Lands, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China; baijie@ms.xjb.ac.cn

<sup>6</sup> Key Laboratory of RS & GIS Application Xinjiang, Urumqi 830011, China

<sup>7</sup> Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; wt.yu@siat.ac.cn

<sup>8</sup> Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; liyue97@mail.bnu.edu.cn

\* Correspondence: liuq03@pcl.ac.cn

† These authors contributed equally to this work.

## Highlights

### What are the main findings?

- A significant global shift occurred around the year 2000: The rate of vegetation greening accelerated significantly after this point, coinciding with a reversal in soil moisture trends from general drying (1982–1999) to wetting (2000–2020). Notably, this post-2000 wetting trend was most pronounced in temperate and tropical climate zones compared to cold and arid regions.
- The drivers of soil moisture are complex, and vegetation's regulatory role is growing: The primary factors controlling soil moisture vary by season, climate zone, and soil depth. For instance, for rootzone soil moisture, vegetation (LAI) is dominant in winter and spring, while solar radiation becomes the primary driver in summer and autumn. Crucially, the overall influence of vegetation (LAI) on soil moisture has significantly strengthened over time, generally contributing to a wetting trend. Although vegetation is not the absolute dominant factor in all seasons, the growth of its influence is particularly pronounced in specific periods: for rootzone soil moisture, its influence increased the most during winter and summer, while for surface soil moisture, the increase was most significant in winter. This highlights the increasingly important regulatory role of vegetation in the global water cycle.

### What are the implications of the main findings?

- Improved prediction models are needed for water resource management: The finding that vegetation's role in regulating soil moisture is growing stronger means that traditional climate-only models may be insufficient. These results provide a scientific basis for improving global models by better incorporating the dynamic effects of vegetation greening, leading to more accurate predictions of soil moisture and better management of water resource risks under climate change.

- A deeper understanding of Earth's ecohydrological processes: This research clarifies the complex, intertwined relationship between the climate, plants (biosphere), and water (hydrosphere). Understanding that factors like the timing of peak vegetation growth (POS) can directly influence soil moisture provides a more nuanced view of terrestrial ecosystems, which is fundamental for assessing ecological responses to global environmental changes.

**Abstract:** Global soil moisture has undergone significant changes in recent decades due to climate change and vegetation greening. However, the seasonal and climate zonal variations in soil moisture dynamics at different depths, driven by both climate and vegetation, remain insufficiently explored. This study provides a comprehensive analysis of the global patterns in rootzone and surface soil moisture and leaf area index (LAI) across different seasons and climate zones, utilizing satellite observations from 1982 to 2020. We investigate how climatic factors and LAI influence soil moisture variations and quantify their dominant contributions. Furthermore, by employing key vegetation phenological indicators, namely the peak of growing season (POS) and the corresponding maximum LAI ( $LAI_{MAX}$ ), we assess the feedback effects of vegetation phenology on soil moisture dynamics. The results indicate that the greening trend (as reflected by LAI increases) from 2000 to 2020 was significantly stronger than that observed during 1982–1999 across all seasons and climate zones. Both rootzone and surface soil moisture shifted from a decreasing (drying) trend (1982–1999) to an increasing (wetting) trend (2000–2020). From 1982 to 2020, the LAI induced moistening trends in both surface and rootzone soil moisture. In arid and temperate zones, precipitation drove rootzone soil moisture increases only during the summer. Among all seasons and climate zones, solar radiation induced the strongest surface soil drying in tropical summers, with a rate of  $-0.04 \times 10^{-3} \text{ m}^3\text{m}^{-3}/\text{Wm}^{-2}$ . For rootzone soil moisture, LAI dominated over individual climatic factors in winter and spring globally. In contrast, solar radiation became the primary driver during summer and autumn, followed by precipitation. For surface soil moisture, precipitation exhibited the strongest control in winter, but solar radiation surpassed it as the dominant factor from spring through autumn. In the tropical autumn, the sensitivity of rootzone and surface soil moisture to POS (and  $LAI_{MAX}$ ) was highest, at  $0.059 \text{ m}^3\text{m}^{-3}\cdot\text{d}^{-1}$  ( $0.256 \text{ m}^3\text{m}^{-3}/\text{m}^2\text{m}^{-2}$ ) and  $0.052 \text{ m}^3\text{m}^{-3}\cdot\text{d}^{-1}$  ( $0.232 \text{ m}^3\text{m}^{-3}/\text{m}^2\text{m}^{-2}$ ), respectively. This research deepens the understanding of how climate and vegetation regulate soil moisture across different climate zones and seasons. It also provides a scientific basis for improving global soil moisture prediction models and managing water resource risks in the context of climate change.

**Keywords:** rootzone and surface soil moisture; vegetation greening; climate factors; phenological feedback

## 1. Introduction

Soil moisture is a crucial component of the terrestrial–atmospheric hydrological cycle, exerting significant influences on surface temperature, precipitation patterns, and vegetation growth [1]. Optimal soil water content not only sustains the physiological metabolism of terrestrial plants but also enhances photosynthetic efficiency, nutrient uptake, and carbon sequestration capacity [2]. Through transpiration processes, terrestrial vegetation actively participates in global carbon–water cycles by releasing substantial amounts of soil water absorbed from various depths back into the atmosphere [3]. Global climate warming is

posing a profound threat to global terrestrial ecosystems through multiple transmission mechanisms. Rising temperatures increase saturation vapor pressure, intensifying atmospheric evaporative demand to trigger or exacerbate “atmospheric drought”, particularly in vulnerable arid regions [4]. At the same time, shifts in extreme weather patterns induced by climate warming are leading to more frequent, intense, and prolonged flash droughts, presenting a severe challenge to tropical and temperate regions like Southern and Eastern Asia and the Amazon Basin [5]. In high-altitude alpine regions highly sensitive to climate change, such as the Tibetan Plateau, sustained warming drives permafrost degradation and thaw, a process that diminishes the soil’s water-holding capacity, reduces surface runoff, and ultimately triggers a series of cascading ecological problems, including wetland shrinkage and lake desiccation in the source region of the Yellow River [6]. Overall, the contraction of moist soil areas and the widespread decline in water content resulting from climate warming collectively constrain improvements in vegetation productivity and the carbon assimilation capacity of ecosystems [7–10]. In response to soil moisture scarcity, vegetation employs adaptive strategies to reduce dependence on soil moisture, such as regulating stomatal conductance to lower photosynthetic intensity and minimize water loss [11]. Soil moisture exerts a limiting effect on vegetation growth, while vegetation reciprocally modulates soil water dynamics. Through canopy interception, plants capture a portion of precipitation, which evaporates back into the atmosphere before infiltrating the soil, thereby reducing effective water recharge. Simultaneously, root systems enhance soil porosity, improve infiltration capacity, and increase surface roughness, collectively altering runoff processes [12]. Moreover, vegetation facilitates the accumulation of soil organic matter, improves soil structure, and enhances water retention capacity. Healthy plant cover mitigates soil erosion, conserves water resources, and ultimately bolsters ecosystem stability and productivity [13]. Soil moisture is further influenced by climatic factors, including precipitation, temperature, radiation, and wind speed. While precipitation directly replenishes soil water, temperature, radiation, and wind indirectly affect moisture levels by altering evapotranspiration rates [14,15]. This coupled vegetation-soil-climate interaction holds profound implications for global carbon sequestration potential, sustainable water resource utilization, and the delivery of ecosystem services.

Satellite observations have confirmed a significant global greening trend in terrestrial vegetation over recent decades. While this vegetation expansion enhances ecosystem services, it concurrently intensifies soil water consumption, with particularly pronounced impacts in water-scarce arid and semi-arid regions [16,17]. As a key indicator of vegetation greening, the LAI is a crucial parameter governing the eco-hydrological interactions between the canopy and soil [18]. Current research primarily examines vegetation-induced asymmetries in diurnal land surface temperature variation, vegetation sensitivity to soil moisture, and the impacts of vegetation changes on evapotranspiration and water yield [19–21]. However, critical knowledge gaps remain regarding the regulatory roles of vegetation and climate on global soil moisture dynamics, particularly their differential responses across climate zones and seasonal scales. Addressing this research gap will provide deeper insights into vegetation-soil moisture interactions, offering a stronger scientific foundation for climate change adaptation and optimized land-water resource management.

The POS, a key metric in vegetation phenology, pinpoints the specific time of year when biomass accumulation reaches its maximum [22]. This timing represents a critical phenological transition, marking the shift from the phase of peak photosynthetic activity to subsequent senescence [23]. Furthermore, the timing of the POS directly influences the annual carbon sequestration capacity of an ecosystem [24,25]. In remote sensing studies, the date of the maximum LAI ( $LAI_{MAX}$ )—a key indicator reflecting canopy greenness and

structure—is widely employed to effectively identify the occurrence of the POS [26–28]. It is important to note, however, that while the POS and the  $LAI_{MAX}$  may overlap in time, they represent two distinct yet related ecological dimensions of vegetation response: the timing of growth and its peak magnitude. Shifts in POS, whether earlier or later, can be seen as a strategic adjustment in vegetation growth and development to adapt to climate change, whereas  $LAI_{MAX}$  is a more direct indicator of the ecosystem’s productivity potential [29,30]. For instance, in the mid-to-high latitudes of the Northern Hemisphere, earlier vegetation green-up and rising temperatures promote an increase in early-season carbon uptake, which in turn advances the POS [23]. However, seasonal droughts may inhibit the growth potential of vegetation, leading to a lower  $LAI_{MAX}$  [31]. This decoupling of timing and magnitude in vegetation response underscores the complexity of plant-environment interactions. Soil moisture dynamics across different soil layers regulate vegetation growth through complex biophysical mechanisms, thereby significantly modulating phenological timing, including POS. In high-latitude ecosystems, shallow-rooted vegetation constrained by permafrost shows particularly pronounced responses: as temperatures decline and the active layer freezes, restricted access to soil moisture and nutrients substantially impedes normal vegetation development [32,33]. In contrast, arid and semi-arid regions demonstrate an inverse relationship, where even marginal soil moisture increases frequently serve as the dominant growth driver, triggering advanced greening phenology [34]. Current research has predominantly examined the unidirectional effects of climatic drivers (e.g., temperature and precipitation) on vegetation phenology, including POS, focusing on how these factors induce phenological shifts [35–38]. However, the feedback mechanisms whereby POS timing and its associated peak biomass influence soil moisture dynamics—with vegetation phenology acting as a mediator—remain substantially understudied. Notably, emerging evidence demonstrates that earlier spring greening in the Northern Hemisphere triggers late-spring soil moisture deficits that persist through summer, consequently intensifying the frequency and magnitude of summer heatwaves [39]. These findings strongly suggest that advanced POS may similarly amplify soil water depletion across multiple soil layers during spring and early summer, thereby exerting prolonged impacts on moisture availability throughout summer and autumn. Nevertheless, the biophysical processes governing this critical vegetation-soil moisture feedback remain poorly quantified in current scientific understanding.

Therefore, this study aims to unravel these complex eco-hydrological interactions from 1982 to 2020. The main objectives of this research are: (1) to analyze the spatiotemporal variations in rootzone and surface soil moisture, as well as LAI, across different global climate zones and seasons using long-term satellite observations; (2) to assess the regulatory effects of LAI and climate factors on soil moisture patterns and to determine the dominant controls of soil moisture under different seasonal and climatic conditions; (3) to investigate the spatiotemporal patterns of POS and  $LAI_{MAX}$  and their linkages with climate factors and soil moisture.

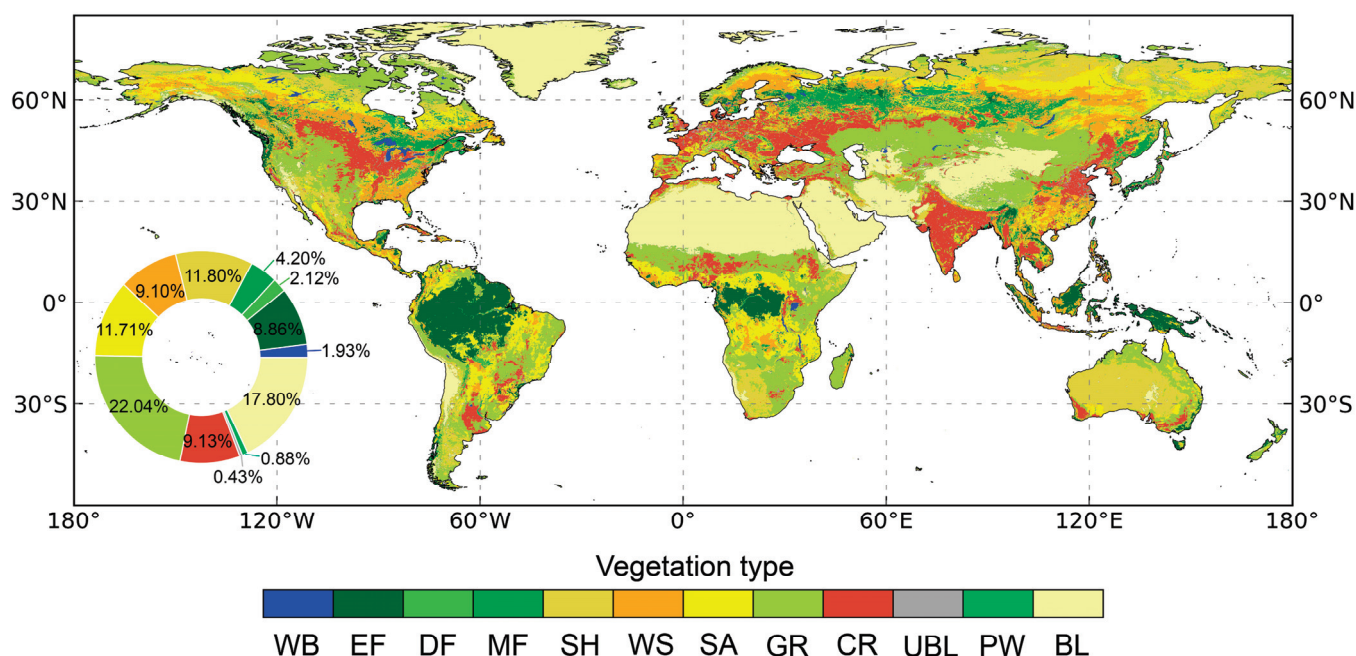
## 2. Materials and Methods

### 2.1. Study Area

The study area encompasses all continental landmasses (excluding Antarctica), characterized by diverse vegetation types (Figure 1). The land cover is dominated by grasslands (22.04%), followed by barren lands (17.80%), shrublands (11.80%), savannas (11.71%), croplands (9.13%), woody savannas (9.10%), evergreen forests (8.86%), mixed forests (4.20%), deciduous forests (2.12%), water bodies (1.93%), permanent wetlands (0.88%), and urban and built-up lands (0.43%). The study area features diverse climate types (Figure S1),



including cold (32.86%), arid (27.60%), tropical (15.99%), polar (12.65%), and temperate climates (10.90%). The area of plains below 200 m accounts for approximately 23.25% of the total land area. The climatic factors in the study area exhibit significant seasonal variation. The global average precipitation for winter, spring, summer, and autumn is 25.98 mm, 34.98 mm, 61.22 mm, and 46.94 mm, respectively. Summer precipitation is the most abundant, approximately 2.4 times that of winter. The corresponding average temperatures during these seasons are  $-13.93^{\circ}\text{C}$ ,  $-4.15^{\circ}\text{C}$ ,  $7.12^{\circ}\text{C}$ , and  $-5.27^{\circ}\text{C}$ . Solar radiation also exhibits significant seasonal variation, with average values of  $73.42\text{ W/m}^2$  in winter,  $203.50\text{ W/m}^2$  in spring,  $264.37\text{ W/m}^2$  in summer, and  $100.35\text{ W/m}^2$  in autumn. The intensity of solar radiation in summer is approximately 3.6 times higher than in winter. In contrast, the variation in wind speed is more moderate, with average values of 3.69 m/s, 3.98 m/s, 3.46 m/s, and 3.52 m/s for winter, spring, summer, and autumn, respectively. The diverse natural climatic conditions and complex ecosystem structure of the study area provide an ideal research setting for in-depth analysis of global soil moisture dynamics under climate change, as well as vegetation and climatic regulation on soil moisture.



**Figure 1.** The spatial distribution of global land cover types. The donut chart inset in the lower left corner represents the proportion of each land cover category worldwide. WB: Water Bodies; EF: Evergreen Forests; DF: Deciduous Forests; MF: Mixed Forests; SH: Shrublands; WS: Woody Savannas; SA: Savannas; GR: Grasslands; CR: Croplands; UBL: Urban and Built-up Lands; PW: Permanent Wetlands; BL: Barren Lands.

## 2.2. Data Sources

### 2.2.1. Satellite Dataset

The LAI serves as a crucial metric in quantifying vegetation dynamics [40–42]. To investigate the regulatory effect of long-term vegetation dynamics on soil moisture, this study utilized the GIMMS LAI4g dataset. The dataset, which spans the period from 1982 to 2020 with a spatial resolution of  $1/12^{\circ}$  and a 15-day temporal resolution, was generated by training a back propagation neural network (BPNN) model with 3.6 million high-quality samples and can be accessed via doi:10.5281/zenodo.7649107 [43]. Owing to its superior performance and widespread adoption, this dataset serves as a robust data source for analyzing long-term vegetation dynamics across multiple scales [44–47]. Global seasonal

LAI datasets were generated by merging data from the Northern (NH) and Southern (SH) Hemispheres to harmonize the seasons, defined as spring (NH: March–May; SH: September–November), summer (NH: June–August; SH: December–February), autumn (NH: September–November; SH: March–May), and winter (NH: December–February; SH: June–August) [48]. While this approach achieves a conceptual unification of seasons, it is a simplified model primarily intended for macro-scale analysis. It is important to recognize that this integration does not imply perfect global seasonal synchrony. That is, it does not reflect the precise, biophysically synchronous events at local levels, but rather provides a useful framework for global comparisons and trend analysis.

### 2.2.2. Soil Moisture and Meteorological Datasets

To investigate the dynamics of soil moisture at different depths and the regulatory effects of climate and vegetation, this study utilized two datasets: rootzone soil moisture and surface soil moisture. Both datasets were obtained from the National Earth System Science Data Center (<https://www.geodata.cn>). The dataset, which spanned from 1982 to 2020 with a spatial resolution of  $0.1^\circ$ , integrated multi-source soil moisture products, including ERA5-Land, the four-layer CFSR, and MERRA-2 (5 cm and 100 cm depths) [49]. The rootzone and surface soil moisture data from the dataset have been successfully used to construct the standardized ecological drought index (SEDI) to investigate the spatiotemporal dynamics and evolutionary trends of ecological drought in China [50]. Furthermore, its surface soil moisture data have also been applied to quantify the influence of soil moisture on vegetation's sensitivity to precipitation and temperature [51]. These successful applications demonstrate the value of this dataset; therefore, we selected it as the foundational data for our study and processed it as follows to meet our analytical requirements. To ensure consistent spatial resolution, the soil moisture dataset, originally at a  $0.1^\circ$  resolution, was resampled to  $1/12^\circ$ . Following the same methodology used for the LAI dataset to align seasons between hemispheres, we integrated soil moisture data from corresponding seasons in the Northern and Southern Hemispheres. Finally, global spring, summer, autumn, and winter datasets were generated for both rootzone and surface soil moisture.

Monthly meteorological data, including precipitation, temperature difference (the difference between monthly maximum and minimum temperatures), solar radiation, and wind speed, were sourced from the TerraClimate dataset (<https://www.climatologylab.org/>). The dataset has a native spatial resolution of 4.6 km and a temporal coverage from 1982 to 2020. To match the spatial resolution of other datasets and harmonize the seasons between the Northern and Southern Hemispheres, we first resampled the meteorological data to a  $1/12^\circ$  resolution. Subsequently, the data were aggregated based on the seasonal alignment between the hemispheres to construct a global seasonal meteorological dataset for spring, summer, autumn, and winter.

### 2.2.3. Land Cover and Climate Classification Datasets

The climate zone classification was based on the latest Köppen–Geiger climate classification system (<https://www.gloh2o.org/koppen/>, accessed on 1 May 2025) [52]. While this dataset provides classifications for historical, current, and future periods, this study selected the 1991–2020 map as it represents the climatic conditions of our study period. The classification scheme categorizes the globe into five major climate groups: Tropical, Arid, Temperate, Cold, and Polar. Due to its low vegetation cover, the Polar region was excluded from our analysis, focusing on the remaining four climate zones.

The MCD12C1 land cover dataset, derived from MODIS and based on the International Geosphere-Biosphere Programme (IGBP) classification system, was reclassified into twelve major types for the period from 2001 to 2020: water bodies, evergreen forests, deciduous forests, mixed forests, shrublands, woody savannas, savannas, grasslands, croplands, urban and built-up lands, permanent wetlands, and barren lands (see Table S1 for detailed classification information). To ensure our analysis focused on the characteristics of stable environments, we exclusively selected areas where the land cover type remained unchanged throughout the 2001–2020 period for subsequent analysis.

### 2.3. Methods

#### 2.3.1. Trend Analysis Based on Linear Regression

To analyze the seasonal trends of LAI and soil moisture, a linear regression model was employed, and the statistical significance of these trends was assessed using the Mann–Kendall test [19]. Specifically, the trend analysis involved calculating the slope of the linear regression for each pixel over three distinct periods: 1982–2020, 1982–1999, and 2000–2020, using Equation (1). The formula for the linear regression slope is as follows:

$$\text{Slope} = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n \sum_{i=1}^n X_i^2 - \left( \sum_{i=1}^n X_i \right)^2} \quad (1)$$

In the formula,  $X_i$  and  $Y_i$  represent the values of the independent and dependent variables for year  $i$ , respectively, and  $n$  is the total number of years in the study period. A slope value greater than 0 indicates an increasing trend in the dependent variable, whereas a value less than 0 indicates a decreasing trend.

#### 2.3.2. Soil Moisture Changes Induced by LAI and Climate Factors

The sensitivity of soil moisture to LAI and climatic factors was analyzed using a multiple linear regression model, as shown in Equation (2). In this model, soil moisture served as the response variable, while LAI and climatic factors were the explanatory variables. The sensitivity coefficient was defined as the partial derivative of soil moisture with respect to each explanatory variable. This approach aims to decouple the association between soil moisture and each explanatory variable from the confounding effects of other covariates [53]. To eliminate non-stationarity in the time series data, all variables were detrended prior to the multiple regression analysis. Additionally, the Variance Inflation Factor (VIF) was used as a key diagnostic tool to assess the impact of multicollinearity on the robustness of the regression model (Equation (3)) [54].

$$\Delta SM = \frac{\partial SM}{\partial LAI} \Delta LAI + \frac{\partial SM}{\partial PPT} \Delta PPT + \frac{\partial SM}{\partial Srad} \Delta Srad + \frac{\partial SM}{\partial Tem-D} \Delta Tem - D + \frac{\partial SM}{\partial WS} \Delta WS \quad (2)$$

$$VIF_i = \frac{1}{1 - R_i^2} \quad (3)$$

In Equation (2),  $\frac{\partial SM}{\partial LAI}$ ,  $\frac{\partial SM}{\partial PPT}$ ,  $\frac{\partial SM}{\partial Srad}$ ,  $\frac{\partial SM}{\partial Tem-D}$ , and  $\frac{\partial SM}{\partial WS}$  represent the sensitivity coefficients of soil moisture to LAI, precipitation (PPT), solar radiation (Srad), temperature difference (Tem-D), and wind speed (WS), respectively. Subsequently, the changes in soil moisture induced by LAI and climate factors ( $\delta SM^{LAI}$ ,  $\delta SM^{PPT}$ ,  $\delta SM^{Srad}$ ,  $\delta SM^{Tem-D}$ ,  $\delta SM^{WS}$ ) are calculated by multiplying the long-term trends from Equation (1) by their corresponding sensitivity coefficients [53]. The  $R_i^2$  in Equation (3) represents the coefficient of determination obtained by regressing one predictor variable (e.g., LAI) against all others (PPT, Srad, Tem-D, WS). The higher the  $R_i^2$ , the stronger the linear association

with other variables, thus yielding a larger VIF [55]. While  $VIF > 5$  can indicate potential multicollinearity, a threshold of  $VIF > 10$  is commonly adopted to diagnose severe cases [56]. Therefore, this study employs  $VIF > 10$  to identify severe multicollinearity.

### 2.3.3. Shapley Additive Explanations (SHAP)

To better understand the impacts of LAI and climatic factors on soil moisture at the pixel scale, this study employed a Random Forest (RF) model [57]. Given the limited sample size, we implemented a 5-fold cross-validation (CV) scheme to train and evaluate the RF model, ensuring its stability and the reliability of the results. This approach facilitates better utilization of the entire dataset and effectively mitigates the incidental bias arising from a single data partition. The hyperparameter tuning process was focused on preventing overfitting and optimizing stability. To strictly control the complexity of individual decision trees, we held several key parameters constant: `max_depth` was set to 3, `min_samples_split` to 3, and `min_samples_leaf` to 1. Concurrently, the number of trees (`n_estimators`) was designated as the core hyperparameter for optimization, with a search space of [30, 50, 70, 100, 120]. To further reduce the influence of random fluctuations, the entire 5-fold CV process was repeated 50 times for each candidate `n_estimators` value. The model's performance was evaluated using dual metrics—the coefficient of determination ( $R^2$ ) and the Root Mean Squared Error (RMSE). The `n_estimators` value that yielded the highest average  $R^2$  and the lowest average RMSE was selected as optimal, ensuring both the model's predictive accuracy and robustness.

Upon establishing the optimal model, we employed the SHAP (Shapley Additive Explanations) method for model interpretation [58]. SHAP is an explainable machine learning method that quantifies the contribution of each feature to a model's prediction by calculating SHAP values; a higher SHAP value indicates greater feature importance [59–61]. We applied SHAP analysis to assess the contribution of each factor to soil moisture and identified the dominant controlling factor for each pixel by selecting the feature with the highest SHAP score [62]. The SHAP value is calculated using the following formula:

$$\Phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(p - |S| - 1)!}{p!} [v(S \cup \{j\}) - v(S)] \quad (4)$$

where  $\Phi_j$  denotes the SHAP value for feature  $j$ ,  $S$  represents any possible subset of features that excludes feature  $j$ .  $N \setminus \{j\}$  represents the collection of all subsets that exclude feature  $j$ .  $p$  denotes the total number of features ( $|N| = p$ ).  $v(S)$  represents the prediction output of the model when using only the feature subset  $S$ .  $v(S \cup \{j\}) - v(S)$  represents the change in the model's prediction when feature  $j$  is added to the subset  $S$ . To ensure the reliability of the SHAP analysis conclusions, we conducted a consistency validation. This procedure involved fixing the optimal hyperparameters while varying only the `random_state` and retraining the model 20 times. We then calculated the frequency with which the initially identified top contributing variable maintained its first-place rank across these 20 runs, thereby verifying the consistency and reliability of our feature importance findings. The entire computational and validation workflow was executed on the Pengcheng Cloud Brain II high-performance computing platform.

### 2.3.4. LAI<sub>MAX</sub> and POS Extraction Methods

Snow cover can contaminate satellite-observed signals, which severely interferes with the time-series fitting process and, in turn, significantly reduces the accuracy of phenological information extraction [26,63]. Given that the GIMMS LAI4g dataset has already flagged snow-contaminated pixels during its production, our methodology proceeds as follows.

First, we performed gap-filling using a background value set to 5% of all snow-free observations. Subsequently, based on repeated testing and literature review, the Savitzky–Golay (SG) filter algorithm with a window size of 9 was applied to smooth the time series and filter out inherent noise [36]. Finally, a daily LAI time series was reconstructed using the cubic spline interpolation method (Equation (5)) [36,63].  $LAI_{MAX}$  is defined as the maximum value of the generated daily LAI time series, and POS is the day of the year on which  $LAI_{MAX}$  occurs.

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad (5)$$

where  $S_i(x)$  is the fitted daily LAI time series,  $x$  and  $x_i$  are the smoothed LAI values for two neighboring points, and  $a_i$ ,  $b_i$ ,  $c_i$ , and  $d_i$  are the fitting coefficients.

### 2.3.5. Correlation Analysis

In this study, Pearson correlation and partial correlation analyses were employed to investigate the linear associations among variables. Specifically, Pearson correlation analysis was used to assess the pairwise relationships between phenological metrics (POS,  $LAI_{MAX}$ ) and soil moisture at different depths (surface and rootzone). Partial correlation analysis was then applied to determine the relationship between individual climate factors (precipitation, solar radiation, temperature difference, and wind speed) and the phenological metrics (POS,  $LAI_{MAX}$ ), while controlling for the effects of other climatic variables. The coefficients of both partial and Pearson correlation analyses range from  $-1$  to  $1$ . A positive value indicates a positive correlation, while a negative value signifies a negative correlation. The magnitude of the absolute value reflects the strength of the correlation. The correlation is considered statistically significant when the  $p$ -value is less than  $0.05$ . The formulas for Pearson correlation analysis (Equation (6)) and partial correlation analysis (Equation (7)) are as follows:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6)$$

$$R_{xy \times z} = \frac{R_{xy} - R_{xz}R_{yz}}{\sqrt{(1 - R_{xz}^2)(1 - R_{yz}^2)}} \quad (7)$$

In Equation (6),  $r$  represents the Pearson correlation coefficient.  $x_i$  denotes the POS( $LAI_{MAX}$ ) for year  $i$ .  $y_i$  represents the soil moisture (rootzone and surface) for year  $i$ .  $n$  indicates the number of years, and  $\bar{x}$  and  $\bar{y}$  represent the corresponding variable averages over the study period. In Equation (7),  $R_{xy \times z}$  refers to the correlation between soil moisture ( $x$ ) and a single climate factor ( $y$ ), while controlling for other climatic variables ( $z$ ).  $R_{xz}$  denotes the correlation between variables  $x$  and  $z$ .  $R_{yz}$  represents the correlation between  $y$  and  $z$ , and  $R_{xy}$  indicates the correlation between variables  $x$  and  $y$ .

### 2.3.6. Sensitivity of Rootzone and Surface Soil Moisture to POS and $LAI_{MAX}$

The sensitivity of rootzone and surface soil moisture to POS and  $LAI_{MAX}$  was quantified using the slope of a linear regression (Equation (8)) [26,64]. The sensitivity coefficient quantifies the magnitude of soil moisture variation (unit:  $m^3/m^3 \cdot d^{-1}$  or  $m^3m^{-3}/m^2m^{-2}$ ) induced by changes in POS and  $LAI_{MAX}$ . A negative value indicates an inverse relationship, meaning that an increase in POS and  $LAI_{MAX}$  is associated with a decrease in rootzone and surface soil moisture, and vice versa.

$$Y = Sensitivity \times Var + \varepsilon \quad (8)$$



where  $Var$  represents POS and  $LAI_{MAX}$ ,  $Y$  denotes the rootzone soil moisture and surface soil moisture, and  $\varepsilon$  is the random error.

### 2.3.7. Hurst Exponent

The Hurst exponent, an indicator for evaluating the persistence (or long-term memory) and consistency within long-term time series data, has found extensive application in the fields of climatology and vegetation science [65–67]. The main computational procedure is as follows:

- (1) Divide the long time series into sub-series, and for each sub-series.
- (2) Define the mean sequence of the time series.

$$\bar{V} = \frac{1}{\tau} \sum_{t=1}^{\tau} \chi_{(t)}, \tau = 1, 2, \dots, n \quad (9)$$

- (3) Calculate the accumulated deviation.

$$\chi_{(t,\tau)} = \sum_{u=1}^t (V_{(u)} - \bar{V}_{\tau}), 1 \leq t \leq \tau \quad (10)$$

- (4) Define the range sequence.

$$R_{(\tau)} = \max_{1 \leq t \leq \tau} \chi_{(t,\tau)} - \min_{1 \leq t \leq \tau} \chi_{(t,\tau)}, \tau = 1, 2, \dots, n \quad (11)$$

- (5) Define the standard deviation sequence.

$$S_{(\tau)} = \left[ \frac{1}{\tau} \sum_{t=1}^{\tau} (V_{(t)} - \bar{V}_{(\tau)})^2 \right]^{1/2}, \tau = 1, 2, \dots, n \quad (12)$$

- (6) Compute the Hurst exponent.

$$\frac{R_{(\tau)}}{S_{(\tau)}} = (c\tau) \quad (13)$$

- (7) The H value (the Hurst exponent) is obtained by the least squares fit formula:

$$\log(R/S)_n = a + H \times \log(n) \quad (14)$$

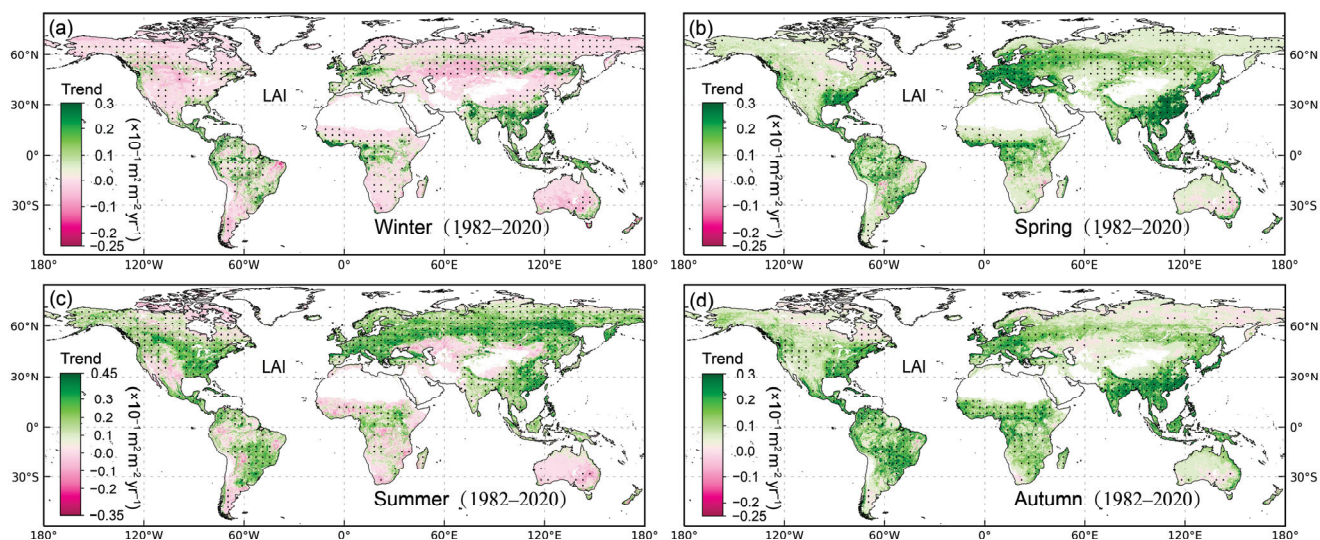
The Hurst exponent (H) ranges between 0 and 1. When  $H < 0.5$ , it indicates that future trends in the time series will exhibit anti-persistence (a tendency to reverse relative to the current state), with greater anti-persistence as H decreases. Conversely, when  $H > 0.5$ , it suggests future trends will maintain persistence (continue in the same direction), with stronger consistency as H increases. A value of  $H = 0.5$  signifies that future changes are uncorrelated with the current state (random fluctuations). To further analyze the temporal variation characteristics of vegetation phenology (POS and  $LAI_{MAX}$ ) from 1982 to 2020, this study classified the Hurst exponent into four levels: strong anti-persistence (0–0.3), weak anti-persistence (0.3–0.45), random fluctuation (0.45–0.55), and strong persistence (0.55–1.0).

## 3. Results

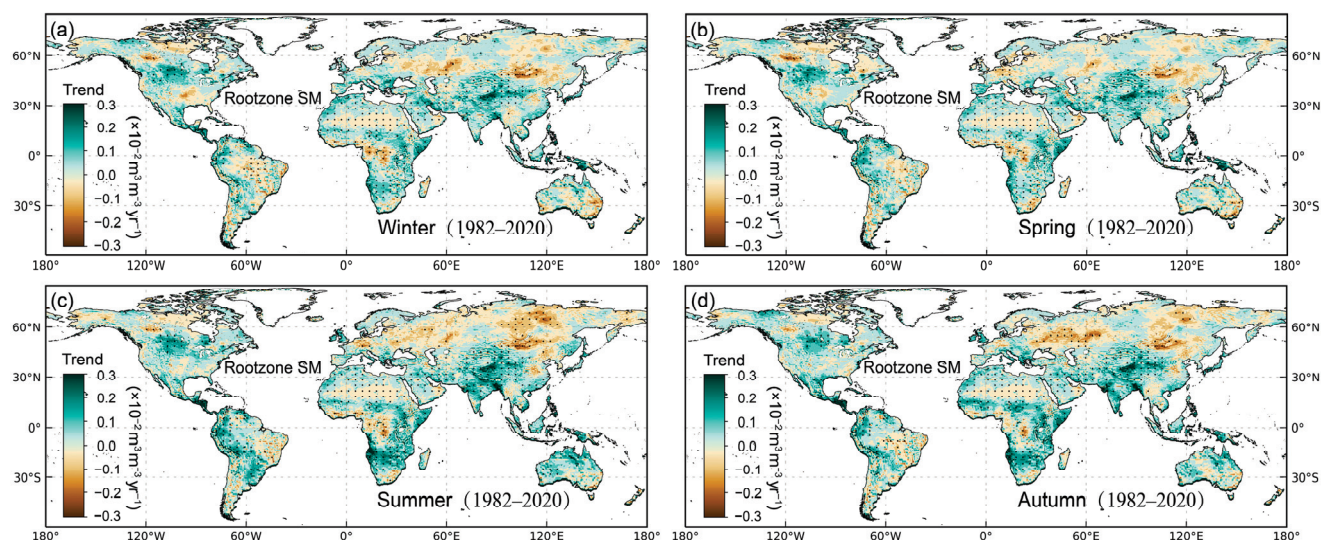
### 3.1. Spatiotemporal Variation in Soil Moisture and LAI

From 1982 to 2020, both LAI and soil moisture exhibited significant seasonal and regional variations. Regarding LAI (Figure 2), a notable trend of greening (increase in LAI) during

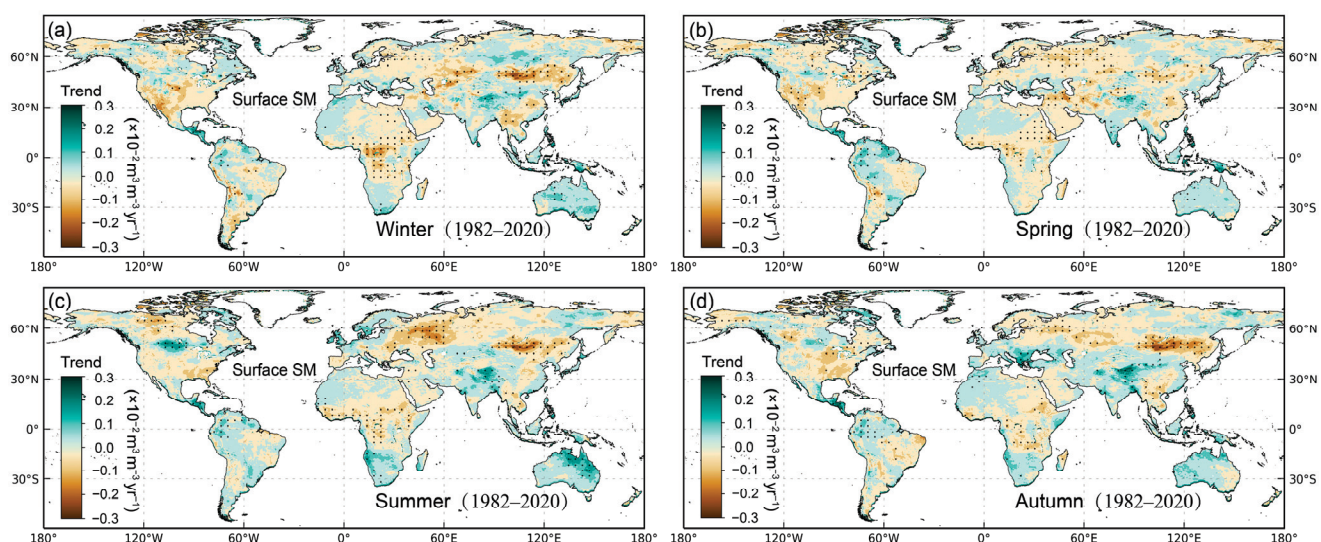
winter was observed across multiple regions, with Southeast Asia showing the most pronounced increase ( $0.13 \times 10^{-1} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ ), followed by Central Europe ( $0.10 \times 10^{-1} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ ), India ( $0.09 \times 10^{-1} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ ), Southeast Australia ( $0.083 \times 10^{-1} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ ), and the tropical rainforest regions of Africa and South America ( $0.08 \times 10^{-1}$  and  $0.078 \times 10^{-1} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ , respectively). A significant greening trend was also evident in China ( $0.06 \times 10^{-1} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ ). In spring, a widespread greening trend was observed across most regions globally, with an average change rate of  $0.058 \times 10^{-1} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ , while localized browning (decrease in LAI) trends were relatively weak (average  $-0.044 \times 10^{-1} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ ). During summer, the greening trend peaked across the entire Northern Hemisphere, except in arid regions, with an average change rate of  $0.10 \times 10^{-1} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ ; conversely, Australia exhibited a clear browning trend (average  $-0.03 \times 10^{-1} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ ). In autumn, a greening trend was similarly maintained across most regions globally, with an average change rate of  $0.062 \times 10^{-1} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ . As shown for rootzone soil moisture (Figure 3), winter and spring featured spatially consistent global patterns of wetting and drying. The average rate of increase was  $0.089 \times 10^{-2} \text{ m}^3\text{m}^{-3} \text{ yr}^{-1}$  in winter and  $0.088 \times 10^{-2} \text{ m}^3\text{m}^{-3} \text{ yr}^{-1}$  in spring, while the average rate of decrease was  $-0.065 \times 10^{-2}$  and  $-0.063 \times 10^{-2} \text{ m}^3\text{m}^{-3} \text{ yr}^{-1}$ , respectively. Summer trends demonstrated pronounced regional variations. Eurasia was characterized by a dipole pattern, with drying in the north-central region ( $-0.067 \times 10^{-2} \text{ m}^3\text{m}^{-3} \text{ yr}^{-1}$ ) and wetting in the south-central region ( $0.076 \times 10^{-2} \text{ m}^3\text{m}^{-3} \text{ yr}^{-1}$ ). Meanwhile, significant wetting occurred in parts of the Southern Hemisphere, including southern Africa ( $0.11 \times 10^{-2} \text{ m}^3\text{m}^{-3} \text{ yr}^{-1}$ ), northwestern South America ( $0.077 \times 10^{-2} \text{ m}^3\text{m}^{-3} \text{ yr}^{-1}$ ), and southeastern South America ( $0.10 \times 10^{-2} \text{ m}^3\text{m}^{-3} \text{ yr}^{-1}$ ). In autumn, the drying zone across north-central Eurasia shrank ( $-0.063 \times 10^{-2} \text{ m}^3\text{m}^{-3} \text{ yr}^{-1}$ ). In contrast, surface soil moisture trends (Figure 4) were generally weaker in magnitude. Notably, large swaths of the Northern Hemisphere experienced a consistent drying trend across all seasons (rates of  $-0.037 \times 10^{-2}$ ,  $-0.035 \times 10^{-2}$ ,  $-0.037 \times 10^{-2}$ , and  $-0.037 \times 10^{-2}$  for winter, spring, summer, and autumn, respectively; unit:  $\text{m}^3\text{m}^{-3} \text{ yr}^{-1}$ ). Conversely, Australia in the Southern Hemisphere displayed a persistent wetting trend throughout the four seasons (rates of  $0.043 \times 10^{-2}$ ,  $0.027 \times 10^{-2}$ ,  $0.052 \times 10^{-2}$ , and  $0.043 \times 10^{-2}$  for winter, spring, summer, and autumn, respectively; unit:  $\text{m}^3\text{m}^{-3} \text{ yr}^{-1}$ ).



**Figure 2.** The spatial distribution of global seasonal LAI trends from 1982 to 2020. Panels show the trends for (a) winter, (b) spring, (c) summer, and (d) autumn. Black dots indicate trend pixels that have passed the significance MK test.



**Figure 3.** The spatial distribution of seasonal trends in global rootzone soil moisture from 1982 to 2020. Panels show the trends for (a) winter, (b) spring, (c) summer, and (d) autumn. Black dots indicate trend pixels that have passed the significance MK test.



**Figure 4.** The spatial distribution of seasonal trends in global surface soil moisture from 1982 to 2020. Panels show the trends for (a) winter, (b) spring, (c) summer, and (d) autumn. Black dots indicate trend pixels that have passed the significance MK test.

A comparative analysis of global LAI data for two distinct periods (1982–1999 vs. 2000–2020) revealed a significantly accelerated greening trend across all seasons during 2000–2020 compared to the earlier period (Figures S2 and S3). Regarding soil moisture, changes in rootzone soil moisture exhibited distinct spatiotemporal heterogeneity. In central Eurasia, large areas that experienced soil moisture decrease across all seasons during 1982–1999 shifted to a significant increasing trend during 2000–2020. Conversely, northeastern Asia showed an opposite pattern of change (Figures S4 and S5). Furthermore, regions in Africa and South America transitioned from an earlier decrease in rootzone soil moisture to a significant increasing trend during the 2000–2020 period. Overall, the rate of rootzone soil moisture increase in the 21st century was faster than that during 1982–1999. For surface soil moisture, most regions globally showed a general decreasing trend during



1982–1999. After the year 2000, however, this decreasing trend became mainly concentrated in the high-latitude regions of the Northern Hemisphere (Figures S6 and S7).

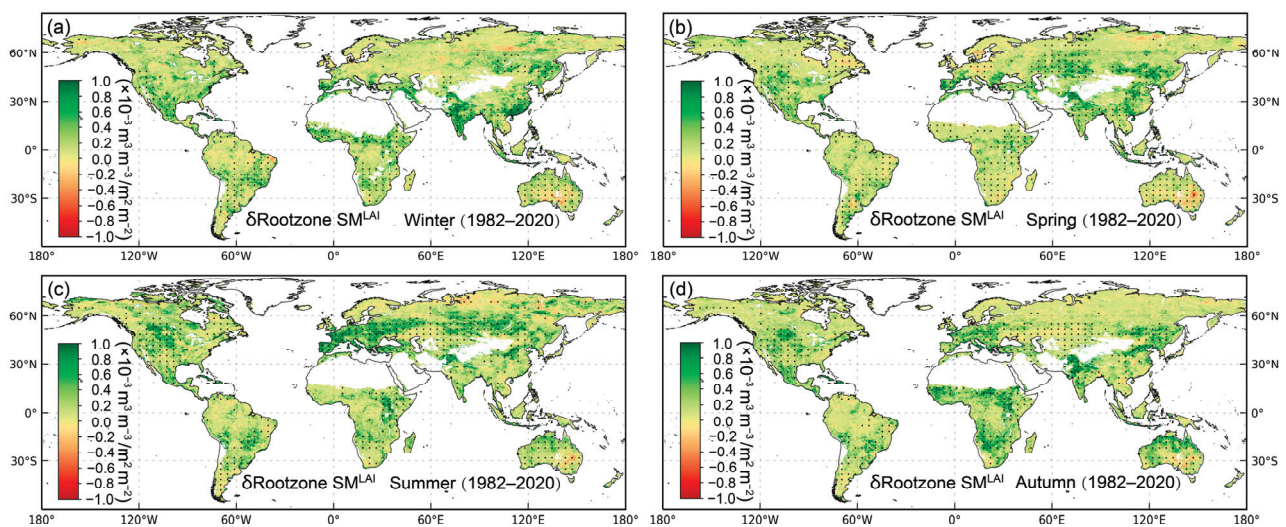
From 1982 to 2020, the average LAI for all major vegetation types globally peaked in summer. Specifically, deciduous forests exhibited the highest LAI ( $4.86 \text{ m}^2\text{m}^{-2}$ ), followed by evergreen forests ( $4.60 \text{ m}^2\text{m}^{-2}$ ), mixed forests ( $4.27 \text{ m}^2\text{m}^{-2}$ ), woody savannas ( $3.02 \text{ m}^2\text{m}^{-2}$ ), savannas ( $2.31 \text{ m}^2\text{m}^{-2}$ ), croplands ( $1.94 \text{ m}^2\text{m}^{-2}$ ), grasslands ( $1.02 \text{ m}^2\text{m}^{-2}$ ), and shrublands ( $0.93 \text{ m}^2\text{m}^{-2}$ ) (Table S2). Furthermore, across all four seasons, both rootzone and surface soil moisture in mixed forests were generally higher than in other vegetation types. Regarding seasonal greening trends (Table S3), deciduous and mixed forests showed the most rapid LAI increase in spring, at a rate of approximately  $0.69 \times 10^{-2} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ . In summer, the increasing trend was most pronounced in mixed forests ( $0.72 \times 10^{-2} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ ), whereas grasslands exhibited the slowest trend ( $0.18 \times 10^{-2} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ ). During autumn, deciduous and mixed forests continued to show strong growth rates of  $0.64 \times 10^{-2} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$  and  $0.48 \times 10^{-2} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ , respectively. In terms of long-term changes in soil moisture (Table S3), an increasing trend was observed across all seasons for both rootzone and surface soil moisture in evergreen forests, deciduous forests, woody savannas, savannas, and croplands. Among these, evergreen forests demonstrated the largest increase, followed by deciduous forests. In contrast, mixed forests and shrublands experienced a decreasing trend in surface soil moisture across all seasons, while their rootzone soil moisture showed a decline only in specific seasons. For grasslands, rootzone soil moisture increased across all four seasons, whereas surface soil moisture only showed an increasing trend during summer and autumn.

From 1982 to 2020, the average LAI trend across all climate zones and seasons indicated a widespread greening (Figure S8). The LAI trend was lowest in arid zones, while the average spring LAI trend in temperate zones was the highest, reaching  $0.05 \times 10^{-1} \text{ m}^2\text{m}^{-2} \text{ yr}^{-1}$ . In tropical regions, the average LAI trends for spring and summer were lower than those for winter and autumn. During this period, both surface and rootzone soil moisture showed a decreasing average trend in cold climate zones. In contrast, these soil moisture layers generally exhibited an increasing trend across all seasons in other climate zones, including arid, temperate, and tropical regions (Figure S8). From 1982 to 1999 and from 2000 to 2020 (Figure S8), all climate zones exhibited a greening trend in LAI across all seasons, with the rate of change generally higher in the post-21st-century period (2000–2020) compared to the earlier period (1982–1999). In arid, cold, temperate, and tropical climate zones, rootzone soil moisture shifted from a decreasing trend during 1982–1999 to an increasing trend during 2000–2020. Similarly, surface soil moisture overall transitioned from a decreasing trend in 1982–1999 to an increasing trend in 2000–2020 across all climate zones. From 2000 to 2020, the increasing trends in rootzone and surface soil moisture were more pronounced in temperate and tropical climate zones than in cold and arid zones (Figures S8 and S9).

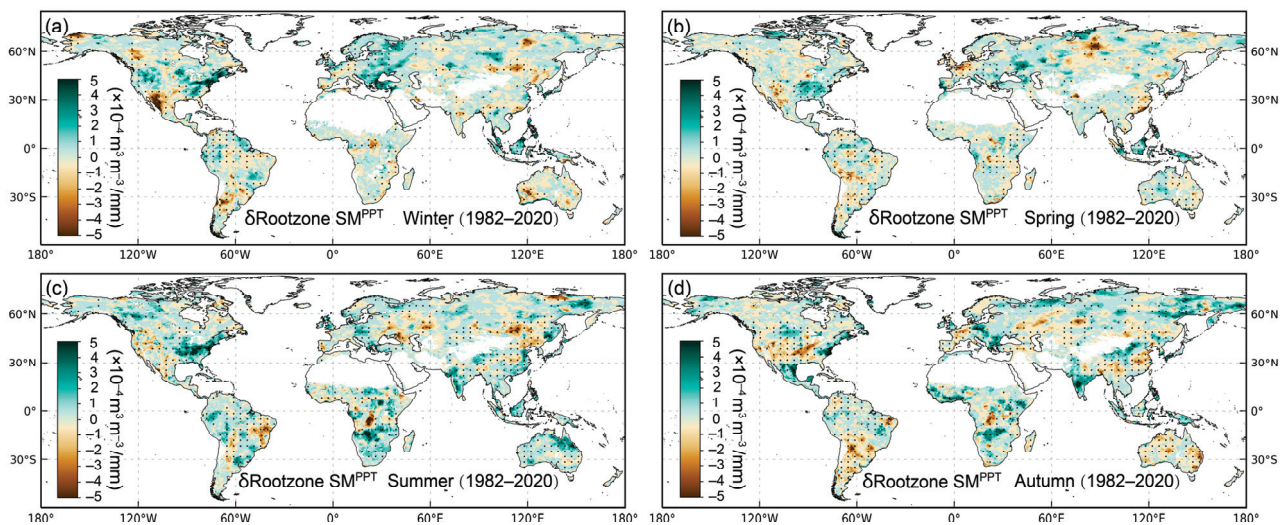
### 3.2. Changes in Soil Moisture Induced by LAI and Climate Factors

Table S4 presents the results of the Variance Inflation Factor (VIF) test for each variable across the entire study period (1982–2020) and two sub-periods (1982–1999 and 2000–2020). For all periods, the proportion of pixels with  $\text{VIF} < 10$  for each variable is generally above 95%. This indicates that the degree of multicollinearity among the selected variables is weak. Consequently, the use of a multiple linear regression model to analyze the sensitivity of soil moisture to LAI and climatic factors is both robust and appropriate. In the spatial distribution of rootzone and surface soil moisture changes induced by LAI (Figures 5 and S10), LAI caused a more pronounced increasing trend in rootzone soil moisture than in surface soil moisture in China and India during winter, as well as in central Eurasia during

summer. Precipitation exerted concentrated effects on both rootzone and surface soil moisture at regional scales, driving either increases or decreases (Figures 6 and S11). Summer solar radiation primarily reduced rootzone soil moisture in Eurasia and central Africa, while enhancing it in India, northern North America, Australia, southern Africa, and central-northern South America (Figure 7). In autumn, solar radiation induced a significant increasing trend in rootzone soil moisture but a decreasing trend in surface soil moisture in India (Figure S12). Winter temperature differences led to a marked increase in rootzone soil moisture accompanied by decreased surface soil moisture in north-central Eurasia (Figures 8 and S13). During summer, temperature differences caused reduced rootzone soil moisture in central Europe while moderating the declining trend of surface soil moisture in the same region. Furthermore, summer wind speed created stratified soil moisture patterns in northeastern South America, characterized by a significant decreasing trend in rootzone soil moisture but an increasing trend in surface soil moisture (Figures 9 and S14).

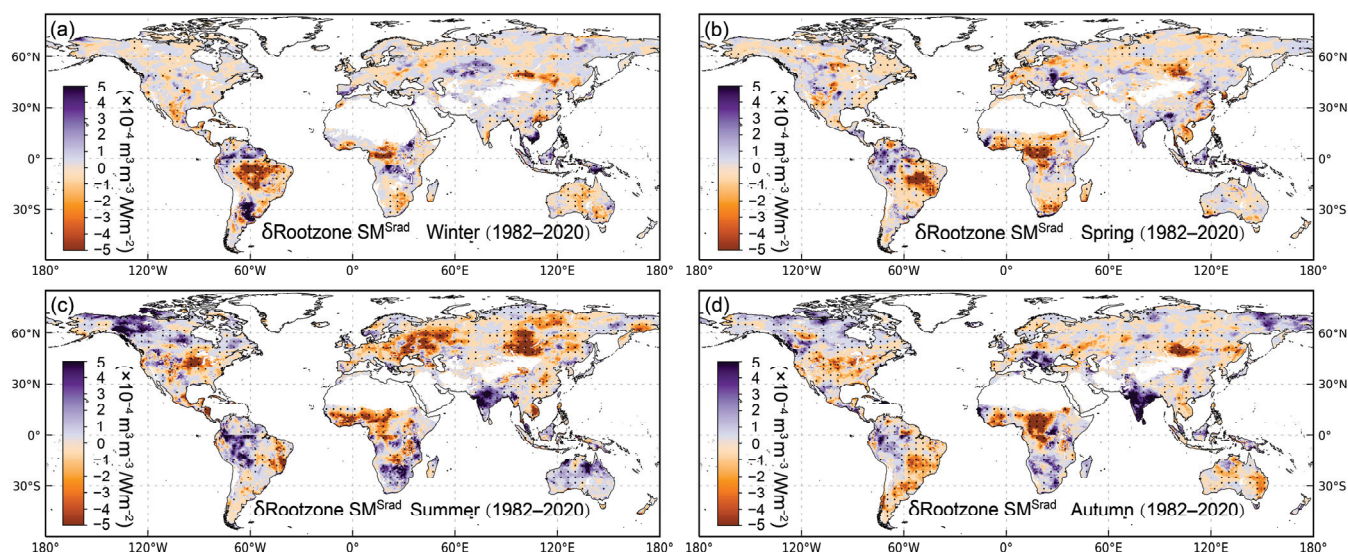


**Figure 5.** The spatial distribution of rootzone soil moisture changes induced by LAI from 1982 to 2020. The panels show the results for (a) winter, (b) spring, (c) summer, and (d) autumn. Black dots indicate pixels that passed the significance test ( $p < 0.05$ ).

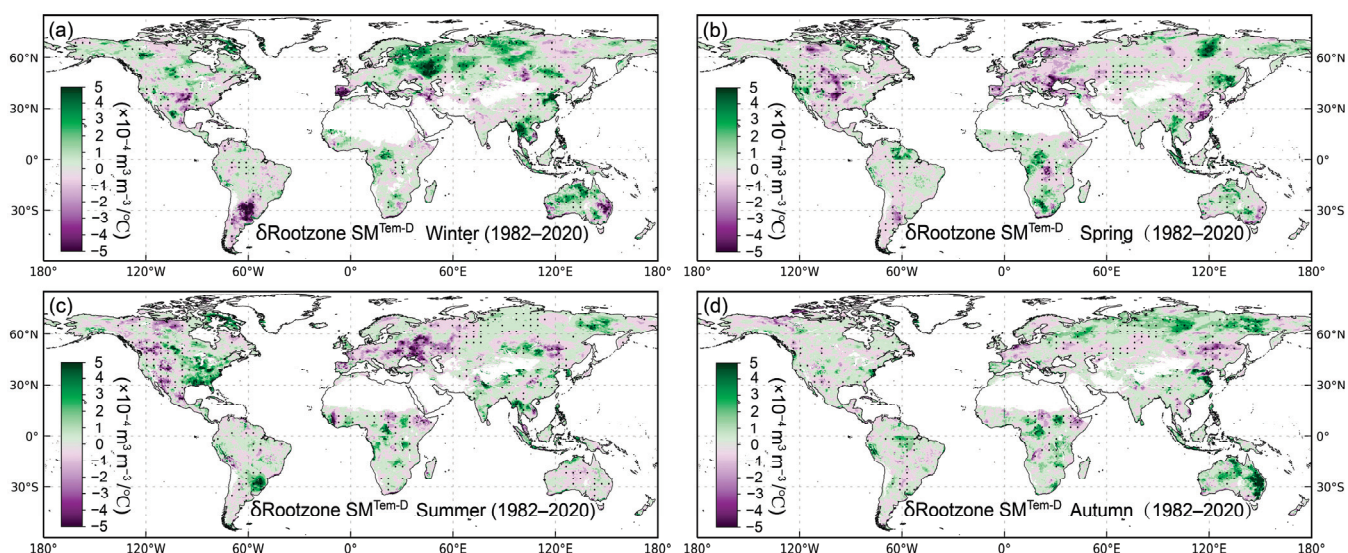


**Figure 6.** The spatial distribution of rootzone soil moisture changes induced by precipitation from 1982 to 2020. The panels show the results for (a) winter, (b) spring, (c) summer, and (d) autumn. Black dots indicate pixels that passed the significance test ( $p < 0.05$ ).





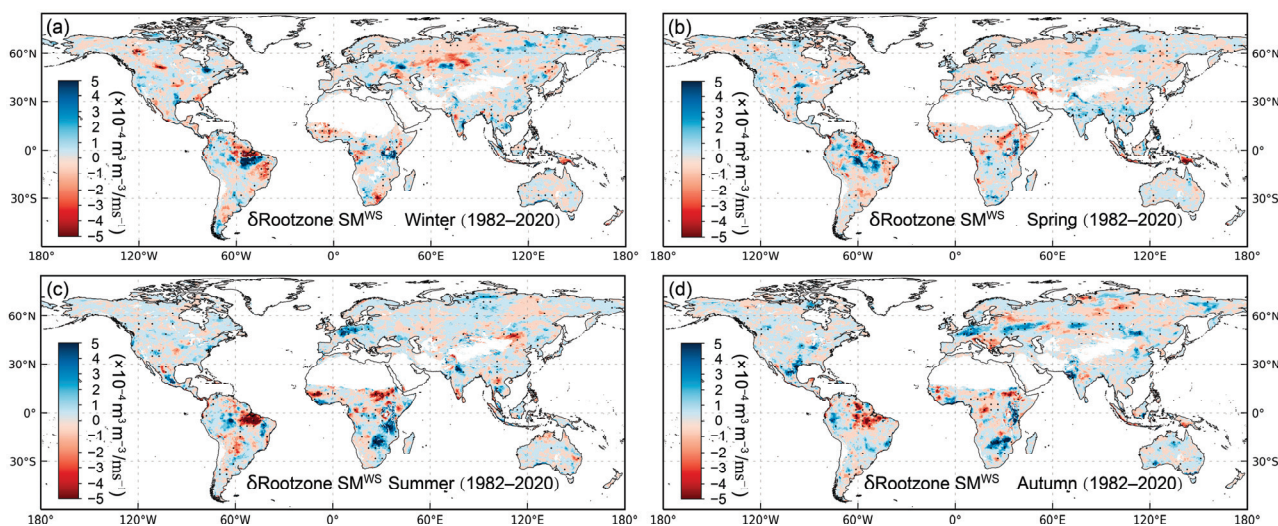
**Figure 7.** The spatial distribution of rootzone soil moisture changes induced by solar radiation from 1982 to 2020. The panels show the results for (a) winter, (b) spring, (c) summer, and (d) autumn. Black dots indicate pixels that passed the significance test ( $p < 0.05$ ).



**Figure 8.** The spatial distribution of rootzone soil moisture changes induced by temperature differences from 1982 to 2020. The panels show the results for (a) winter, (b) spring, (c) summer, and (d) autumn. Black dots indicate pixels that passed the significance test ( $p < 0.05$ ).

The comparative analysis of two periods (1982–1999 vs. 2000–2020) found that in south-eastern China during winter, the increasing trend in LAI-induced rootzone soil moisture was more pronounced during 2000–2020 than in the 1982–1999 period (Figures S15 and S20). For the Northern Hemisphere winter, precipitation-driven soil moisture increases exhibited a notable northward shift in spatial distribution during 2000–2020 compared to the earlier period (Figures S16 and S21). In spring across Eurasia, regions showing precipitation-induced moistening during 1982–1999 transitioned to drying trends by 2000–2020, with particularly severe rootzone moisture declines in some areas. Similarly, summer solar radiation-induced moistening areas in Eurasia during 1982–1999 were replaced by widespread drying trends in the 2000–2020 period (Figures S17 and S22). Summer temperature difference-induced moistening trends observed across most mid-to-high latitude Northern Hemisphere regions during

1982–1999 reversed to drying trends by 2000–2020 (Figures S18 and S23), with analogous transitions occurring in autumn across central-southern Europe. Wind speed-induced rootzone moisture patterns displayed fragmented distributions of both high-moisture and high-dryness areas across both periods, with relatively few extensive regions showing coherent moistening or drying trends (Figures S19 and S24).



**Figure 9.** The spatial distribution of rootzone soil moisture changes induced by wind speed from 1982 to 2020. The panels show the results for (a) winter, (b) spring, (c) summer, and (d) autumn. Black dots indicate pixels that passed the significance test ( $p < 0.05$ ).

Based on a comparative analysis of 1982–1999 and 2000–2020, LAI-driven surface moisture variations showed relatively minor differences between these two periods, with only limited areas exhibiting either strong moistening or drying trends (Figures S25 and S30). In contrast, precipitation-induced surface moisture changes were more pronounced, displaying distinct clustered spatial patterns (Figures S26 and S31). During spring in mid-to-high latitude regions of the Northern Hemisphere, solar radiation shifted from inducing moistening trends during 1982–1999 to causing drying trends in 2000–2020 (Figures S27 and S32). Temperature differences generated more dramatic surface moisture variations in summer and autumn than in winter and spring during 1982–1999, while their impacts weakened during 2000–2020 (Figures S28 and S33). Furthermore, autumn wind speed effects during 1982–1999 produced extreme localized moistening or drying trends in certain regions, but these influences diminished in the 2000–2020 period (Figures S29 and S34).

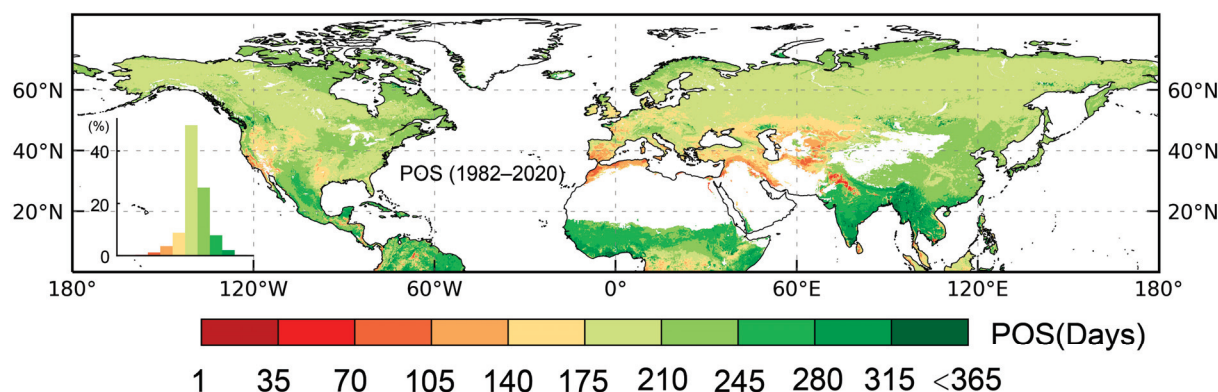
Across all climate zones, LAI consistently enhanced rootzone soil moisture in all seasons, with the strongest moistening trend ( $0.13 \times 10^{-3} \text{ m}^3 \text{ m}^{-3} / \text{m}^2 \text{ m}^{-2}$ ) occurring in temperate winters and the weakest ( $0.02 \times 10^{-3} \text{ m}^3 \text{ m}^{-3} / \text{m}^2 \text{ m}^{-2}$ ) in cold climate winters (Figure S35). Precipitation showed seasonal and zonal variability—it increased rootzone soil moisture only during summer in arid and temperate zones but exerted year-round positive effects in cold climates. Solar radiation universally reduced rootzone soil moisture across all climate zones and seasons. Temperature differences caused maximum drying in temperate springs ( $-0.06 \times 10^{-3} \text{ m}^3 \text{ m}^{-3} / ^\circ \text{C}$ ) but strongest moistening in cold climate winters ( $0.04 \times 10^{-3} \text{ m}^3 \text{ m}^{-3} / ^\circ \text{C}$ ). Wind speed induced more pronounced drying in tropical regions than elsewhere (Figure S36). For surface soil moisture, LAI consistently promoted moistening globally. Precipitation generated peak moistening in temperate summers ( $0.013 \times 10^{-3} \text{ m}^3 \text{ m}^{-3} / \text{mm}$ ), while solar radiation caused maximum drying in tropical summers ( $-0.04 \times 10^{-3} \text{ m}^3 \text{ m}^{-3} / \text{W m}^{-2}$ ). Temperature differences led to the



strongest surface moistening in arid zone springs ( $0.024 \times 10^{-3} \text{ m}^3 \text{ m}^{-3} / ^\circ\text{C}$ ), with wind speed showing the greatest impacts in tropical regions compared to other climate zones.

### 3.3. Spatiotemporal Variations in POS and LAI<sub>MAX</sub> in the Northern Hemisphere

Figure 10 shows the spatial distribution of POS in the Northern Hemisphere from 1982 to 2020. Areas with an earlier POS are mainly concentrated in arid regions and along the Mediterranean coast, while in cold climate zones, the POS is concentrated between 175 and 210 days. In the Northern Hemisphere, areas with LAI<sub>MAX</sub> in the range of  $0\text{--}2 \text{ m}^2/\text{m}^2$  account for 45.55% and are mainly distributed in arid regions and near the Arctic Circle (Figure S37). During the period 1982–2020, the POS for major vegetation types in the Northern Hemisphere exhibited an advancing trend. Among forest types, the average POS for evergreen, deciduous, and mixed forests was 201.32 d, 202.46 d, and 201.78 d, with advancement rates of  $0.037 \text{ d yr}^{-1}$ ,  $0.13 \text{ d yr}^{-1}$ , and  $0.025 \text{ d yr}^{-1}$ , respectively. For savannas and grasslands, the average POS for woody savannas, savannas, and grasslands was 204.93 d, 207.45 d, and 201.34 d, advancing at rates of  $0.011 \text{ d yr}^{-1}$ ,  $0.022 \text{ d yr}^{-1}$ , and  $0.036 \text{ d yr}^{-1}$ , respectively. Furthermore, the average POS for shrublands and croplands was 206.48 d and 201.62 d, with corresponding advancement rates of  $0.05 \text{ d yr}^{-1}$  and  $0.03 \text{ d yr}^{-1}$  (Figure S38).

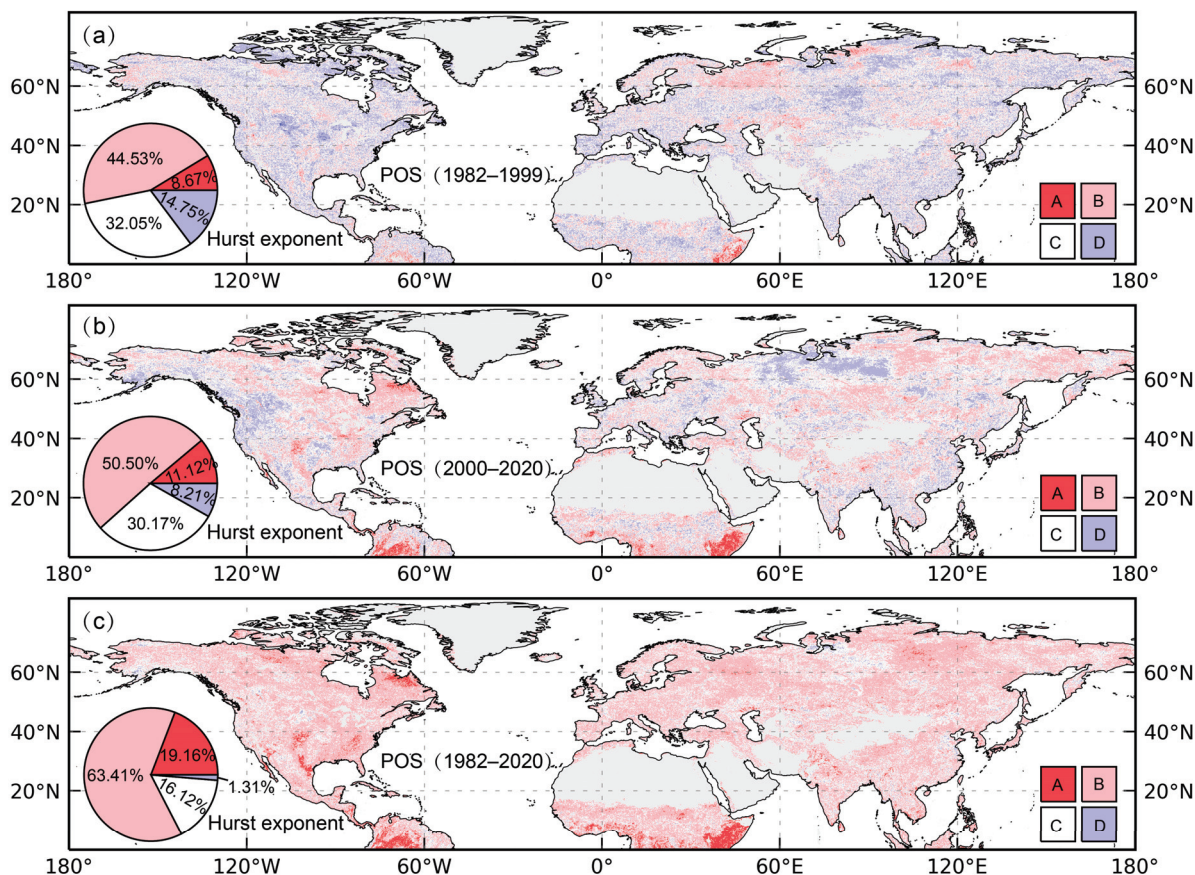


**Figure 10.** The spatial distribution of the average POS in the Northern Hemisphere from 1982 to 2020. The histogram in the bottom-left corner represents the frequency distribution of the corresponding dates.

Among all climate zones and time periods, the tropical region showed the largest POS advancement trend from 1982 to 1999, advancing at an average of  $0.33 \text{ d yr}^{-1}$  (Figure S39). In contrast, arid regions during the same period (1982–1999) had the largest POS delay trend, with an annual extension of 0.11 d. For LAI<sub>MAX</sub>, the increasing trend from 2000 to 2020 in all climate zones was more than twice that of the 1982–1999 period (Arid: 2.30 times, Cold: 2.89 times, Temperate: 2.56 times, Tropical: 4.50 times). The growth trend of LAI<sub>MAX</sub> was strongest in the temperate climate zone, with an annual growth rate of  $0.14 \times 10^{-1} \text{ m}^2/\text{m}^2$  from 2000 to 2020. In the Northern Hemisphere, from 1982 to 2020, 56.7% of the area exhibited an advancing trend in the POS at an average rate of  $0.27 \text{ days yr}^{-1}$ , while 46.3% showed a delaying trend at a rate of  $0.29 \text{ days yr}^{-1}$ . Concurrently, for the LAI<sub>MAX</sub>, 72.3% of the area demonstrated an increasing trend with a mean rate of  $0.012 \text{ m}^2 \text{ m}^{-2} \text{ yr}^{-1}$ , whereas the remaining 27.7% experienced a decreasing trend at a rate of  $0.010 \text{ m}^2 \text{ m}^{-2} \text{ yr}^{-1}$  (Figures S40 and S41).

Figure 11 shows the spatial distribution of the Hurst exponent for POS for different time spans. In both the 1982–1999 and 2000–2020 periods, more than 50% of the area in the Northern Hemisphere had a future POS trend opposite to the trend observed during the study period. For the entire period from 1982 to 2020, this proportion of areas with a

reversing trend reached 82.57%, while only 1.31% of the area showed a trend consistent with the past. For  $LAI_{MAX}$ , the area with a reversing future trend was 57.96% for the 1982–2020 period (Figure S42). Additionally, across all climate zones and time periods, the Hurst exponent for  $LAI_{MAX}$  was closer to 0.5 compared to that of POS (Figure S43).



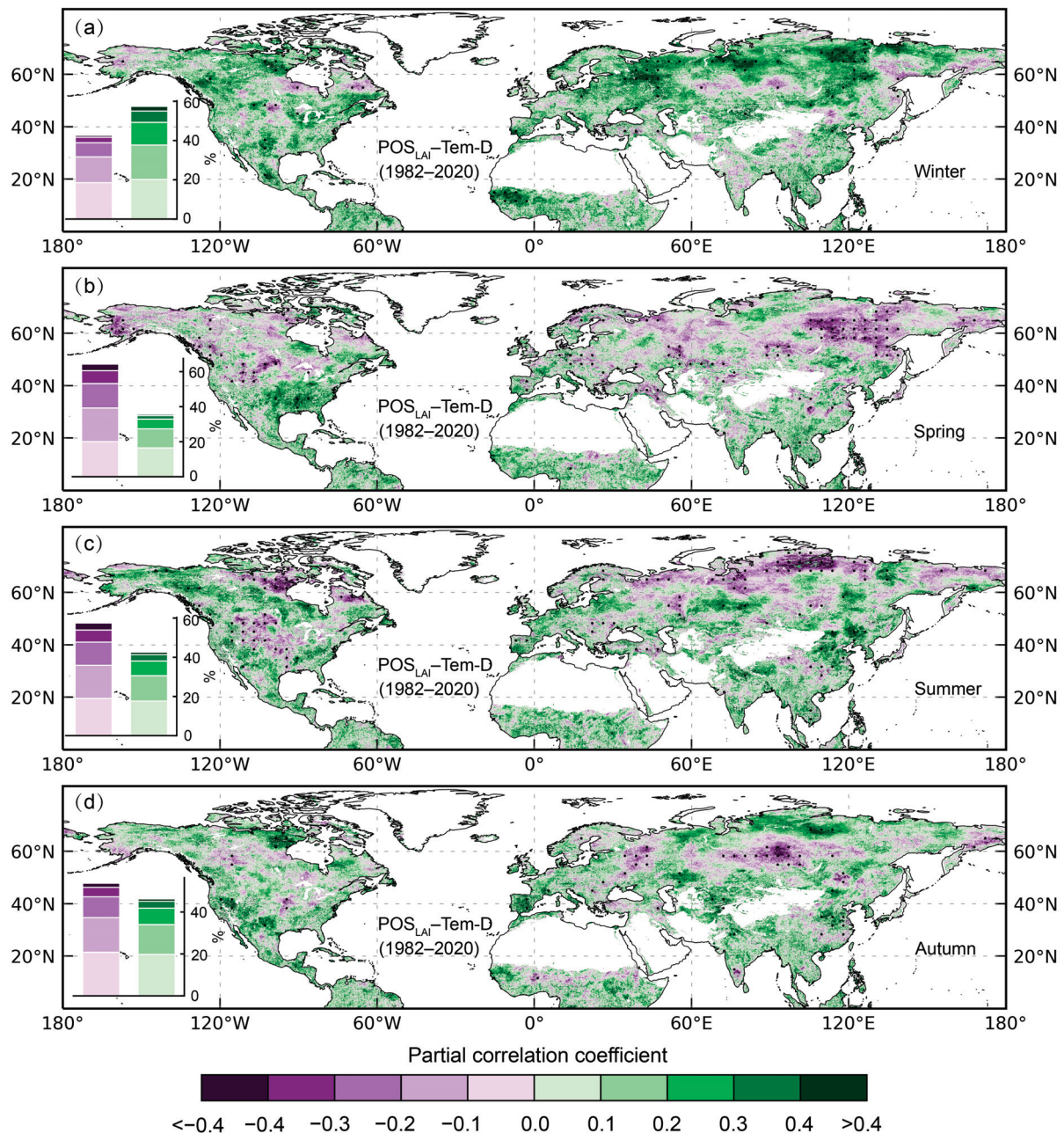
**Figure 11.** The spatial patterns of the Hurst exponent for POS in the Northern Hemisphere for the periods (a) 1982–1999, (b) 2000–2020, and (c) 1982–2020. The color of the squares indicates the nature of the trend: strong anti-persistent (dark red, A; 0–0.3), weak anti-persistent (light red, B; 0.3–0.45), random fluctuations (white, C; 0.45–0.55), or strong persistent (light blue, D; 0.55–1.0).

### 3.4. Correlation of POS and $LAI_{MAX}$ with Climate Factors

In the spatial distribution of partial correlation coefficients between climatic factors and POS across seasons in the Northern Hemisphere from 1982 to 2020 (Figures S44–S46 and 12), the differences in the proportions of positive and negative correlation regions between POS and precipitation, solar radiation, and wind speed were generally within 10%, exhibiting relatively homogeneous spatial heterogeneity. Notably, an asymmetric mechanism was observed between POS and seasonal temperature differences: the disparity in the proportions of positive and negative correlation regions reached as high as 28.56% in spring, and 14.80% and 14.86% in winter and summer, respectively. In the spatial distribution of partial correlation coefficients between climatic factors and  $LAI_{MAX}$  in the Northern Hemisphere from 1982 to 2020 (Figures S55–S58),  $LAI_{MAX}$  showed a higher proportion of positive correlations with precipitation than negative correlations across all seasons—winter (59.57% vs. 40.43%), spring (62.48% vs. 37.52%), summer (61.77% vs. 38.23%), and autumn (58.62% vs. 41.38%). The proportion of positive correlation regions between summer solar radiation and  $LAI_{MAX}$  reached 63.24%, higher than in other seasons, possibly due to seasonal variations in light-use efficiency. At the regional scale, seasonal solar radiation and  $LAI_{MAX}$  in



South Asia exhibited negative correlations of varying magnitudes. In the Northern Hemisphere, except for spring, where the difference in the proportions of positive (58.62%) and negative (41.38%) correlations between temperature differences and  $LAI_{MAX}$  was relatively large, the proportions of positive and negative correlation regions in other seasons were nearly balanced.  $LAI_{MAX}$  displayed a predominantly negative correlation with wind speed in winter (54.20% vs. 45.80%), spring (52.82% vs. 47.18%), summer (60.22% vs. 39.78%), and autumn (57.35% vs. 42.65%), which may reflect the inhibitory effect of wind-induced mechanical abrasion on  $LAI_{MAX}$ .



**Figure 12.** The spatial distribution of the partial correlation coefficients between  $POS$  and seasonal temperature difference in the Northern Hemisphere from 1982 to 2020. Panels (a), (b), (c), and (d) show the results for winter, spring, summer, and autumn, respectively. The histogram in the bottom-left corner of each panel shows the frequency distribution of the corresponding partial correlation coefficients. Black dots indicate pixels where the correlation is statistically significant ( $p < 0.05$ ).



A comparative analysis of climatic factors in relation to POS and LAI<sub>MAX</sub> across seasons in the Northern Hemisphere was conducted, comparing the periods 1982–1999 and 2000–2020 (Figures S47–S54 and S59–S66). Compared to 1982–1999, the spatial extent of the negative correlation between POS and the temperature difference in spring, summer, and autumn expanded significantly across the high latitudes of the Northern Hemisphere during 2000–2020. Over Eurasia, the negative correlation between summer wind speed and LAI<sub>MAX</sub> displayed a more fragmented spatial distribution in the 2000–2020 period than in 1982–1999. Furthermore, during 2000–2020, LAI<sub>MAX</sub> showed a significantly stronger positive correlation with summer solar radiation in mid-to-high latitude regions of Eurasia, with the correlation strength being more pronounced than in the earlier period (1982–1999).

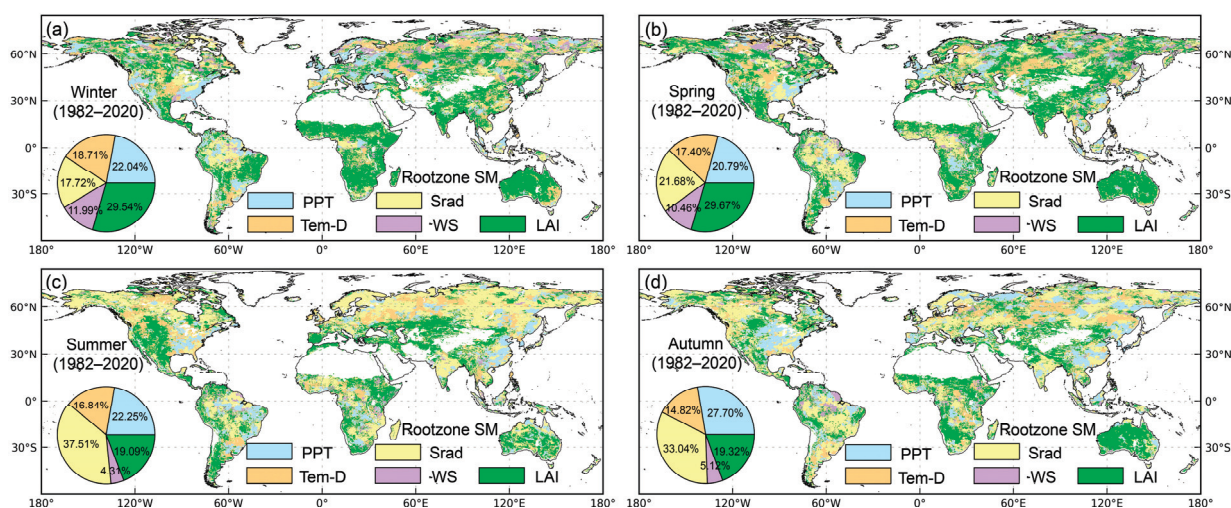
## 4. Discussion

### 4.1. Assessing Soil Moisture Changes and Dominant Contributing Factors

Surface and rootzone soil moisture display distinct seasonal dynamics, a result of their fundamentally different hydro-regulatory mechanisms in the soil profile. The surface soil layer is tightly coupled with the atmosphere [68]. Its shallow nature facilitates rapid water and energy exchanges, making its moisture content primarily controlled by precipitation and evaporation [69]. Consequently, surface moisture responds swiftly to rainfall but is also susceptible to rapid depletion from intense evaporation. In contrast, the rootzone soil moisture dynamics are more complex and buffered. As the main region of root activity, its water balance is governed by two dominant processes: root water uptake to support plant growth and transpiration [70], and water fluxes involving infiltration from the surface or percolation to deeper layers. This mechanistic disparity leads to a dynamic relationship characterized by alternating periods of coupling and decoupling. While the layers are strongly coupled during prolonged wet periods [71], they often decouple during droughts [72] or after light precipitation events, leading to divergent moisture responses. For instance, this decoupling is observed in northeastern South America during its summer. In this region, high temperatures, low relative humidity, and dense vegetation drive strong transpiration, as wind disrupts the leaf boundary layer and increases the vapor pressure deficit. This process causes continuous water extraction from the root zone, resulting in a decreasing moisture trend. Concurrently, prevailing summer winds transport Atlantic moisture inland, which can increase near-surface humidity and the probability of light precipitation events [73]. However, this precipitation is often insufficient in magnitude to effectively infiltrate and recharge the root zone.

The dominant drivers regulating rootzone and surface soil moisture across seasonal scales from 1982 to 2020 were systematically assessed (Figures 13 and 14), with a comparative analysis conducted for two distinct periods: 1982–1999 and 2000–2020 (Figures S67–S70). To ensure the validity of these findings, we assessed the stability of the factor rankings. Through 20 repeated validation runs (with optimal parameters fixed and only the random seed varied), the top-ranked contributing factor to soil moisture, as identified by the Random Forest-SHAP method, consistently demonstrated high stability in its ranking. Specifically, for each season, the most critical factor appeared in the first position in over 85% of the runs (Table S5). This result indicates that despite the stochasticity arising from different random initializations, the dominant role of the core influencing factors is not disturbed by chance, thereby confirming the reliability and robustness of our contribution analysis conclusions. Following this validation, vegetation control, quantified by LAI, supersedes climatic factors (precipitation, solar radiation, temperature difference, and wind speed) in dominating rootzone soil moisture variations during winter and spring. This pattern is particularly pronounced in tropical savannas during winter dry seasons, where LAI-mediated moisture regulation becomes predominant

as precipitation-driven water recharge (via rainfall and surface runoff) approaches negligible levels. The rootzone's enhanced water retention capacity—attributable to greater depth and structural stability—renders it less vulnerable to short-term climatic variability compared to surface soil [74]. These properties establish rootzone soil moisture as the pivotal hydrological reservoir for vegetation, serving essential functions in drought resilience and physiological maintenance during water-limited periods [75–77]. Notably, vegetation exerts stronger nonlinear regulation on rootzone soil moisture than solar radiation's indirect effects through surface heating [78]. During spring, as temperatures rise, vegetation greening occurs across most global regions, with rapid plant growth remaining the primary factor controlling rootzone soil moisture in deeper soil layers. In summer, solar radiation dominates rootzone soil moisture variation due to maximum radiation intensity and prolonged sunshine duration, followed by precipitation as the secondary factor. Although solar radiation weakens in autumn, residual heat accumulated in soils during summer maintains relatively high temperatures [79], making solar radiation still the predominant influence on rootzone soil moisture. Wind speed plays a more dominant role in winter and spring than in summer and autumn, particularly across high-latitude Northern Hemisphere regions where Arctic cold air brings both snowfall and subfreezing temperatures that immobilize soil moisture through ground freezing.

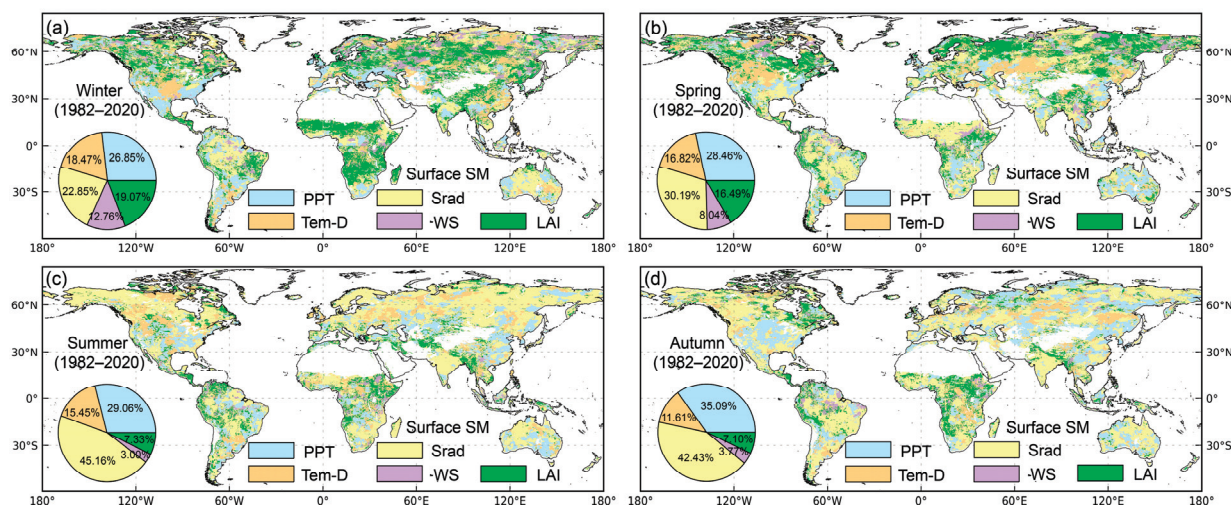


**Figure 13.** The spatial distribution of the dominant contributors to seasonal variations in rootzone soil moisture across the globe. Panels (a), (b), (c), and (d) show the results for winter, spring, summer, and autumn, respectively. PPT, Sradi, Tem-D, and WS denote precipitation, solar radiation, temperature difference, and wind speed, respectively. The pie charts illustrate the percentage of the total area dominated by each driving factor.

The dominant control of seasonal temperature differences on rootzone soil moisture dynamics remained remarkably stable, consistently accounting for 14–19% of the total contribution. A further comparison between the two periods (1982–1999 vs. 2000–2020) reveals a widespread increasing trend in the dominant contribution of LAI to rootzone soil moisture. Specifically, compared to the 1982–1999 period, the dominance of LAI in 2000–2020 increased by 10.47%, 6.65%, 10.37%, and 6.77% during winter, spring, summer, and autumn, respectively. This change can likely be attributed to the accelerated rate of vegetation greening since the 21st century [80,81], which has in turn enhanced the regulatory role of LAI on rootzone soil moisture.

Globally, precipitation is the most dominant seasonal driver of surface soil moisture in winter, followed by solar radiation, as the season's weaker solar radiation, lower temperatures, and reduced evaporation allow most of it to infiltrate the surface layer and become the key

regulator. From spring to autumn (particularly in summer and autumn), solar radiation surpasses precipitation as the dominant influence by more directly heating the topsoil and enhancing evaporation [82]. LAI maintains a consistent 16–20% dominance in controlling surface soil moisture during winter and spring, but declines to less than half this level in summer and autumn. Wind speed reaches its peak influence in winter, accounting for 12.76% of soil moisture regulation. Between 1982 and 1999 and 2000–2020, LAI's dominant control over surface soil moisture increased across all seasons, with winter exhibiting the most significant gain—an 9.10 percentage point increase relative to the 1982–1999 baseline.



**Figure 14.** The spatial distribution of the dominant contributors to seasonal variations in surface soil moisture across the globe. Panels (a), (b), (c), and (d) show the results for winter, spring, summer, and autumn, respectively. PPT, Srad, Tem-D, and WS denote precipitation, solar radiation, temperature difference, and wind speed, respectively. The pie charts illustrate the percentage of the total area dominated by each driving factor.

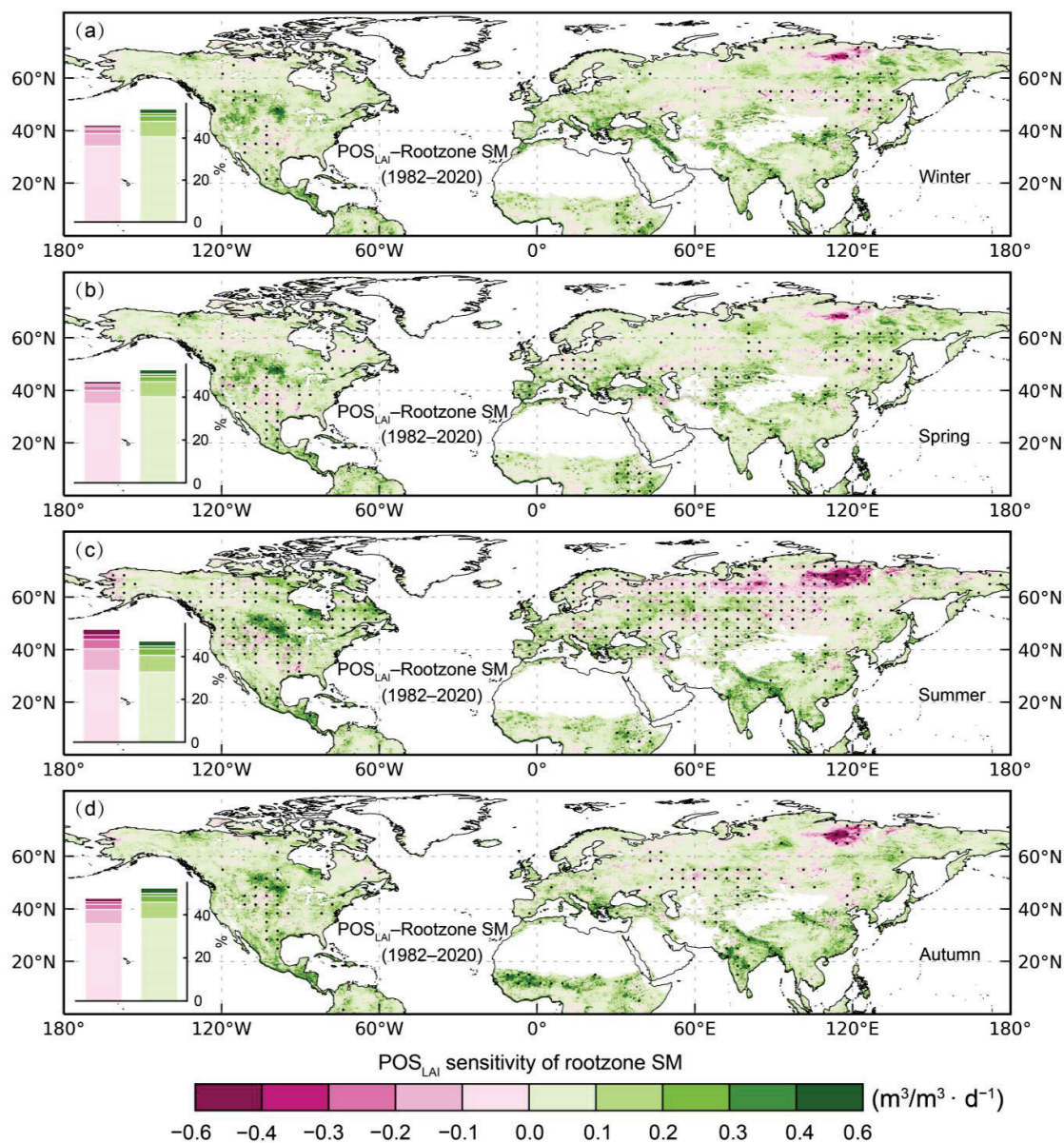
#### 4.2. The Relationships of POS and LAI<sub>MAX</sub> with Soil Moisture and Climatic Factors

In the Northern Hemisphere, the spring diurnal temperature range exhibits a high proportion of negative correlation with the POS for vegetation growth (Figure 12). On the one hand, a larger temperature range, accompanied by high daytime temperatures and strong radiation, significantly promotes the accumulation of effective accumulated temperature, enabling vegetation to reach the required thermal threshold for growth more quickly [83]. On the other hand, against the backdrop of global warming, cool nighttime temperatures effectively suppress plant respiratory consumption, thereby optimizing the carbon balance. This “warm day, cool night” climate pattern allocates more photosynthetic products to the rapid accumulation of biomass, thus powerfully advancing the POS [84]. Furthermore, high spring temperatures not only accelerate snowmelt but also promote the thawing of the permafrost active layer in high-latitude regions [85]. This process simultaneously releases valuable early-season water and nutrients for vegetation green-up, collectively contributing to the significant advancement of the growth peak.

The sensitivity and correlation of seasonal soil moisture to POS and LAI<sub>MAX</sub> from 1982 to 2020 were investigated. From 1982 to 2020, the sensitivity of rootzone and surface soil moisture to POS was mainly concentrated in the range of  $-0.1$  to  $0.1 \text{ m}^3/\text{m}^3\cdot\text{d}^{-1}$  (Figures 15 and S72). Northern Central Siberia showed consistently negative rootzone soil moisture sensitivity to POS throughout all seasons, with peak negative sensitivity coverage in summer. The high-latitude environment, characterized by light/temperature limitations and consequent condensed growing seasons, drives vegetation to intensively



deplete soil moisture during brief summer growth periods [86–88]. The sensitivity of root-zone and surface soil moisture to  $LAI_{MAX}$  was mainly concentrated in the range of  $-0.3$  to  $0.3 \text{ m}^3\text{m}^{-3}/\text{m}^2\text{m}^{-2}$  (Figures S71 and S73). A further comparison between two periods (1982–1999 vs. 2000–2020) reveals distinct patterns in soil moisture sensitivity. Across all seasons in the Northern Hemisphere, the spatial extent of positive sensitivity to POS in both rootzone and surface soil moisture significantly exceeded negative sensitivity during 2000–2020 (Figures S74, S76, S78 and S80). In contrast to the sensitivity to POS, the sensitivity to  $LAI_{MAX}$  showed a dominance of positive over negative responses in both periods, and this dominance was more pronounced after 2000 (Figures S75, S77, S79 and S81).



**Figure 15.** The spatial distribution of the sensitivity of rootzone soil moisture to POS in the Northern Hemisphere from 1982 to 2020. Panels (a), (b), (c), and (d) show the results for winter, spring, summer, and autumn, respectively. The histogram in the lower-left corner of each panel represents the frequency distribution of the corresponding sensitivity coefficient. Black dots indicate pixels that passed the significance test ( $p < 0.05$ ).

In the Northern Hemisphere, the analysis of correlations between POS and soil moisture (1982–2020) reveals spatially consistent patterns for both rootzone and surface layers (Figure 16 and Figure S91). Seasonal variations indicate that rootzone soil moisture displayed slightly higher proportions of negative than positive correlations with POS in winter and spring, while summer and autumn showed substantially dominant positive correlations. For surface soil moisture, winter exhibited nearly balanced positive/negative correlation proportions, whereas spring through autumn demonstrated markedly higher prevalence of positive correlations. During 1982–2020,  $LAI_{MAX}$  exhibited extensive significant negative correlations with summer and autumn rootzone and surface soil moisture across high-latitude regions, particularly in Eurasia, while mid-latitude areas showed predominantly positive correlations (Figures S90 and S92). This spatial pattern confirms that in high latitudes, vegetation's peak growth period coincides with substantial water demand and strong transpiration, where summer/autumn precipitation proves insufficient to compensate for combined vegetation and soil moisture losses. In contrast, mid-latitude regions demonstrate synchronized vegetation growth and soil moisture dynamics due to abundant warm-season precipitation. For winter and spring conditions, positive correlations between rootzone soil moisture and  $LAI_{MAX}$  dominated across most Northern Hemisphere regions, whereas summer and autumn displayed approximately balanced proportions of positive and negative correlation areas.

The spatiotemporal dynamics of soil moisture are profoundly reshaped by phenological shifts, which regulate vegetation transpiration and canopy interception of precipitation. Studies indicate that from 1982 to 2020, an earlier POS accelerated soil moisture depletion in tropical and temperate regions, exhibiting positive sensitivity coefficients that peaked during the tropical autumn ( $0.059 \text{ m}^3\text{m}^{-3} \text{ d}^{-1}$  and  $0.052 \text{ m}^3\text{m}^{-3} \text{ d}^{-1}$  for rootzone and surface soil moisture, respectively) (Figures S82–S85). In contrast, during certain seasons in arid and cold climate regions, an earlier POS contributed to increased soil moisture, manifesting as negative sensitivity, with the minimum value occurring in the cold climate summer ( $-0.0297 \text{ m}^3\text{m}^{-3} \text{ d}^{-1}$  and  $-0.0293 \text{ m}^3\text{m}^{-3} \text{ d}^{-1}$  for rootzone and surface, respectively). Concurrently, an increase in the  $LAI_{MAX}$  demonstrated a more ubiquitous positive effect. Except for summer in cold climate zones,  $LAI_{MAX}$  was significantly and positively correlated with soil moisture across all climatic zones, with the sensitivity reaching its maximum in the tropical autumn ( $0.256 (\text{m}^3/\text{m}^3)/(\text{m}^2/\text{m}^2)$  and  $0.232 (\text{m}^3/\text{m}^3)/(\text{m}^2/\text{m}^2)$  for rootzone and surface, respectively). Coupled with the global trend of vegetation greening (i.e., increasing  $LAI_{MAX}$ ), this reveals a critical mechanism: although a denser canopy enhances transpirational water loss, its “water conservation effect”—achieved by suppressing surface evaporation and improving soil infiltration—often dominates, thereby facilitating a large-scale increase in soil moisture across most regions globally.

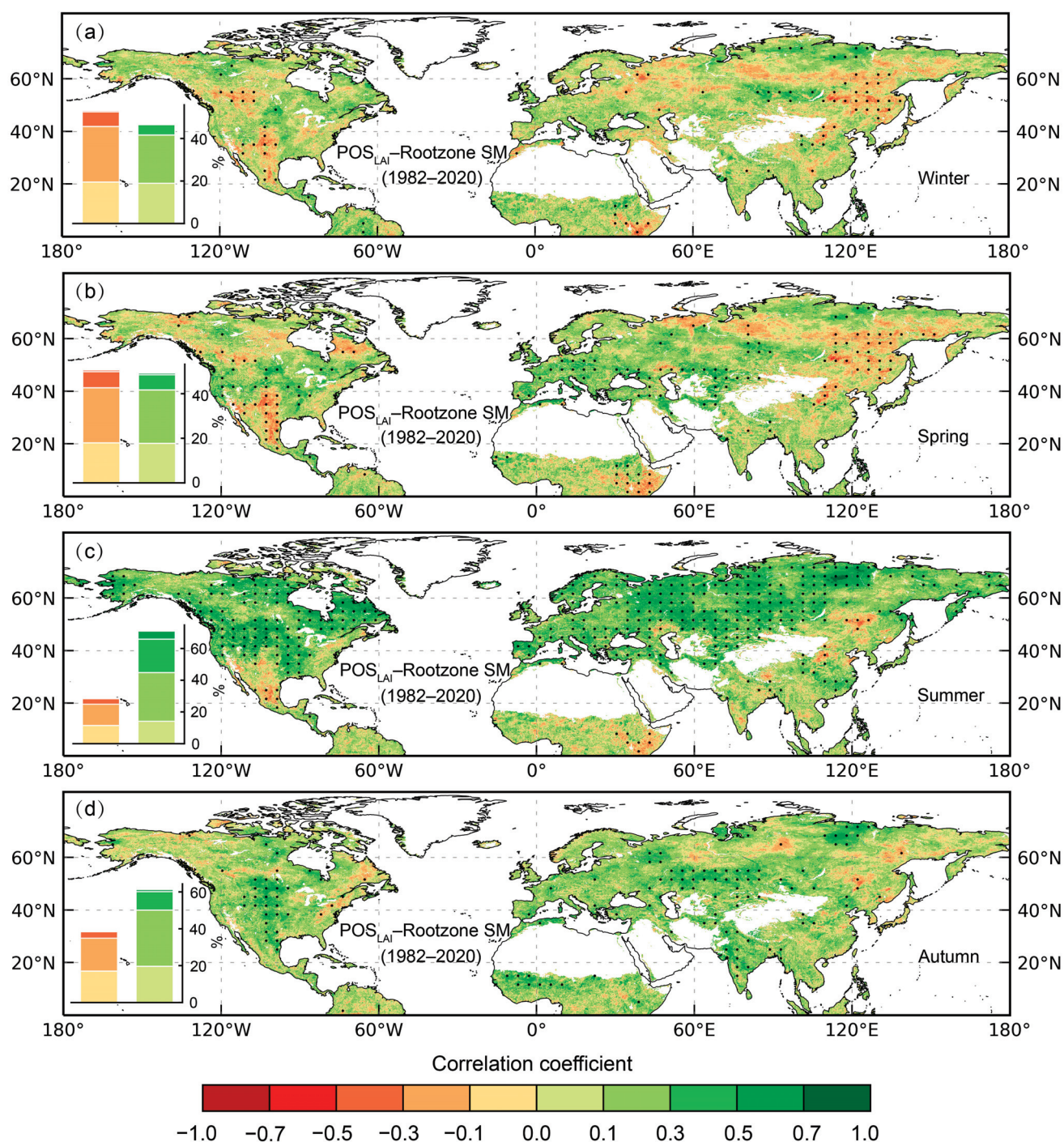
Amidst the ongoing trend of global greening, elevated atmospheric  $\text{CO}_2$  concentration is the dominant driver at a global scale [89]. The  $\text{CO}_2$  fertilization effect stimulates vegetation growth through two core mechanisms: first, by directly enhancing photosynthetic efficiency, as  $\text{CO}_2$  is a key raw material for photosynthesis [90]; and second, by improving vegetation water-use efficiency (WUE) through the regulation of stomatal conductance, which allows for less water consumption per unit of carbon fixed [91]. In addition, other anthropogenic activities play a vital role. For instance, increased nitrogen deposition provides additional nutrients to many nitrogen-limited ecosystems, thereby promoting vegetation growth [92]. More direct are the impacts of changes in land use and management practices. Studies have shown that the greening trends in China and India are especially prominent globally, largely attributable to China's large-scale afforestation programs and the intensive agricultural practices in both countries (e.g., multi-cropping, irrigation, and



fertilization) [93], which directly increase vegetation cover and productivity. The dominant limiting factors for vegetation growth vary by region. In the high-latitude regions of the Northern Hemisphere, temperature is the primary limiting factor [33,94]. Consequently, climate warming, which leads to an extended growing season and increased thermal resources, is the principal driver of greening in this area. In global arid and semi-arid regions, water is the absolute limiting factor [37]. Here, the extent of greening is co-determined by the enhanced WUE from the CO<sub>2</sub> fertilization effect and shifts in precipitation patterns [95]. In contrast, within intensively managed agricultural ecosystems, the influence of direct human interventions (such as irrigation and fertilization) often surpasses the background effects of climate change and elevated CO<sub>2</sub> [93]. Crucially, the same factors responsible for vegetation greening can also explain the variations in its resource use efficiency. It is probable that the combined influence of the CO<sub>2</sub> fertilization effect, climate warming, and nitrogen deposition has boosted the capacity of vegetation across mid-to-high Eurasian latitudes to utilize solar radiation, achieved through an increased light saturation point and the alleviation of environmental stress [96,97]. On a macroscopic scale, this improvement in physiological efficiency transcends the limitation of light saturation, which has led to the pronounced strengthening of the positive correlation between LAI<sub>MAX</sub> and solar radiation observed since the year 2000.

However, the anti-persistence revealed by the Hurst exponent ( $H < 0.5$ ) indicates a potential for future trend reversals, posing a formidable challenge to our previous findings and potentially triggering a “double blow” to soil moisture regulation. On one hand, should the LAI<sub>MAX</sub>-driven “global greening” stagnate or reverse, the ubiquitous soil “water conservation effect” observed across most regions would be severely weakened, likely exacerbating the risk of large-scale soil drought. On the other hand, future fluctuations in the POS would disrupt current regional hydrological patterns, introducing new and more unpredictable water stress, irrespective of whether regions currently benefit or suffer from an earlier POS.

This serves as a critical warning that a simple linear extrapolation of vegetation trends observed over the past few decades is highly unreliable, posing a fundamental challenge to both hydrological forecasting and ecosystem management. In the field of hydrological forecasting, long-term models that rely on historical greening trends for parameterization are prone to significant biases. Such models will systematically overestimate future evapotranspiration and underestimate available water resources by overestimating LAI, thereby misleading regional water resource planning. Notably, some studies have already recognized this, suggesting that the impact of future vegetation greening on evapotranspiration may diminish [98]. Furthermore, a potential reversal in the POS trend would alter the seasonal rhythm of peak vegetation water demand, further reducing the accuracy of seasonal drought predictions. From an ecosystem management perspective, this uncertainty is equally critical. First, it challenges climate change mitigation strategies by questioning the stability of the terrestrial carbon sink. A potential reversal in the LAI trend implies that the carbon sink, currently enhanced by “global greening,” is not permanently stable, introducing new uncertainties for achieving carbon neutrality targets. Therefore, the findings of this study call for a departure from the static mindset of trend extrapolation and a move towards more resilient and adaptive management frameworks capable of addressing the high uncertainty inherent in vegetation dynamics.



**Figure 16.** The spatial distribution of the correlation between rootzone soil moisture and POS in the Northern Hemisphere from 1982 to 2020. Panels (a), (b), (c), and (d) show the results for winter, spring, summer, and autumn, respectively. The histogram in the lower-left corner of each panel represents the frequency distribution of the corresponding correlation coefficient. Black dots indicate pixels that passed the significance test ( $p < 0.05$ ).

#### 4.3. Uncertainties and Limitations

The three reanalysis soil moisture datasets employed in this study—ERA5-Land, CFSR, and MERRA-2—exhibit good agreement with In Situ observations in many Regions [99–101]. However, their reliability faces significant challenges in densely vegetated areas, such as tropical rainforests, a limitation common to various remote sensing

technologies. Although optical remote sensing provides high-resolution surface information, its signals cannot penetrate the vegetation canopy and are susceptible to interference from clouds and rain, often resulting in data gaps [102]. In contrast, microwave remote sensing has become the mainstream approach for all-weather soil moisture monitoring, owing to its ability to penetrate clouds and, to some extent, vegetation. Nevertheless, even the more penetrative L-band signal experiences severe attenuation under dense vegetation cover. For instance, in regions like the Amazon rainforest, the scarcity of ground observations makes it challenging to accurately calibrate key parameters within radiative transfer models (e.g., vegetation water content and optical depth), thereby limiting the effectiveness of corrections for vegetation interference [103]. Furthermore, commonly used L-band and C-band microwaves can only sense the top ~5 cm of the soil, rendering them incapable of measuring deeper rootzone soil moisture [71].

Estimating rootzone soil moisture, whether through indirect retrievals from surface information or direct simulations by land surface models, is subject to significant uncertainties. The land surface models within the reanalysis datasets used in this study are constrained by the accuracy of their parameterization schemes and meteorological forcing data. This can lead to common systematic biases across multiple products, particularly in structurally complex forest ecosystems. Moreover, the ensemble averaging method employed by the dataset, while intended to integrate multi-source information, may mask these inter-product systematic biases and is susceptible to the influence of individual outlier products [49]. Additionally, the process of downscaling coarse-resolution reanalysis data to a uniform high resolution can introduce further smoothing errors, failing to fully reproduce true local-scale spatial heterogeneity [49]. Despite these prevalent limitations, we contend that the multi-source fusion soil moisture product selected for this study—owing to its global coverage, long-term time series (1982–2020), spatial continuity, and inclusion of both surface and rootzone layers—remains one of the most reliable and comprehensive data sources currently available for investigating large-scale vegetation-hydrology interactions, providing invaluable support for such scientific inquiries. Looking forward, P-band microwave remote sensing, with its superior penetration capability (up to 40 cm or deeper), presents a promising new avenue for directly sensing rootzone soil moisture while effectively mitigating interference from the vegetation canopy [70]. Future research should focus on integrating such novel remote sensing data with advanced algorithms and models, alongside standardized global In Situ observation networks, to produce higher-quality global soil moisture datasets.

## 5. Conclusions

In this study, the LAI dataset, meteorological dataset, and surface and rootzone soil moisture datasets were utilized to investigate the regulatory effects of vegetation and meteorological factors on rootzone and surface soil moisture across different seasons and climate zones. Additionally, the POS and LAI<sub>MAX</sub> were employed as indirect metrics to explore the feedback effects of vegetation phenology on soil moisture. The study found that since the beginning of the 21st century, global LAI has exhibited a significant and accelerated greening trend across all seasons. From 1982 to 2020, the temperate climate zone showed the strongest spring LAI trend among all climate zones, with a rate of  $0.05 \times 10^{-1} \text{ m}^2\text{m}^{-2}\cdot\text{yr}^{-1}$ . During the period from 2000 to 2020, the rate of increase in rootzone soil moisture was faster than that observed from 1982 to 1999. For surface soil moisture, a general decreasing trend prevailed across most global regions from 1982 to 1999, while during 2000–2020, the decline became more concentrated in high-latitude regions of the Northern Hemisphere. From 2000 to 2020, the increasing trends in rootzone



soil moisture and surface soil moisture were more pronounced in temperate and tropical climate zones compared to cold and arid zones. The change in soil moisture induced by LAI exhibited a wetting trend in the rootzone across all seasons from 1982 to 2020. The average wetting trend was highest in the temperate winter ( $0.13 \times 10^{-3} \text{ m}^3\text{m}^{-3}/\text{m}^2\text{m}^{-2}$ ) and lowest in the cold climate winter ( $0.02 \times 10^{-3} \text{ m}^3\text{m}^{-3}/\text{m}^2\text{m}^{-2}$ ). In cold climates, the precipitation-induced trend in rootzone soil moisture was positive throughout all seasons from 1982 to 2020. In contrast, solar radiation generally drove a drying trend across all climate zones and seasons. The temperature difference exerted its strongest drying effect in the temperate spring ( $-0.06 \times 10^{-3} \text{ m}^3\text{m}^{-3}/^\circ\text{C}$ ) and its strongest wetting effect in the cold climate winter ( $0.04 \times 10^{-3} \text{ m}^3\text{m}^{-3}/^\circ\text{C}$ ). Regarding surface soil moisture from 1982 to 2020, precipitation induced the strongest wetting trend in the temperate summer, at a rate of  $0.013 \times 10^{-3} \text{ m}^3\text{m}^{-3}/\text{mm}$ . The wetting trend driven by temperature difference was most significant in the arid spring, reaching  $0.024 \times 10^{-3} \text{ m}^3\text{m}^{-3}/^\circ\text{C}$ . Comparing the 1982–1999 and 2000–2020 periods, the dominant influence of LAI on global soil moisture has significantly strengthened. For rootzone soil moisture, this effect increased most notably in winter (+10.47%) and summer (+10.37%). For surface soil moisture, the increase was most prominent in winter, at 9.10 percentage points. From 1982 to 2020, the average sensitivity coefficients of both rootzone and surface soil moisture to POS were positive in temperate and tropical zones. The lowest sensitivity of rootzone and surface soil moisture to POS occurred in the cold-climate summer, with values of  $-0.0297 \text{ m}^3\text{m}^{-3}\cdot\text{d}^{-1}$  and  $-0.0293 \text{ m}^3\text{m}^{-3}\cdot\text{d}^{-1}$ , respectively. Regarding  $\text{LAI}_{\text{MAX}}$ , with the exception of summer in cold climates, the sensitivity coefficients for both soil layers were positive across all seasons in all climate zones. This study also investigated the spatiotemporal patterns of POS and  $\text{LAI}_{\text{MAX}}$  across the Northern Hemisphere, along with their responses to meteorological factors. The research provides a fundamental theoretical framework for predicting the spatiotemporal dynamics of soil moisture under global climate change and vegetation greening scenarios. It also deepens our understanding of global terrestrial ecohydrological processes.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs17223714/s1>.

**Author Contributions:** Conceptualization, H.Y. (Hanmin Yin), X.L. and Q.L.; methodology, H.Y. (Hanmin Yin), X.L. and Q.L.; software, H.Y. (Hanmin Yin), Q.L., W.Y., Y.L., J.W. and J.Y.; resources, J.B. and Q.L.; data curation, H.Y. (Hanmin Yin); writing—original draft preparation, H.Y. (Hanmin Yin); writing—review and editing, H.Y. (Hanmin Yin) and Q.L.; supervision, X.L. and H.Y. (Huping Ye); project administration, Q.L.; funding acquisition, Q.L. and J.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Major Key Project of PCL and the National Natural Science Foundation of China [Grant No. 41871226].

**Data Availability Statement:** The rootzone and surface soil moisture datasets used in this study are publicly available from the National Earth System Science Data Center at [https://www.geodata.cn/main/face\\_scientist](https://www.geodata.cn/main/face_scientist) (accessed on 1 May 2025). The LAI dataset is available in the Zenodo repository at <https://doi.org/10.5281/zenodo.7649107> (accessed on 1 May 2025).

**Acknowledgments:** We gratefully acknowledge the following data sources that were essential to this study: the rootzone and surface soil moisture datasets from the National Earth System Science Data Center, the monthly maximum/minimum temperature, precipitation, solar radiation, and wind speed data from TerraClimate, the GIMMS LAI4g product, and the updated Köppen–Geiger climate classification maps. Special thanks are extended to NASA’s Land Processes Distributed Active



Archive Center (LPDAAC) for providing the MCD12C1 product. These high-quality datasets formed the foundation of our research, and we sincerely appreciate the support from these organizations.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Meng, F.; Luo, M.; Sa, C.; Wang, M.; Bao, Y. Quantitative assessment of the effects of climate, vegetation, soil and groundwater on soil moisture spatiotemporal variability in the Mongolian Plateau. *Sci. Total Environ.* **2022**, *809*, 152198. [CrossRef]
2. Tian, J.; Zhang, B.; He, C.; Han, Z.; Bogena, H.R.; Huisman, J.A. Dynamic response patterns of profile soil moisture wetting events under different land covers in the Mountainous area of the Heihe River Watershed, Northwest China. *Agric. For. Meteorol.* **2019**, *271*, 225–239. [CrossRef]
3. Li, W.; Duveiller, G.; Wieneke, S.; Forkel, M.; Gentine, P.; Reichstein, M.; Niu, S.; Migliavacca, M.; Orth, R. Regulation of the global carbon and water cycles through vegetation structural and physiological dynamics. *Environ. Res. Lett.* **2024**, *19*, 073008. [CrossRef]
4. Fang, Z.; Zhang, W.; Brandt, M.; Abdi, A.M.; Fensholt, R. Globally Increasing Atmospheric Aridity Over the 21st Century. *Earth's Future* **2022**, *10*, e2022EF003019. [CrossRef]
5. Zeng, Z.; Wu, W.; Peñuelas, J.; Li, Y.; Jiao, W.; Li, Z.; Ren, X.; Wang, K.; Ge, Q. Increased risk of flash droughts with raised concurrent hot and dry extremes under global warming. *npj Clim. Atmos. Sci.* **2023**, *6*, 134. [CrossRef]
6. Jin, X.; Jin, H.; Luo, D.; Sheng, Y.; Wu, Q.; Wu, J.; Wang, W.; Huang, S.; Li, X.; Liang, S.; et al. Impacts of Permafrost Degradation on Hydrology and Vegetation in the Source Area of the Yellow River on Northeastern Qinghai-Tibet Plateau, Southwest China. *Front. Earth Sci.* **2022**, *10*, 845824. [CrossRef]
7. Jiao, W.; Wang, L.; Smith, W.K.; Chang, Q.; Wang, H.; D'Odorico, P. Observed increasing water constraint on vegetation growth over the last three decades. *Nat. Commun.* **2021**, *12*, 3777. [CrossRef]
8. Berg, A.; Sheffield, J.; Milly, P.C.D. Divergent surface and total soil moisture projections under global warming. *Geophys. Res. Lett.* **2017**, *44*, 236–244. [CrossRef]
9. Zhang, Y.; Parazoo, N.C.; Williams, A.P.; Zhou, S.; Gentine, P. Large and projected strengthening moisture limitation on end-of-season photosynthesis. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9216–9222. [CrossRef] [PubMed]
10. Yuan, W.; Zheng, Y.; Piao, S.; Ciais, P.; Lombardozzi, D.; Wang, Y.; Ryu, Y.; Chen, G.; Dong, W.; Hu, Z.; et al. Increased atmospheric vapor pressure deficit reduces global vegetation growth. *Sci. Adv.* **2019**, *5*, eaax1396. [CrossRef]
11. Konings, A.G.; Williams, A.P.; Gentine, P. Sensitivity of grassland productivity to aridity controlled by stomatal and xylem regulation. *Nat. Geosci.* **2017**, *10*, 284–288. [CrossRef]
12. Bai, P.; Liu, X.; Zhang, Y.; Liu, C. Assessing the Impacts of Vegetation Greenness Change on Evapotranspiration and Water Yield in China. *Water Resour. Res.* **2020**, *56*, e2019WR027019. [CrossRef]
13. Igwe, P.; Ezeukwu, J.; Edoka, N.; Ejie, O.; Ifi, G. A review of vegetation cover as a natural factor to soil erosion. *Int. J. Rural Dev. Environ. Health Res.* **2017**, *1*, 21–28.
14. Davarzani, H.; Smits, K.; Tolene, R.M.; Illangasekare, T. Study of the effect of wind speed on evaporation from soil through integrated modeling of the atmospheric boundary layer and shallow subsurface. *Water Resour. Res.* **2014**, *50*, 661–680. [CrossRef] [PubMed]
15. Wang, Y.; Yang, J.; Chen, Y.; Wang, A.; De Maeyer, P. The Spatiotemporal Response of Soil Moisture to Precipitation and Temperature Changes in an Arid Region, China. *Remote Sens.* **2018**, *10*, 468. [CrossRef]
16. Zhang, W.; Wei, F.; Horion, S.; Fensholt, R.; Forkel, M.; Brandt, M. Global quantification of the bidirectional dependency between soil moisture and vegetation productivity. *Agric. For. Meteorol.* **2022**, *313*, 108735. [CrossRef]
17. Jiao, Z.; Shen, Z. School choice with priority-based affirmative action: A responsive solution. *J. Math. Econ.* **2021**, *92*, 1–9. [CrossRef]
18. Xu, Z.-X.; Li, L.; Zhao, J. A distributed eco-hydrological model and its application. *Water Sci. Eng.* **2017**, *10*, 257–264. [CrossRef]
19. Yu, L.; Liu, Y.; Li, X.; Yan, F.; Lyne, V.; Liu, T. Vegetation-induced asymmetric diurnal land surface temperatures changes across global climate zones. *Sci. Total Environ.* **2023**, *896*, 165255. [CrossRef]
20. Sun, S.; Liu, Y.; Chen, H.; Ju, W.; Xu, C.-Y.; Liu, Y.; Zhou, B.; Zhou, Y.; Zhou, Y.; Yu, M. Causes for the increases in both evapotranspiration and water yield over vegetated mainland China during the last two decades. *Agric. For. Meteorol.* **2022**, *324*, 109118. [CrossRef]
21. Li, W.; Migliavacca, M.; Forkel, M.; Denissen, J.M.C.; Reichstein, M.; Yang, H.; Duveiller, G.; Weber, U.; Orth, R. Widespread increasing vegetation sensitivity to soil moisture. *Nat. Commun.* **2022**, *13*, 3959. [CrossRef]
22. Gonsamo, A.; Chen, J.M.; Ooi, Y.W. Peak season plant activity shift towards spring is reflected by increasing carbon uptake by extratropical ecosystems. *Glob. Change Biol.* **2018**, *24*, 2117–2128. [CrossRef]

23. Liu, Z.; Fu, Y.H.; Chen, D.; Crowther, T.W.; Ding, R.; Hao, F.; Vitasse, Y.; Zhang, X.; Zhou, S.; Zohner, C.M. Increased early-season productivity drives earlier peak of vegetation photosynthesis across the Northern Hemisphere. *Commun. Earth Environ.* **2025**, *6*, 157. [CrossRef]
24. Xia, J.; Niu, S.; Ciais, P.; Janssens, I.A.; Chen, J.; Ammann, C.; Arain, A.; Blanken, P.D.; Cescatti, A.; Bonal, D.; et al. Joint control of terrestrial gross primary productivity by plant phenology and physiology. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 2788–2793. [CrossRef]
25. Huang, Z.; Zhou, L.; Chi, Y. Spring phenology rather than climate dominates the trends in peak of growing season in the Northern Hemisphere. *Glob. Change Biol.* **2023**, *29*, 4543–4555. [CrossRef] [PubMed]
26. Yin, H.; Ma, X.; Liao, X.; Ye, H.; Yu, W.; Li, Y.; Wei, J.; Yuan, J.; Liu, Q. Linking Vegetation Phenology to Net Ecosystem Productivity: Climate Change Impacts in the Northern Hemisphere Using Satellite Data. *Remote Sens.* **2024**, *16*, 4101. [CrossRef]
27. Luo, Y.; Zhang, Z.; Chen, Y.; Li, Z.; Tao, F. ChinaCropPhen1km: A high-resolution crop phenological dataset for three staple crops in China during 2000–2015 based on leaf area index (LAI) products. *Earth Syst. Sci. Data* **2020**, *12*, 197–214. [CrossRef]
28. Zhang, L.; Lei, H.; Shen, H.; Cong, Z.; Yang, D.; Liu, T. Evaluating the Representation of Vegetation Phenology in the Community Land Model 4.5 in a Temperate Grassland. *J. Geophys. Res. Biogeosci.* **2019**, *124*, 187–210. [CrossRef]
29. Xu, X.; Du, H.; Fan, W.; Hu, J.; Mao, F.; Dong, H. Long-term trend in vegetation gross primary production, phenology and their relationships inferred from the FLUXNET data. *J. Environ. Manag.* **2019**, *246*, 605–616. [CrossRef]
30. Park, T.; Chen, C.; Macias-Fauria, M.; Tømmervik, H.; Choi, S.; Winkler, A.; Bhatt, U.S.; Walker, D.A.; Piao, S.; Brovkin, V.; et al. Changes in timing of seasonal peak photosynthetic activity in northern ecosystems. *Glob. Change Biol.* **2019**, *25*, 2382–2395. [CrossRef]
31. Xu, C.; McDowell, N.G.; Fisher, R.A.; Wei, L.; Sevanto, S.; Christoffersen, B.O.; Weng, E.; Middleton, R.S. Increasing impacts of extreme droughts on vegetation productivity under climate change. *Nat. Clim. Change* **2019**, *9*, 948–953. [CrossRef]
32. Bonfils, C.J.W.; Phillips, T.J.; Lawrence, D.M.; Cameron-Smith, P.; Riley, W.J.; Subin, Z.M. On the influence of shrub height and expansion on northern high latitude climate. *Environ. Res. Lett.* **2012**, *7*, 015503. [CrossRef]
33. Li, G.; Wu, C.; Chen, Y.; Huang, C.; Zhao, Y.; Wang, Y.; Ma, M.; Ding, Z.; Yu, P.; Tang, X. Increasing temperature regulates the advance of peak photosynthesis timing in the boreal ecosystem. *Sci. Total Environ.* **2023**, *882*, 163587. [CrossRef] [PubMed]
34. Luo, M.; Meng, F.; Sa, C.; Duan, Y.; Bao, Y.; Liu, T.; De Maeyer, P. Response of vegetation phenology to soil moisture dynamics in the Mongolian Plateau. *CATENA* **2021**, *206*, 105505. [CrossRef]
35. Wang, X.; Wu, C. Estimating the peak of growing season (POS) of China's terrestrial ecosystems. *Agric. For. Meteorol.* **2019**, *278*, 107639. [CrossRef]
36. Liu, Y.; Wu, C.; Wang, X.; Jassal, R.S.; Gonsamo, A. Impacts of global change on peak vegetation growth and its timing in terrestrial ecosystems of the continental US. *Glob. Planet. Change* **2021**, *207*, 103657. [CrossRef]
37. Gao, X.; Zhao, D. Impacts of climate change on vegetation phenology over the Great Lakes Region of Central Asia from 1982 to 2014. *Sci. Total Environ.* **2022**, *845*, 157227. [CrossRef]
38. Li, P.; Peng, C.; Wang, M.; Luo, Y.; Li, M.; Zhang, K.; Zhang, D.; Zhu, Q. Dynamics of vegetation autumn phenology and its response to multiple environmental factors from 1982 to 2012 on Qinghai-Tibetan Plateau in China. *Sci. Total Environ.* **2018**, *637–638*, 855–864. [CrossRef]
39. Lian, X.; Piao, S.; Li, L.Z.X.; Li, Y.; Huntingford, C.; Ciais, P.; Cescatti, A.; Janssens, I.A.; Peñuelas, J.; Buermann, W.; et al. Summer soil drying exacerbated by earlier spring greening of northern vegetation. *Sci. Adv.* **2020**, *6*, eaax0255. [CrossRef] [PubMed]
40. Piao, S.; Yin, G.; Tan, J.; Cheng, L.; Huang, M.; Li, Y.; Liu, R.; Mao, J.; Myneni, R.B.; Peng, S.; et al. Detection and attribution of vegetation greening trend in China over the last 30 years. *Glob. Change Biol.* **2015**, *21*, 1601–1609. [CrossRef]
41. Hu, Y.; Zhou, J.; Deng, J.; Li, Y.; Yang, C.; Li, D. River Bars and Vegetation Dynamics in Response to Upstream Damming: A Case Study of the Middle Yangtze River. *Remote Sens.* **2023**, *15*, 2324. [CrossRef]
42. Xu, T.; Yan, K.; He, Y.; Gao, S.; Yang, K.; Wang, J.; Liu, J.; Liu, Z. Spatio-Temporal Variability Analysis of Vegetation Dynamics in China from 2000 to 2022 Based on Leaf Area Index: A Multi-Temporal Image Classification Perspective. *Remote Sens.* **2023**, *15*, 2975. [CrossRef]
43. Cao, S.; Li, M.; Zhu, Z.; Wang, Z.; Zha, J.; Zhao, W.; Duanmu, Z.; Chen, J.; Zheng, Y.; Chen, Y.; et al. Spatiotemporally consistent global dataset of the GIMMS leaf area index (GIMMS LAI4g) from 1982 to 2020. *Earth Syst. Sci. Data* **2023**, *15*, 4877–4899. [CrossRef]
44. Chen, Y.; Zhu, Z.; Zhao, W.; Li, M.; Cao, S.; Zheng, Y.; Tian, F.; Myneni, R.B. The direct and indirect effects of the environmental factors on global terrestrial gross primary productivity over the past four decades. *Environ. Res. Lett.* **2024**, *19*, 014052. [CrossRef]
45. Han, X.; Chen, Y.; Fang, G.; Li, Z.; Li, Y.; Di, Y. Spatiotemporal Variations and Driving Factors of Water Availability in the Arid and Semiarid Regions of Northern China. *Remote Sens.* **2024**, *16*, 4318. [CrossRef]

46. Liao, J.; Xu, Y.; Pi, J.; Li, Y.; Ke, C.; Zhan, W.; Chen, J. Widespread Sensitivity of Vegetation to the Transition From Normal Droughts to Flash Droughts. *Geophys. Res. Lett.* **2025**, *52*, e2024GL114321. [CrossRef]
47. Tang, J.; Niu, B.; Fu, G.; Peng, J.; Hu, Z.; Zhang, X. Shifted trend in drought sensitivity of vegetation productivity from 1982 to 2020. *Agric. For. Meteorol.* **2025**, *362*, 110388. [CrossRef]
48. Li, J.; Bevacqua, E.; Chen, C.; Wang, Z.; Chen, X.; Myneni, R.B.; Wu, X.; Xu, C.-Y.; Zhang, Z.; Zscheischler, J. Regional asymmetry in the response of global vegetation growth to springtime compound climate events. *Commun. Earth Environ.* **2022**, *3*, 123. [CrossRef]
49. Liu, Y.; Yang, Y.; Song, J. Variations in Global Soil Moisture During the Past Decades: Climate or Human Causes? *Water Resour. Res.* **2023**, *59*, e2023WR034915. [CrossRef]
50. Zhu, Y.; Jiang, S.; Ren, L.; Guo, J.; Zhong, F.; Du, S.; Cui, H.; He, M.; Duan, Z. Three-dimensional ecological drought identification and evaluation method considering eco-physiological status of terrestrial ecosystems. *Sci. Total Environ.* **2024**, *951*, 175423. [CrossRef]
51. Bai, X.; Zhang, Z.; Gu, D. Driving mechanism of natural vegetation response to climate change in China from 2001 to 2022. *Environ. Res.* **2025**, *276*, 121529. [CrossRef]
52. Beck, H.E.; Zimmermann, N.E.; McVicar, T.R.; Vergopolan, N.; Berg, A.; Wood, E.F. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data* **2018**, *5*, 180214. [CrossRef]
53. Forzieri, G.; Alkama, R.; Miralles, D.G.; Cescatti, A. Satellites reveal contrasting responses of regional climate to the widespread greening of Earth. *Science* **2017**, *356*, 1180–1184. [CrossRef] [PubMed]
54. Abeyesiriwardana, H.D.; Gomes, P.I.A. Integrating vegetation indices and geo-environmental factors in GIS-based landslide-susceptibility mapping: Using logistic regression. *J. Mt. Sci.* **2022**, *19*, 477–492. [CrossRef]
55. Midi, H.; Sarkar, S.K.; Rana, S. Collinearity diagnostics of binary logistic regression model. *J. Interdiscip. Math.* **2010**, *13*, 253–267. [CrossRef]
56. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer Texts in Statistics; Springer: New York, NY, USA, 2013; Volume 103, pp. 971–978.
57. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
58. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777.
59. Jabeur, S.B.; Mefteh-Wali, S.; Viviani, J.-L. Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Ann. Oper. Res.* **2024**, *334*, 679–699. [CrossRef]
60. Zhang, W.; Fan, Z.; Jin, C.; Jiao, Y.; Di, K.; Feng, M.; Lu, Y.; Zhao, K.; Zhao, H.; Hao, S.; et al. Reversal of the sensitivity of vegetation productivity to precipitation in global terrestrial biomes over the recent decade. *Agric. For. Meteorol.* **2025**, *370*, 110598. [CrossRef]
61. Bai, T.; Wang, X.-S.; Han, P.-F. Controls of groundwater-dependent vegetation coverage in the yellow river basin, China: Insights from interpretable machine learning. *J. Hydrol.* **2024**, *631*, 130747. [CrossRef]
62. Chen, H.; Liu, J.; He, W.; Xu, P.; Nguyen, N.T.; Lv, Y.; Huang, C. Shifted vegetation resilience from loss to gain driven by changes in water availability and solar radiation over the last two decades in Southwest China. *Agric. For. Meteorol.* **2025**, *368*, 110543. [CrossRef]
63. Yin, H.; Liu, Q.; Liao, X.; Ye, H.; Li, Y.; Ma, X. Refined Analysis of Vegetation Phenology Changes and Driving Forces in High Latitude Altitude Regions of the Northern Hemisphere: Insights from High Temporal Resolution MODIS Products. *Remote Sens.* **2024**, *16*, 1744. [CrossRef]
64. Chen, A.; Meng, F.; Mao, J.; Ricciuto, D.; Knapp, A.K. Photosynthesis phenology, as defined by solar-induced chlorophyll fluorescence, is overestimated by vegetation indices in the extratropical Northern Hemisphere. *Agric. For. Meteorol.* **2022**, *323*, 109027. [CrossRef]
65. Jiang, L.; Bao, A.; Guo, H.; Ndayisaba, F. Vegetation dynamics and responses to climate change and human activities in Central Asia. *Sci. Total Environ.* **2017**, *599*, 967–980. [CrossRef] [PubMed]
66. Bashir, B.; Cao, C.; Naeem, S.; Zamani Joharestani, M.; Bo, X.; Afzal, H.; Jamal, K.; Mumtaz, F. Spatio-Temporal Vegetation Dynamic and Persistence under Climatic and Anthropogenic Factors. *Remote Sens.* **2020**, *12*, 2612. [CrossRef]
67. Chang, J.; Liu, Q.; Wang, S.; Huang, C. Vegetation Dynamics and Their Influencing Factors in China from 1998 to 2019. *Remote Sens.* **2022**, *14*, 3390. [CrossRef]
68. Hirschi, M.; Mueller, B.; Dorigo, W.; Seneviratne, S.I. Using remotely sensed soil moisture for land–atmosphere coupling diagnostics: The role of surface vs. root-zone soil moisture variability. *Remote Sens. Environ.* **2014**, *154*, 246–252. [CrossRef]
69. Akbar, R.; Short Gianotti, D.; McColl, K.A.; Haghighi, E.; Salvucci, G.D.; Entekhabi, D. Hydrological Storage Length Scales Represented by Remote Sensing Estimates of Soil Moisture and Precipitation. *Water Resour. Res.* **2018**, *54*, 1476–1492. [CrossRef]

70. Li, M.; Sun, H.; Zhao, R. A Review of Root Zone Soil Moisture Estimation Methods Based on Remote Sensing. *Remote Sens.* **2023**, *15*, 5361. [CrossRef]
71. Feldman, A.F.; Short Gianotti, D.J.; Dong, J.; Akbar, R.; Crow, W.T.; McColl, K.A.; Konings, A.G.; Nippert, J.B.; Tumber-Dávila, S.J.; Holbrook, N.M.; et al. Remotely Sensed Soil Moisture Can Capture Dynamics Relevant to Plant Water Uptake. *Water Resour. Res.* **2023**, *59*, e2022WR033814. [CrossRef]
72. Wei, L.; Qiu, Z.; Zhou, G.; Zuecco, G.; Liu, Y.; Wen, Y. Soil water hydraulic redistribution in a subtropical monsoon evergreen forest. *Sci. Total Environ.* **2022**, *835*, 155437. [CrossRef] [PubMed]
73. Moscati, M.D.L.; Gan, M.A. Rainfall variability in the rainy season of semiarid zone of Northeast Brazil (NEB) and its relation to wind regime. *Int. J. Climatol.* **2007**, *27*, 493–512. [CrossRef]
74. Li, B.-B.; Li, P.-P.; Zhang, W.-T.; Ji, J.-Y.; Liu, G.-B.; Xu, M.-X. Deep soil moisture limits the sustainable vegetation restoration in arid and semi-arid Loess Plateau. *Geoderma* **2021**, *399*, 115122. [CrossRef]
75. Ansley, R.J.; Boutton, T.W.; Jacoby, P.W. Root Biomass and Distribution Patterns in a Semi-Arid Mesquite Savanna: Responses to Long-Term Rainfall Manipulation. *Rangel. Ecol. Manag.* **2014**, *67*, 206–218. [CrossRef]
76. Su, B.; Shangguan, Z. Decline in soil moisture due to vegetation restoration on the Loess Plateau of China. *Land Degrad. Dev.* **2019**, *30*, 290–299. [CrossRef]
77. Wang, Y.; Shao, M.a.; Liu, Z. Large-scale spatial variability of dried soil layers and related factors across the entire Loess Plateau of China. *Geoderma* **2010**, *159*, 99–108. [CrossRef]
78. Ruichen, M.; Jinxi, S.; Bin, T.; Wenjin, X.; Feihe, K.; Haotian, S.; Yuxin, L. Vegetation variation regulates soil moisture sensitivity to climate change on the Loess Plateau. *J. Hydrol.* **2023**, *617*, 128763. [CrossRef]
79. Gao, S.; Chen, Y.; Li, K.; He, B.; Hou, P.; Guo, Z. Frequent heatwaves limit the indirect growth effect of urban vegetation in China. *Sustain. Cities Soc.* **2023**, *96*, 104662. [CrossRef]
80. Cortés, J.; Mahecha, M.D.; Reichstein, M.; Myneni, R.B.; Chen, C.; Brenning, A. Where Are Global Vegetation Greening and Browning Trends Significant? *Geophys. Res. Lett.* **2021**, *48*, e2020GL091496. [CrossRef]
81. Zhu, Z.; Piao, S.; Myneni, R.B.; Huang, M.; Zeng, Z.; Canadell, J.G.; Ciais, P.; Sitch, S.; Friedlingstein, P.; Arneeth, A.; et al. Greening of the Earth and its drivers. *Nat. Clim. Change* **2016**, *6*, 791–795. [CrossRef]
82. Khamidov, M.; Ishchanov, J.; Hamidov, A.; Shermatov, E.; Gafurov, Z. Impact of Soil Surface Temperature on Changes in the Groundwater Level. *Water* **2023**, *15*, 3865. [CrossRef]
83. Piao, S.; Tan, J.; Chen, A.; Fu, Y.H.; Ciais, P.; Liu, Q.; Janssens, I.A.; Vicca, S.; Zeng, Z.; Jeong, S.-J.; et al. Leaf onset in the northern hemisphere triggered by daytime temperature. *Nat. Commun.* **2015**, *6*, 6911. [CrossRef] [PubMed]
84. Peng, S.; Piao, S.; Ciais, P.; Myneni, R.B.; Chen, A.; Chevallier, F.; Dolman, A.J.; Janssens, I.A.; Peñuelas, J.; Zhang, G.; et al. Asymmetric effects of daytime and night-time warming on Northern Hemisphere vegetation. *Nature* **2013**, *501*, 88–92. [CrossRef]
85. Smith, S.L.; O'Neill, H.B.; Isaksen, K.; Noetzli, J.; Romanovsky, V.E. The changing thermal state of permafrost. *Nat. Rev. Earth Environ.* **2022**, *3*, 10–23. [CrossRef]
86. Rina, W.; Bao, G.; Hai, Q.; Chen, J.; Guo, E.; Li, F.; Bao, Y.; Miao, L.; Huang, X. Parallel acceleration of vegetation growth rate and senescence rate across the Northern Hemisphere from 1982 to 2015. *Glob. Ecol. Conserv.* **2023**, *46*, e02622. [CrossRef]
87. Flynn, D.F.B.; Wolkovich, E.M. Temperature and photoperiod drive spring phenology across all species in a temperate forest community. *New Phytol.* **2018**, *219*, 1353–1362. [CrossRef]
88. Vitasse, Y.; Signarbieux, C.; Fu, Y.H. Global warming leads to more uniform spring phenology across elevations. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 1004–1008. [CrossRef]
89. Piao, S.; Wang, X.; Park, T.; Chen, C.; Lian, X.; He, Y.; Bjerke, J.W.; Chen, A.; Ciais, P.; Tømmervik, H.; et al. Characteristics, drivers and feedbacks of global greening. *Nat. Rev. Earth Environ.* **2020**, *1*, 14–27. [CrossRef]
90. Farquhar, G.D.; Sharkey, T.D. Stomatal conductance and photosynthesis. *Annu. Rev. Plant Physiol.* **1982**, *33*, 317–345. [CrossRef]
91. Keenan, T.F.; Hollinger, D.Y.; Bohrer, G.; Dragoni, D.; Munger, J.W.; Schmid, H.P.; Richardson, A.D. Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise. *Nature* **2013**, *499*, 324–327. [CrossRef] [PubMed]
92. Greaver, T.L.; Clark, C.M.; Compton, J.E.; Vallano, D.; Talhelm, A.F.; Weaver, C.P.; Band, L.E.; Baron, J.S.; Davidson, E.A.; Tague, C.L.; et al. Key ecological responses to nitrogen are altered by climate change. *Nat. Clim. Change* **2016**, *6*, 836–843. [CrossRef]
93. Chen, C.; Park, T.; Wang, X.; Piao, S.; Xu, B.; Chaturvedi, R.K.; Fuchs, R.; Brovkin, V.; Ciais, P.; Fensholt, R.; et al. China and India lead in greening of the world through land-use management. *Nat. Sustain.* **2019**, *2*, 122–129. [CrossRef]
94. Piao, S.; Friedlingstein, P.; Ciais, P.; Viovy, N.; Demarty, J. Growing season extension and its impact on terrestrial carbon cycle in the Northern Hemisphere over the past 2 decades. *Glob. Biogeochem. Cycles* **2007**, *21*. [CrossRef]
95. Donohue, R.J.; Roderick, M.L.; McVicar, T.R.; Farquhar, G.D. Impact of CO<sub>2</sub> fertilization on maximum foliage cover across the globe's warm, arid environments. *Geophys. Res. Lett.* **2013**, *40*, 3031–3035. [CrossRef]



96. Ainsworth, E.A.; Long, S.P. What have we learned from 15 years of free-air CO<sub>2</sub> enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO<sub>2</sub>. *New Phytol.* **2005**, *165*, 351–372. [CrossRef]
97. Hussain, S.; Ulhassan, Z.; Brestic, M.; Zivcak, M.; Weijun, Z.; Allakhverdiev, S.I.; Yang, X.; Safdar, M.E.; Yang, W.; Liu, W. Photosynthesis research under climate change. *Photosynth. Res.* **2021**, *150*, 5–19. [CrossRef]
98. Yang, Y.; Roderick, M.L.; Guo, H.; Miralles, D.G.; Zhang, L.; Fatichi, S.; Luo, X.; Zhang, Y.; McVicar, T.R.; Tu, Z.; et al. Evapotranspiration on a greening Earth. *Nat. Rev. Earth Environ.* **2023**, *4*, 626–641. [CrossRef]
99. Beck, H.E.; Pan, M.; Miralles, D.G.; Reichle, R.H.; Dorigo, W.A.; Hahn, S.; Sheffield, J.; Karthikeyan, L.; Balsamo, G.; Parinussa, R.M.; et al. Evaluation of 18 satellite- and model-based soil moisture products using in situ measurements from 826 sensors. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 17–40. [CrossRef]
100. Reichle, R.H.; Draper, C.S.; Liu, Q.; Girotto, M.; Mahanama, S.P.P.; Koster, R.D.; De Lannoy, G.J.M. Assessment of MERRA-2 Land Surface Hydrology Estimates. *J. Clim.* **2017**, *30*, 2937–2960. [CrossRef]
101. Li, M.; Wu, P.; Ma, Z. A comprehensive evaluation of soil moisture and soil temperature from third-generation atmospheric and land reanalysis data sets. *Int. J. Climatol.* **2020**, *40*, 5744–5766. [CrossRef]
102. Zhang, D.; Zhou, G. Estimation of Soil Moisture from Optical and Thermal Remote Sensing: A Review. *Sensors* **2016**, *16*, 1308. [CrossRef] [PubMed]
103. Cho, K.; Negrón-Juárez, R.; Colliander, A.; Cosio, E.G.; Salinas, N.; de Araujo, A.; Chambers, J.Q.; Wang, J. Calibration of the SMAP Soil Moisture Retrieval Algorithm to Reduce Bias Over the Amazon Rainforest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 8724–8736. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## Article

# Impact Analysis of Climate Change on Hydropower Resource Development in the Vakhsh River Basin of Tajikistan

Hailong Liu <sup>1,2,\*</sup>, Aminjon Gulakhmadov <sup>3</sup> and Firdavs Shaimuradov <sup>3</sup>

<sup>1</sup> Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China

<sup>2</sup> School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China

<sup>3</sup> Institute of Water Problems, Hydropower and Ecology of the National Academy of Sciences of Tajikistan, Dushanbe 734042, Tajikistan; aminjon@ms.xjb.ac.cn (A.G.); sh.firdavs-80@mail.ru (F.S.)

\* Correspondence: liuhl@uestc.edu.cn

**Abstract:** With increasing energy demands and environmental pressures, hydropower, as a clean and renewable energy source, has attracted widespread attention for its development and utilization. However, hydropower systems are highly sensitive to climate change, significantly impacting generation, management, and safety. This study addresses the stability of hydropower resources in the Vakhsh River Basin, Tajikistan, using digital analysis, snowmelt runoff simulation, and soil erosion assessment to estimate spatial distribution. Under three climate scenarios (RCP2.6, RCP4.5, and RCP8.5), hydropower trends were simulated, and soil erosion was quantified. Results show annual hydropower potentials: Garm (55.465 billion kWh/a), Rogun (112.737 billion kWh/a), Nurex (78.853 billion kWh/a). Across all scenarios, runoff and hydropower generation increase (162–328,108 kWh/a), with growth rates following RCP4.5 < RCP2.6 < RCP8.5. Soil erosion simulation results indicate that a one millimeter increase in precipitation could lead to sediment deposition of  $1.57 \times 10^6$  kWh/year in upstream reservoirs. These results demonstrate that climate change has a significant impact on hydropower development in the Vakhsh River Basin. The research provides technical support for hydropower development under climate change.

**Keywords:** hydropower resources; climate change; hydrological model; soil erosion; Vakhsh River

## 1. Introduction

With the growing population and continuous socioeconomic development, energy demand and environmental pressures are increasing [1,2]. As a clean and renewable energy source, hydropower exhibits characteristics such as renewability, cost-effectiveness, accessibility, and high efficiency, making its development and utilization widely recognized [3].

With the continuous growth of energy demand and the escalating impacts of global climate change [4], countries worldwide have prioritized hydropower development as a key focus of energy strategy and a shared choice for addressing climate change and achieving sustainable development [5,6]. Currently, hydropower meets approximately 20% of global electricity demand [7], with more than 55 countries relying on it for over half of their power supply, including 24 nations where hydropower accounts for more than 90% of electricity generation [8,9]. Hydropower is highly sensitive to climatic variability and climate change [6,10], particularly in Central Asian countries vulnerable to

these impacts [11]. In the Gorno-Badakhshan region of Tajikistan, the Fedchenko Glacier has shown significant shrinkage [12]. Research by Hydro-Québec indicates that climate change-induced phenomena (such as earlier spring snowmelt, reduced summer flows, and increased winter discharges) can severely affect power generation capacity, peak regulation, demand management, and dam safety [13,14].

Central Asia is experiencing a climatic divergence trend of ‘west drying and east wetting’ [15,16]. Over the past 30 years, the annual average temperature in the basin has risen by 1.2–1.8 °C, accelerating glacier retreat. Simulations indicate that a 2 °C temperature increase could lead to a short-term surge in runoff by 16.51% [17]. Data from the Climatic Research Unit (CRU) reveal intensified spatiotemporal precipitation variability in the basin from 1970 to 2013, with notable autumn drying trends (regional drought indices decreased by 23%) and increased spring moisture. The escalation of extreme rainfall events has exacerbated reservoir sedimentation, while droughts have heightened winter power shortage risks [18].

As the largest tributary of the Amu Darya in Tajikistan, the Vakhsh River is the country’s core basin for hydropower development, supplying 93% of its electricity [19]. The river’s flow is predominantly fed by alpine glaciers and snowmelt, exhibiting strong seasonality: winter runoff accounts for only 15–20% of the annual total, while summer melt contributes over 60% [20]. This unique replenishment mechanism renders the Vakhsh Basin’s hydropower system highly vulnerable to climate change. Increased seasonal runoff fluctuations have reduced winter generation by 30–40%, severely threatening grid stability. Moreover, sediment deposition has raised annual maintenance costs for power plants like Nurek by more than 50%, shortening their designed lifespans [21]. Climate change has also escalated transboundary water disputes, posing challenges to regional harmony [22].

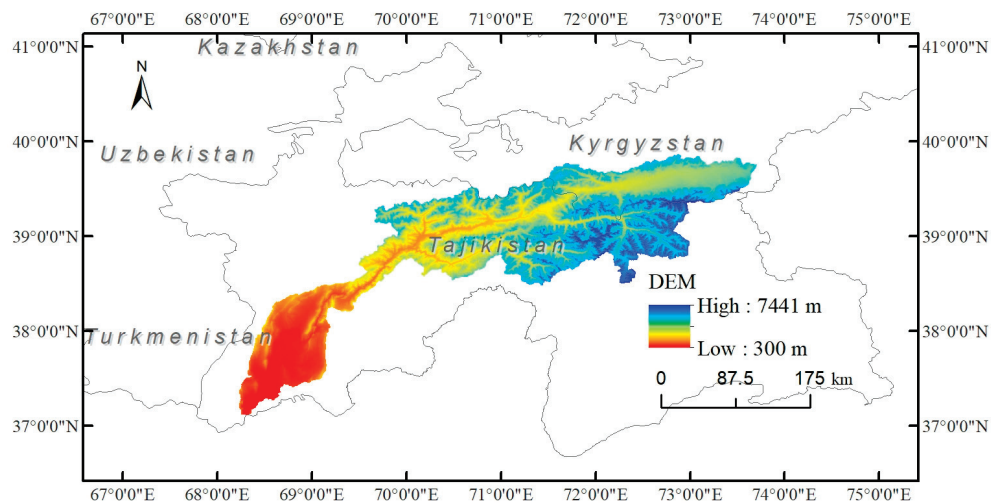
This study will quantitatively analyze the current hydropower resources in the Vakhsh River Basin and their changes under different climate scenarios, while assessing the risks posed by soil erosion on hydropower development. The findings will provide decision-making support for sustainable hydropower development in Central Asia.

This paper focuses on the changes in hydropower resources in the Vakhsh River Basin and primarily conducts the following research: (1) Assessment of hydropower resources in the Vakhsh River Basin; (2) Impact of climate change on runoff and soil erosion; (3) Influence of changes in runoff and soil erosion on hydropower resources.

## 2. Materials and Methods

### 2.1. Study Area

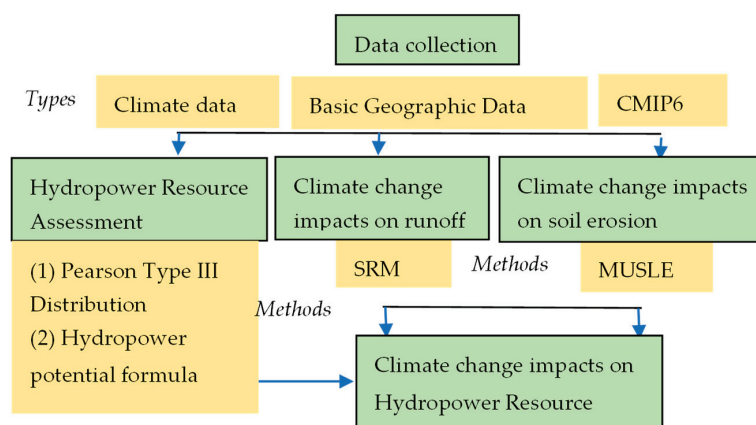
This study takes the Vakhsh River Basin in Tajikistan as the research area. The river originates from the Daumuruk Mountains in southern Kyrgyzstan, flows through Kyrgyzstan and Tajikistan, and merges into the Amu Darya. The Vakhsh River has a total length of 524 km and a drainage area of 39,100 km<sup>2</sup>, with 7900 km<sup>2</sup> located within Kyrgyzstan. The upper reaches are situated in the southeastern part of Kyrgyzstan, bordering China and Tajikistan, spanning 39° N to 40° N and 72° E to 74° E [23] (Figure 1). The source region is characterized by towering mountains and deep valleys, with an average elevation exceeding 3000 m. The climate is predominantly temperate continental, with distinct plateau features. The upper reaches (above 3000 m) have an annual average temperature below 10 °C, while the downstream valley areas approach 18 °C. The southeastern part of the basin (such as the Pamir Plateau) receives an average annual precipitation of about 511 mm, which significantly decreases downstream, with some areas receiving less than 200 mm annually.



**Figure 1.** Study area distribution map.

## 2.2. Research Framework

In order to analyze the impact of climate change on hydropower resource development in the Vakhsh river basin of Tajikistan, we designed the research framework as follows (Figure 2).



**Figure 2.** Methodology flowchart.

## 2.3. Data Collection

The National Hydrometeorology Agency (NHA) (<http://meteo.tj> (accessed on 2 May 2025)) provided temperature and precipitation data for 1975–2016. CMIP6 projections for 2020–2100 were generated for three Representative Concentration Pathways (RCPs): RCP2.6, RCP4.5, and RCP8.5. These correspond to greenhouse gas emission scenarios with radiative forcing values of 2.6, 4.5, and 8.5 W/m<sup>2</sup>, respectively, and project CO<sub>2</sub> concentrations to reach approximately 420, 540, and 940 ppm by 2100 [23]. Snow (1975–2016) and CMIP6 (2020–2100) datasets were sourced from the National Aeronautics and Space Administration (NASA) (<https://www.nccs.nasa.gov/> (accessed on 3 May 2025)) [24].

The 90 m × 90 m SRTM DEM data for the study area was sourced from the Geospatial Cloud (<https://www.gscloud.cn/> (accessed on 3 May 2025)). Land use data and runoff data (1976–1990) were obtained from the National Tibetan Plateau/Third Pole Environment Data Center (<https://data.tpdac.ac.cn/> (accessed on 3 May 2025)).



#### 2.4. Hydropower Resource Assessment

The theoretical hydropower potential of a river represents the long-term average of its hydraulic potential energy, calculated based on the river's mean annual flow and sectional elevation differences. This study employs a digital elevation model (DEM) with 20 km segmentation for computation. The theoretical potential is derived from the elevation difference between upstream and downstream cross-sections and the annual average runoff, using the following formula [25]:

$$E_{TRWater} = \rho \cdot g \cdot Q \cdot \Delta H \cdot t \quad (1)$$

where  $E_{TRWater}$  is theoretical hydropower potential of the river (J);  $\rho$  is water density (taken as 1.0 kg/L in this study);  $g$  is gravitational acceleration (calculated as 9.8014 m/s<sup>2</sup> using the 1979 International Gravity Formula at 39° latitude);  $Q$  is annual average runoff (m<sup>3</sup>/s);  $\Delta H$  is elevation difference between cross-sections (m);  $t$  is time duration (8760 h per year).

#### 2.5. Simulation of Climate Change Impacts on Runoff

Given the significant contribution of snowmelt runoff to the Vakhsh River Basin (accounting for a high proportion of total discharge), the SRM (Snowmelt Runoff Model) was employed to simulate hydrological processes under climate change scenarios.

The Snowmelt Runoff Model (SRM) is a conceptual degree-day model that incorporates snow-covered area as a key input variable. Its governing equation is given by Equation (2) [26].

$$Q_{n+1} = [Csn \cdot a_n(T_n + \Delta T_n)S_n + c_{Rn}P_n] \frac{A \cdot 10,000}{86,400} (1 - k_{n+1}) + Q_n k_{n+1} \quad (2)$$

such that  $Q$  is the average daily discharge [m<sup>3</sup>·s<sup>-1</sup>];  $C$  is the runoff coefficient expressing the losses as a ratio of runoff/precipitation;  $C_s$  is the runoff coefficient of snowmelt;  $C_R$  is the runoff coefficient of rain;  $a$  is the degree-day factor snowmelt depth resulting from 1-degree day [cm·°C<sup>-1</sup>·day<sup>-1</sup>];  $\Delta T$  is the extrapolation of temperature from one station to the average hypsometric elevation of the basin or altitudinal zone [°C];  $S$  is the ratio of snow-covered extend to the total area;  $P$  is the precipitation contributing to runoff;  $A$  is the area of the basin or altitudinal zone [km<sup>2</sup>];  $k$  is the recession coefficient showing the reduction in discharge in a period without snowmelt or rainfall;  $n$  is the sequence of days during the discharge computation period.  $\frac{10,000}{86,400}$  is the conversion of cm/km<sup>2</sup>/day to m<sup>3</sup>/s.

Considering the pronounced effects of climate change in Central Asia, our modeling framework was implemented across all three Representative Concentration Pathways (RCP2.6, RCP4.5, and RCP8.5) using climate projections from the Coupled Model Intercomparison Project Phase 5 (CMIP6) ensemble (<https://esgf-node.llnl.gov/projects/> (accessed on 4 May 2025)). Among them, RCP2.6 corresponds to stringent mitigation targets (requiring net-negative emissions by 2100), RCP4.5 represents moderate policy intervention (stabilizing radiative forcing at 4.5 W/m<sup>2</sup> by 2100), while RCP8.5 exemplifies a high-emission scenario with ineffective mitigation measures (resulting in 8.5 W/m<sup>2</sup> forcing by 2100).

#### 2.6. Impact of Soil Erosion on Hydropower Resources Under Climate Change

Soil erosion was quantified using the Modified Universal Soil Loss Equation (MUSLE), an enhanced erosion model developed by Williams [27] based on the Universal Soil Loss Equation (USLE). MUSLE is a quantitative model that characterizes the relationship between hillslope soil loss and multiple influencing factors. It is widely applied in soil and

water conservation, erosion assessment, and the prediction of soil loss rates on sloping agricultural lands. The MUSLE equation is expressed as [28]:

$$A = R \times K \times LS \times C \times P \quad (3)$$

In the equation,  $A$  is the predicted annual soil erosion amount ( $t/(km^2 \cdot a)$ ),  $R$  is the rainfall erosivity factor ( $MJ \cdot mm/(hm^2 \cdot h)$ ), serving as a kinetic indicator of runoff-induced erosion, where rainfall intensity and duration significantly influence erosion.  $LS$  is the topographic factor (dimensionless), with  $L$  (slope length factor) representing soil erosion normalized to a 22.13 m slope length and  $S$  (slope steepness factor) representing erosion normalized to a  $5.14^\circ$  slope; this study derived  $LS$  factors from DEM (Digital Elevation Model) data.  $K$  is the soil erodibility factor ( $t \cdot hm^2 \cdot h/(hm^2 \cdot MJ \cdot mm)$ ), reflecting soil sensitivity to detachment and transport by erosive forces.  $C$  is the cover and management factor (dimensionless), quantifying the effect of vegetation and management practices on erosion.  $P$  is the support practice factor (dimensionless), representing the ratio of soil loss with conservation measures to that under conventional tillage.

In high-altitude regions, soil erosion rates are often underestimated due to snowmelt erosion and freeze–thaw cycles [29], while being overestimated owing to the steep slope effect. DEM is often used to adjust parameters to accommodate this change. Peng et al. successfully employed the USLE model to estimate soil erosion in the high-altitude Qilian Mountain area in China. Given that our study area closely resembles Qilian Mountain, we adopted the same  $LS$  parameters as Peng's [30].

The rainfall erosivity factor  $R$  reflects the impact of precipitation on soil erosion, and its calculation formula is [31]:

$$R = \sum_{i=1}^{12} 1.735 \times 10^{[(1.5 \times I g_p^{p_i^2} - 0.8188)]} \quad (4)$$

In the equation,  $p_i$  the monthly precipitation (mm), where  $i$  represents the month of the year ( $i = 1, 2, \dots, 12$ );  $p$  is the mean annual rainfall (mm).

$K$  represents the soil erosion per unit area caused by rainfall erosivity, and its calculation formula is as follows [32]:

$$K = \left\{ 0.2 + 0.3 \times \exp \left[ -0.0256 \times SAN \left( 1 - \frac{SIL}{100} \right) \right] \right\} \times \left( \frac{SIL}{CLA + SIL} \right)^{0.3} \times \left( 1 - \frac{0.25 \times C}{C + \exp(3.72 - 2.95 \times C)} \right) \times \left( 1 - \frac{0.7 \times SN1}{SN1 + \exp(-5.51 + 22.9 \times SN1)} \right) \quad (5)$$

Here,  $SAN$ ,  $SIL$ ,  $CLA$  and  $C$ , respectively, represent the mass fractions (%) of sand, silt, clay, and organic carbon. The calculation for  $SN1$  follows:  $SN1 = 1 - (SAN/100)$ .

$LS$  is the slope length-gradient factor (dimensionless), where  $L$  represents the slope length factor and  $S$  denotes the slope gradient factor. Under unchanged conditions,  $LS$  quantifies the ratio of soil loss on a given slope (with specific length and gradient) to that on a standard runoff plot in a typical slope condition. The calculation formula is as follows [33]:

$$S = \begin{cases} 10.8 \times \sin \theta + 0.03, & \theta < 5^\circ \\ 16.8 \times \sin \theta - 0.50, & 5^\circ \leq \theta < 14^\circ \\ 21.9 \times \sin \theta - 0.96, & \theta \geq 14^\circ \end{cases} \quad (6)$$

where  $\theta$  denotes the slope ( $^\circ$ ).

The slope length factor ( $L$ ) is calculated using the classic RUSLE method, with the formula as follows:

$$L = \left( \frac{\lambda}{22.13} \right)^m \quad (7)$$

$$m = \begin{cases} 0.2, & \theta \leq 1^\circ \\ 0.3, & 1^\circ < \theta \leq 3^\circ \\ 0.4, & 3^\circ < \theta \leq 5^\circ \\ 0.5, & \theta > 5^\circ \end{cases} \quad (8)$$

Here,  $\lambda$  denotes the slope length (m), and  $\theta$  represents the slope angle extracted from the DEM (Digital Elevation Model).

The cover and management factor ( $C$ ) quantifies the ratio of soil loss under a specific vegetation cover to that under cultivated conditions, assuming all other factors remain constant. The calculation formula is as follows [34]:

$$C = \begin{cases} 1, & c = 0 \\ 0.6508 - 0.3436 \lg(c), & 0 < c \leq 78.3 \\ 0, & c > 78.3 \end{cases} \quad (9)$$

Here,  $c$  represents the vegetation cover degree, with  $C$  ranging from 0 (indicating no soil erosion) to 1 (dimensionless).

$P$  is the soil conservation practice factor (dimensionless), representing the ratio of soil loss after implementing conservation measures to that under slope farming. Its value ranges from 0 to 1. For built-up areas and water bodies,  $P$  is defaulted to 0 under normal conditions. Based on the study area's land use, USDA Handbook 703, and watershed characteristics,  $P$  values are calibrated for different land use types. The assigned  $P$  factors for each land use category are listed in Table 1.

**Table 1.**  $p$  values for different land use types.

| Land Use Type | $p$  |
|---------------|------|
| Farmland      | 0.01 |
| Forestland    | 0.2  |
| Grassland     | 0.6  |
| Water Body    | 0    |
| Built-up Area | 1    |
| Bare Land     | 0.8  |

### 2.7. Pearson Type III Distribution

The Pearson Type III distribution is widely used in hydrological frequency analysis due to its excellent fit with the frequency distributions of hydrological variables (such as floods and runoff) in most regions, particularly its ability to effectively capture the asymmetric characteristics of extreme events [35].

The Pearson Type III curve is an asymmetric, single-peaked, positively skewed distribution with one finite end and one infinite end, mathematically known as the Gamma distribution. Its probability density function is [36,37]:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} (x - a_0)^{\alpha-1} e^{-\beta(x-a_0)} \quad (10)$$

In the equation,  $\Gamma(\alpha)$  represents the gamma function of  $\alpha$ ;  $\alpha$ ,  $\beta$ , and  $a_0$  are the three parameters characterizing the shape, scale, and location of the Pearson Type III distribution, respectively, where  $\alpha > 0$  and  $\beta > 0$ .

$$\begin{cases} \alpha = \frac{4}{C_s^2} \\ \beta = \frac{2}{\bar{x}C_sC_v} \\ a_0 = \bar{x}(1 - \frac{2C_v}{C_s}) \end{cases} \quad (11)$$

In the equation,  $\bar{x}$  represents the mean,  $C_s$  denotes the coefficient of skewness, and  $C_v$  stands for the coefficient of variation.

In hydrologic calculations, it is generally necessary to determine the random variable  $X_p$  corresponding to a specified frequency  $P$ . This is achieved by integrating the probability density function and solving for the cumulative frequency  $P$ , which is equal to or greater than  $X_p$ . The relationship can be expressed as:

$$P = F(x_p) = P(x \geq x_p) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_{x_p}^{\infty} (x - a_0)^{\alpha-1} e^{-\beta(x-a_0)} dx \quad (12)$$

### 2.8. Simulation and Analysis Software

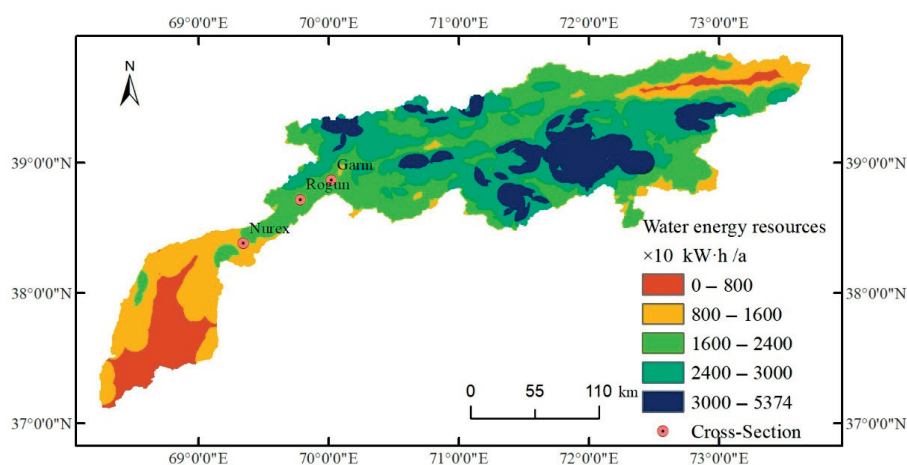
Due to the inclusion of high-altitude areas in the study region, glacier and snowmelt are significant components of runoff. This study introduces the Snowmelt Runoff Model (SRM) for runoff simulation and analysis. SRM is a hydrological model widely used for snowmelt runoff simulation in mountainous watersheds, based on the degree-day method as its core theory. It features a simple structure and a clear physical meaning of parameters [33]. The spatial analysis designed in this paper was conducted using ArcGIS 10.3 software. ArcGIS is a professional Geographic Information System (GIS) software suite developed by ESRI, widely used in spatial data management, analysis, and visualization [38].

## 3. Results and Discussion

### 3.1. Water Energy Resource Assessment in the Vakhsh River Basin

#### 3.1.1. Spatial Distribution of Hydropower Resources

According to the suitable spacing requirements between hydropower station constructions, the elevation drop of unit cells centered at 20 km intervals was evaluated point by point within the basin. The annual average runoff of river sections adopted the multi-year average values from 1976 to 1990. The hydropower resources of the Vakhsh River Basin were calculated using Formula (1), where the water density  $\rho$  was taken as  $1.0 \text{ kg/m}^3$ , the gravitational acceleration  $g$  as  $9.8014 \text{ m/s}^2$ , and the annual hour count  $t$  as 8760 h. The calculation results are shown in Figure 3:



**Figure 3.** Spatial distribution of hydropower resources in the Vakhsh river basin.



As shown in Figure 3, the hydropower resources in the Vakhsh River Basin range from 52 to 537.4 billion kWh/year, with high-value areas primarily distributed in the mountainous upstream region above Nurex and low-value areas concentrated in the downstream plain regions. Further statistical analysis reveals that 8.90% of the regions have less than 80 billion kWh/year, 17.53% fall between 80 and 160 billion kWh/year, 32.97% are between 240 and 300 billion kWh/year, 26.34% between 240 and 300 billion kWh/year, and 14.26% exceed 300 billion kWh/year. Over 90% of the regions surpass 100 billion kWh/year, indicating abundant hydropower resources in the basin.

Based on existing hydropower infrastructure, three hydrological sections (Garm, Rogun, and Nurek) were selected for assessment (spatial locations shown in Figure 2), with results detailed in Table 2.

**Table 2.** Calculation of hydropower resources at typical hydrological sections.

| Name  | Elevation Difference (m) | Mean Annual Flow (m <sup>3</sup> /s) | Hydropower Resources ( $\times 10^8$ kWh/a) |
|-------|--------------------------|--------------------------------------|---------------------------------------------|
| Garm  | 2007                     | 321.92                               | 554.65                                      |
| Rogun | 2036                     | 645.0                                | 1127.37                                     |
| Nurek | 1413                     | 650.05                               | 788.53                                      |

Table 2 shows that the hydropower resources of all three sections exceed 50 billion kWh/year, with the ranking Rogun > Nurek > Garm. The data indicate that the distribution of hydropower resources among different sections is not solely governed by upstream-downstream relationships. Although downstream sections (e.g., Garm) exhibit higher annual runoff, a notable elevation difference in specific areas can exert a more pronounced impact on the outcomes.

### 3.1.2. Frequency Distribution Characteristics of Hydropower Resources

Since annual runoff is highly variable and uncertain, it poses challenges for water energy development. To reduce risks and optimize economic investment, understanding the frequency distribution of hydropower resources is critical. A Pearson Type III distribution analysis was conducted on the hydropower resources of the three representative sections (Garm, Rogun and Nurek), with the results shown below.

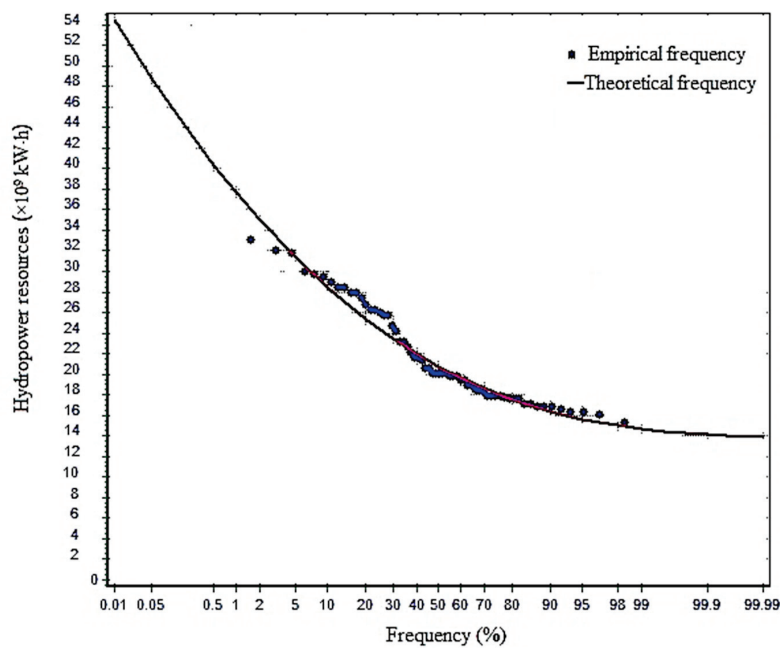
Figure 4 shows that when the guarantee rate is between  $75\% \leq p \leq 90\%$ , the distribution range of hydropower resources is:  $1.81 \times 10^{10}$  kWh/a  $\leq E_{\text{Garm}} \leq 1.64 \times 10^{10}$  kWh/year,  $3.67 \times 10^{10}$  kWh/a  $\leq E_{\text{Rogun}} \leq 3.33 \times 10^{10}$  kWh/year,  $2.57 \times 10^{10}$  kWh/a  $\leq E_{\text{Nurek}} \leq 2.33 \times 10^{10}$  kWh/a. For cascaded hydropower stations, the guaranteed capacity calculation requires coordination between upstream and downstream stations, with the conventional guarantee rate ranging from 75% to 90%, making the three sections suitable for exploitation planning within this range.

### 3.1.3. Intra-Annual Distribution Characteristics of Hydropower Resources

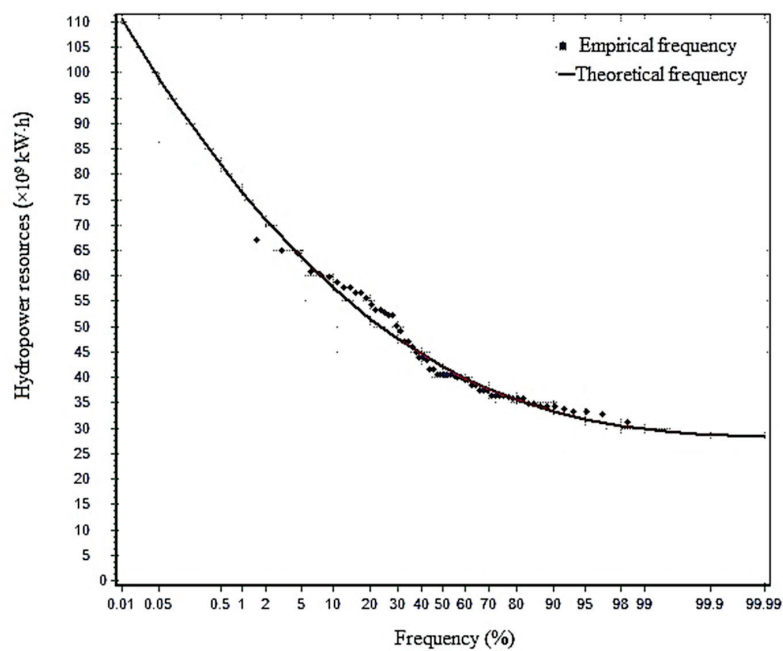
The intra-annual distribution of hydropower resources is a critical basis for designing the storage capacity of diversion reservoirs in power stations to address water regulation during non-flood seasons. Statistical analysis of multi-year (1976–1985) average data yields the following results (Figure 5):

Figure 5 shows that the intra-annual distribution of hydropower resources is highly uneven, with the lowest values occurring in January–March ( $6.03 \times 10^8$  kW·h,  $12.24 \times 10^8$  kW·h,  $8.58 \times 10^8$  kW·h, accounting for 2.77% of the annual total each), and the peak values in June ( $47.83 \times 10^8$  kW·h,  $97.04 \times 10^8$  kW·h,  $68.02 \times 10^8$  kW·h, accounting for

21.96% of the annual total each). Overall, the flood season (May–September) accounts for 75.13%, while the other seven months account for 24.87%.

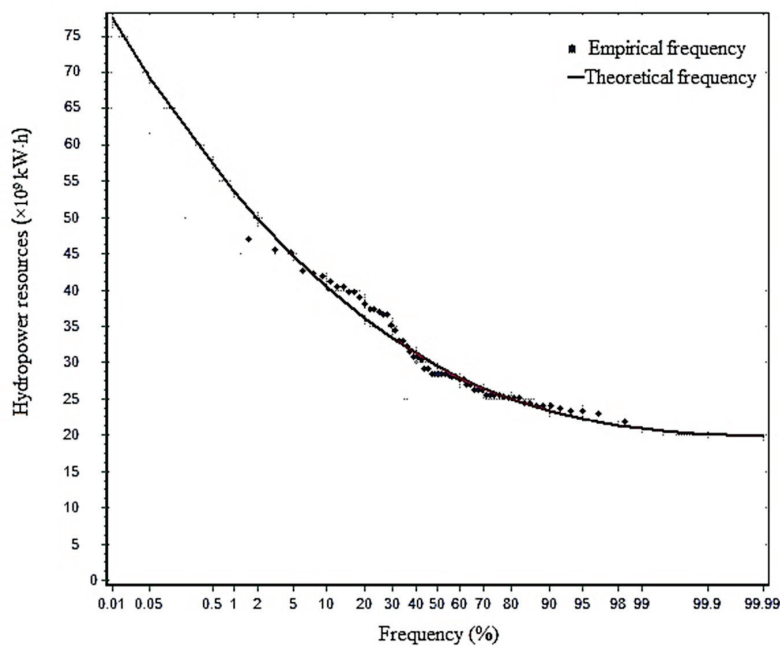


(A) Garm station

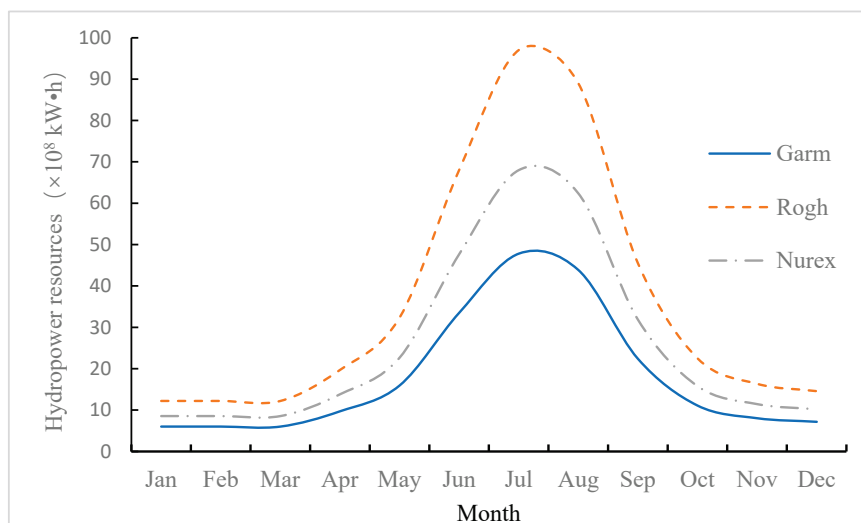


(B) Rogun station

Figure 4. Cont.



(C) Nurek station

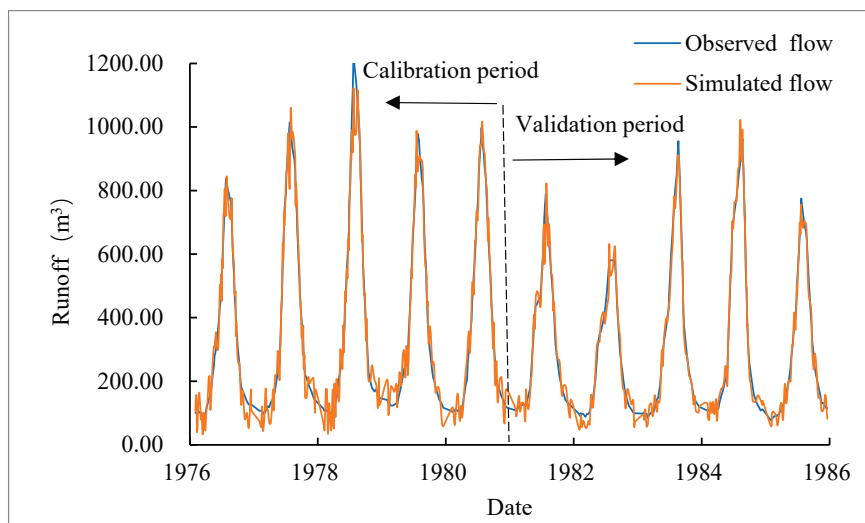
**Figure 4.** Frequency distribution of hydropower resources.**Figure 5.** Intra-annual distribution characteristics of hydropower resources.

### 3.2. Impact of Climate Change on Hydropower Resources

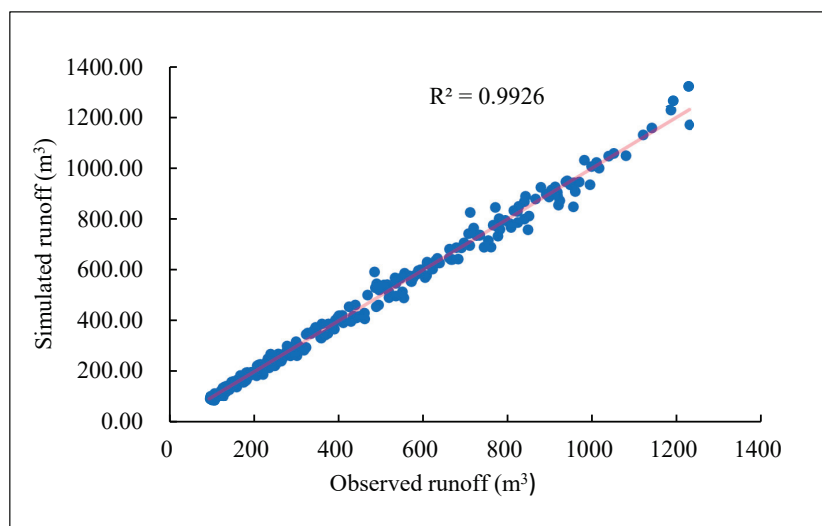
#### 3.2.1. Impact of Climate Change on Runoff

In the Vakhsh River Basin, glacier and snow melt driven by climate change significantly impacts river runoff. Therefore, this study introduces a snowmelt runoff model (SRM) to simulate runoff variations. Based on existing data, the 1976–1980 data from Garm Hydrological Station were used for model calibration, and the 1981–1985 data were used for model validation. The results are shown in Figure 5.

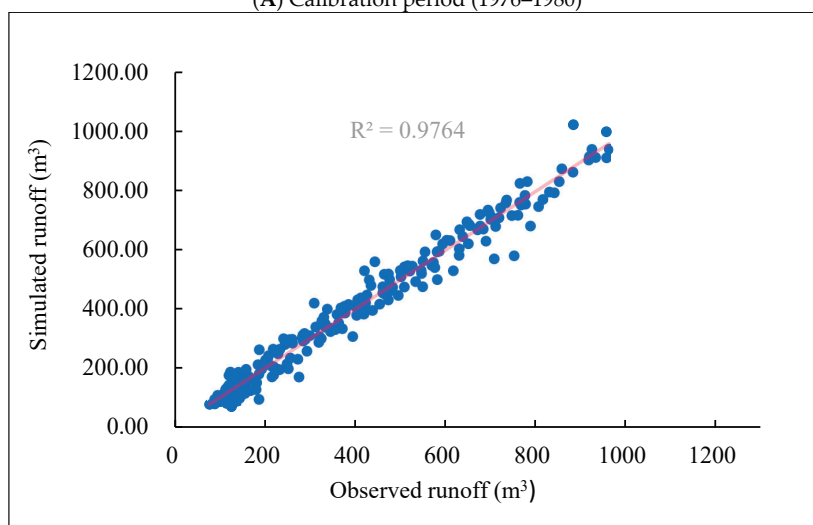
Figure 6 shows excellent consistency between observed and simulated runoff during calibration and validation periods. To further quantitatively analyze its accuracy, the simulation accuracy metrics (Coefficient of Determination  $R^2$  and Nash–Sutcliffe Efficiency Coefficient NSE) were calculated. The results are presented in Figure 7.



**Figure 6.** Runoff simulation of the Garm hydrological section based on the SRM model.



**(A)** Calibration period (1976–1980)



**(B)** Validation period (1981–1985)

**Figure 7.** Comparison of actual and simulated daily runoff using scatter plots.



As shown in Figure 7, during the calibration period, the  $R^2$  was 0.9826 and the NSE was 0.9813, while during the validation period, the  $R^2$  was 0.9764 and the NSE was 0.9763. The SRM model demonstrates high accuracy in simulating hydrological processes at the Garm Hydrological Station, indicating its practical applicability in the Vakhsh River Basin for further predictive studies. Certainly, the high accuracy of the simulation results may also be related to the relatively short time series used for modeling, which introduces significant uncertainty in the outcomes.

The changes in climate data (maximum temperature (TMax), minimum temperature (TMin), average temperature (Tav) and precipitation( $p$ )) for the study area compared to the simulation period are presented in Table 3.

**Table 3.** Fluctuation values of climate scenarios in the Vakhsh River Basin compared to the baseline (1976–1986) period.

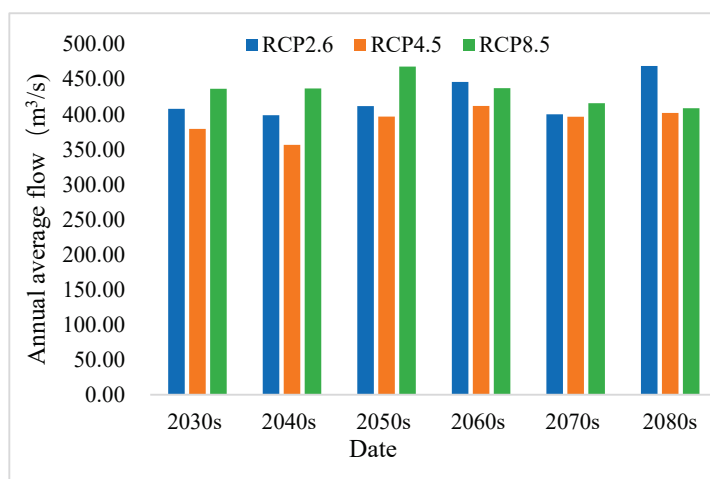
| Variable  | Scenario | 2030s | 2040s | 2050s | 2060s | 2070s | 2080s |
|-----------|----------|-------|-------|-------|-------|-------|-------|
| TMax (°C) | RCP2.6   | 0.19  | 0.18  | 0.33  | 0.46  | 0.44  | 0.27  |
|           | RCP4.5   | 0.20  | 0.28  | 0.36  | 0.52  | 0.62  | 0.61  |
|           | RCP8.5   | 0.21  | 0.51  | 0.76  | 0.83  | 1.19  | 1.40  |
| TMin (°C) | RCP2.6   | 0.22  | 0.22  | 0.35  | 0.40  | 0.44  | 0.27  |
|           | RCP4.5   | 0.26  | 0.32  | 0.33  | 0.50  | 0.49  | 0.51  |
|           | RCP8.5   | 0.19  | 0.46  | 0.63  | 0.70  | 1.04  | 1.15  |
| Tav (°C)  | RCP2.6   | 0.21  | 0.20  | 0.34  | 0.43  | 0.44  | 0.27  |
|           | RCP4.5   | 0.23  | 0.30  | 0.35  | 0.51  | 0.56  | 0.56  |
|           | RCP8.5   | 0.20  | 0.49  | 0.69  | 0.76  | 1.12  | 1.28  |
| $p$ (mm)  | RCP2.6   | 6.67  | 5.96  | 6.97  | 9.60  | 6.06  | 11.41 |
|           | RCP4.5   | 6.27  | 3.79  | 8.18  | 9.86  | 8.16  | 8.75  |
|           | RCP8.5   | 8.19  | 8.22  | 10.44 | 8.25  | 6.72  | 6.21  |

Table 3 shows that over the next 50 years, the Vakhsh River Basin is projected to experience an overall warming trend, with multi-year temperature increases ranging from 0.31 to 0.82 °C and precipitation increases of approximately 8 mm. From a climate scenario perspective, temperature increases follow the order  $RCP2.6 < RCP4.5 < RCP8.5$ , with RCP2.6 showing a slight increase of over 0.3 °C, RCP4.5 showing a slight increase of over 0.4 °C, and RCP8.5 showing an increase exceeding 0.7 °C, while precipitation increases remain within the range of 7.5–8.0 mm across all scenarios.

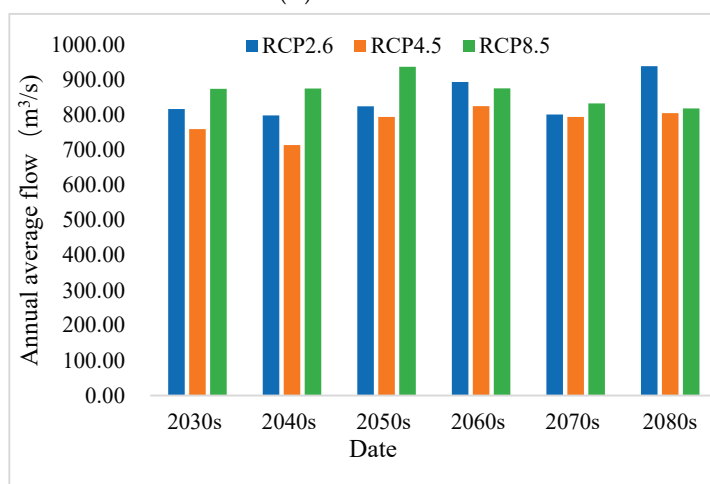
Based on the calibrated parameters, SRM was used to simulate the average monthly predicted values of MaxT, MinT, Average T, and precipitation under the RCP2.6, RCP4.5, and RCP8.5 scenarios. The simulation results are presented in Figure 8.

Figure 8 shows that runoff increases to varying degrees with rising precipitation and temperature. Over the next 50 years, with a 95% confidence level, the total flow rates for the Garm, Rogun, and Nurex sections are  $415.64 \pm 57.77$ ,  $832.78 \pm 114.52$ , and  $839.30 \pm 168.09$  m<sup>3</sup>/s, respectively. The flow increase follows the order  $RCP4.5 < RCP2.6 < RCP8.5$ , which differs from the precipitation trend. This inconsistency occurs because the basin relies heavily on snowmelt runoff, where temperature rise also plays a significant role. At the Garm, Rogun, and Nurex sections, with a 95% confidence level, flow increases by  $93.72 \pm 13.04$ ,  $187.78 \pm 25.84$ , and  $189.25 \pm 37.91$  m<sup>3</sup>/s, respectively, with all three seeing a nearly 30% increase.

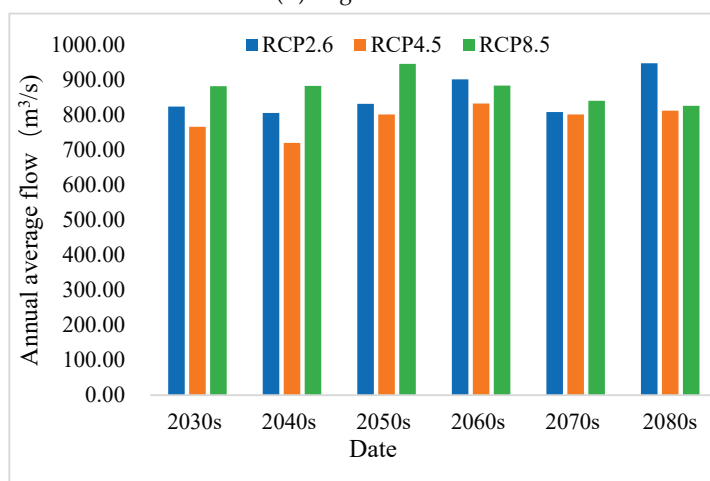
The research findings of this paper are consistent with previous studies. Central Asia is projected to shift toward a warmer and wetter climate, with snowmelt leading to increased runoff near glaciers. Such climatic transitions will result in a rise in flood events across Central Asia [38].



(A) Garm station



(B) Rogun station

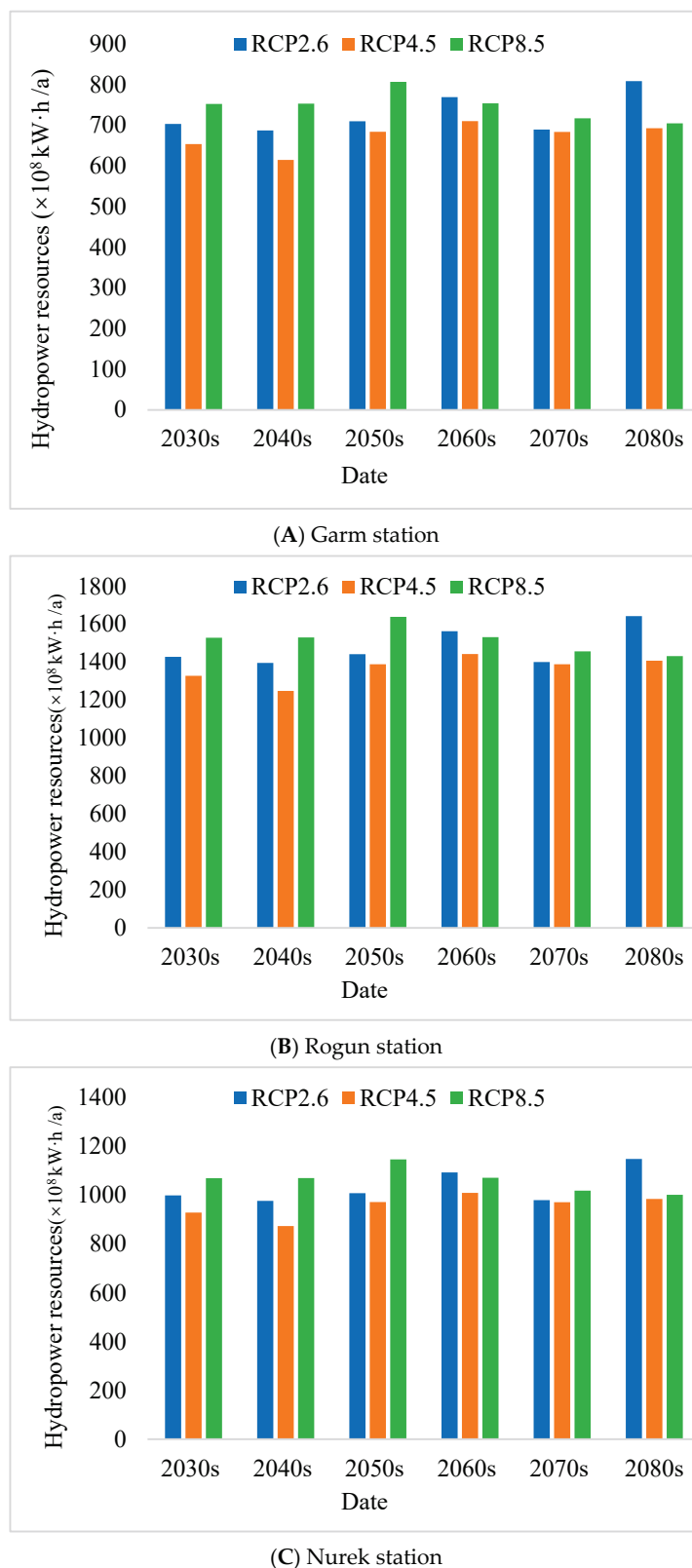


(C) Nurek station

**Figure 8.** Changes in runoff under different climate scenarios.

### 3.2.2. Impact of Climate Change on Hydropower Resources

Based on simulated runoff under different climate scenarios, hydropower resource changes were calculated using Formula (1), with the results shown in Figure 9 below.



**Figure 9.** Changes in hydropower resources under different climate scenarios.

Figure 9 shows that under the three scenarios (RCP2.6, RCP4.5, and RCP8.5), hydropower resources increase with rising runoff, with increments ranging from  $162 \pm 22.54$  to  $328,108 \pm 45,652.24$  kWh/a with a 95% confidence level. In terms of growth magnitude, the order is:  $RCP4.5 < RCP2.6 < RCP8.5$ . From a spatial perspective, Garm exhibits the

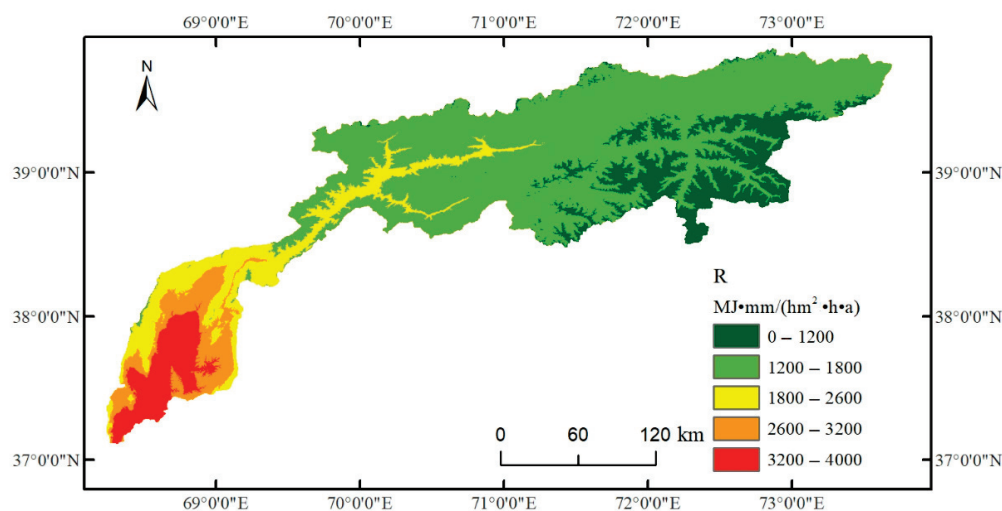
smallest increase, while Nurex shows the largest growth, followed by Rogun. The trend of change aligns with runoff variations.

### 3.2.3. Estimation of Soil Erosion

Changes in precipitation and resulting runoff variations under climate change can influence sediment deposition in hydropower reservoirs. This section employs a modified soil erosion equation to assess soil and water loss in the Vakhsh River Basin.

The Modified Universal Soil Loss Equation (MUSLE) is used to estimate soil erosion in the study area based on rainfall erosivity, soil erodibility, topography (slope length and gradient), vegetation cover and management, and soil conservation practices. The calculation method is detailed in Section 2.5. The evaluation results are presented in the following figure.

Figure 10 indicates that the rainfall erosivity in the study area ranges between 970–3800 MJ·mm/(hm<sup>2</sup>·h·a), showing an overall decreasing trend from the southwest to the northeast. High-value zones (generally exceeding 2600 MJ·mm/(hm<sup>2</sup>·h·a)) are primarily distributed in the downstream plain areas of the basin, while low-value zones (typically below 1800 MJ·mm/(hm<sup>2</sup>·h·a)) are concentrated in the northeastern upstream regions.

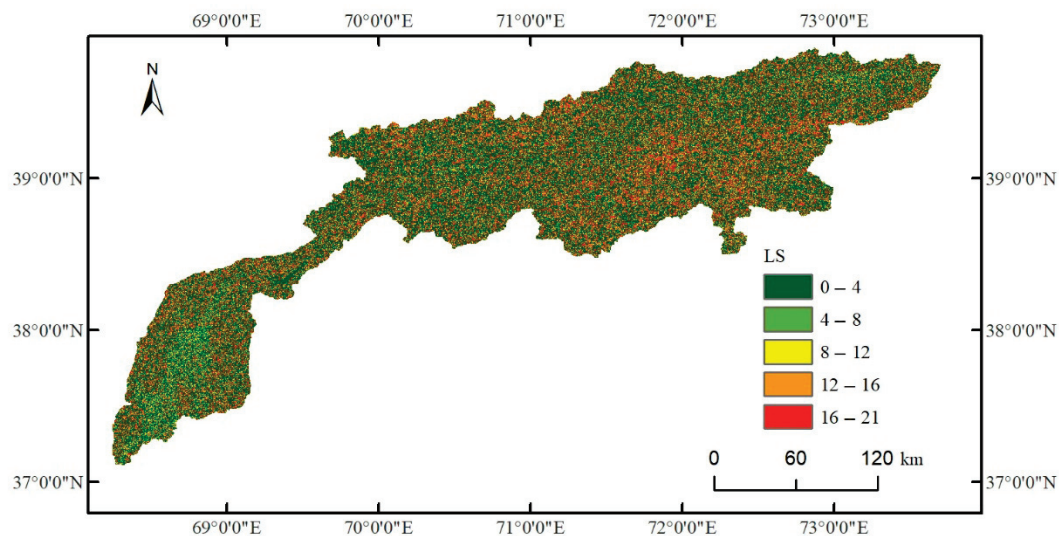


**Figure 10.** Spatial distribution of the R factor in the Vakhsh river basin.

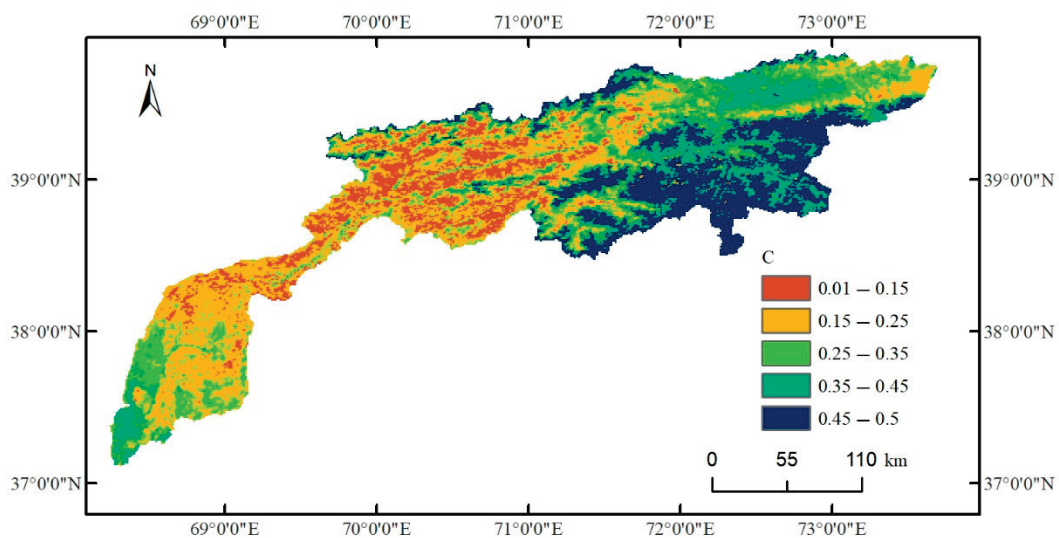
Figure 11 shows that the LS (topographic factor) in the basin ranges from 0 to 21, exhibiting an alternating distribution pattern of high and low values. The western upstream region generally has higher LS values compared to the eastern downstream region. Low-value zones are primarily concentrated near the main river channel, whereas high-value zones are predominantly distributed in the marginal snow-covered areas.

Figure 12 shows that the value of C in the basin ranges between 0.01 and 0.5, displaying an overall trend of higher values in the east and lower values in the west. Most areas in the east exceed 0.25, while most areas in the west are below 0.25. Additionally, there is a local high-value zone in the lower western region, and the western part exhibits a south-to-north decreasing distribution pattern.

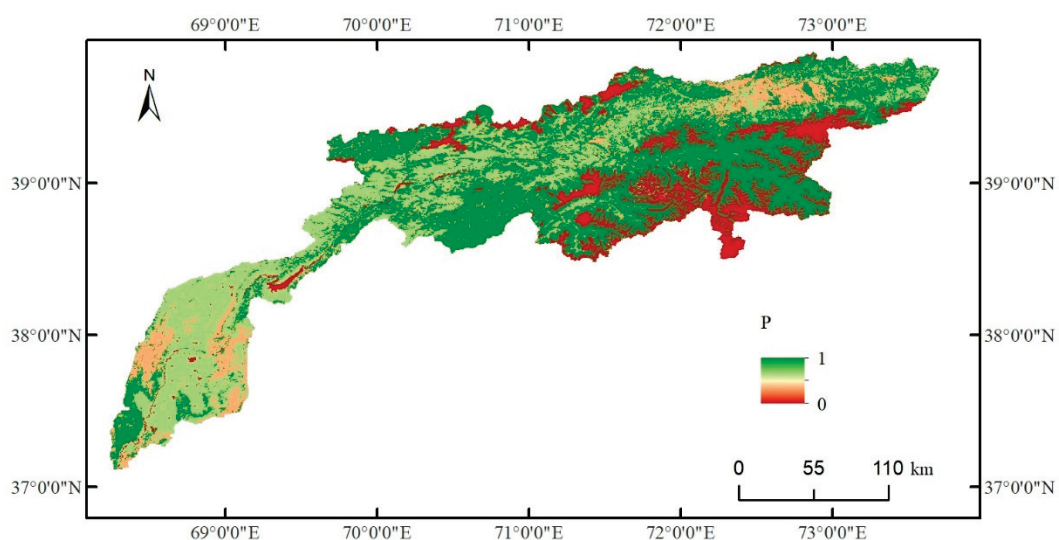
Figure 13 indicates that both the high-value and low-value areas of  $p$  in the basin are concentrated in the eastern region, with most eastern areas exceeding 0.25, while most western areas remain below 0.25. This suggests that soil and water conservation efforts are under greater pressure in the upstream areas.



**Figure 11.** Spatial distribution of the LS factor in the Vakhsh river basin.



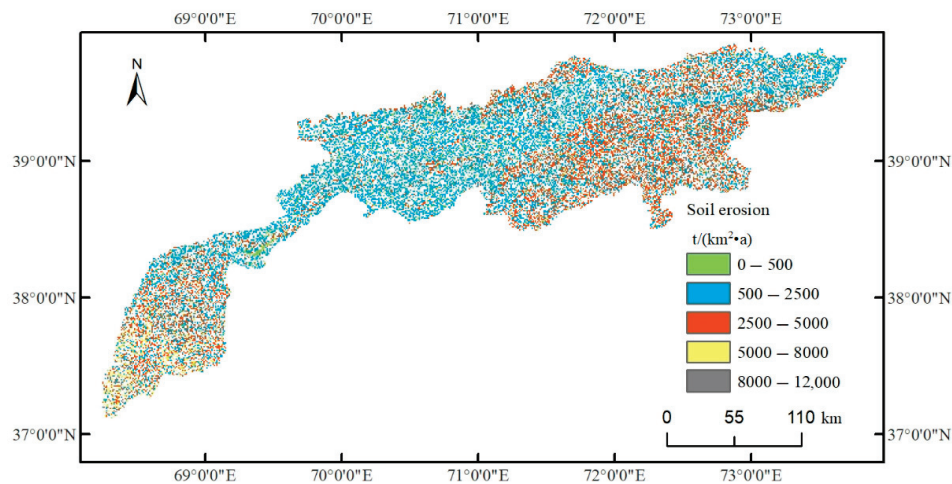
**Figure 12.** Spatial distribution of the C factor in the Vakhsh river basin.



**Figure 13.** Spatial distribution of the  $p$  factor in the Vakhsh river basin.



Figure 14 shows that soil erosion in the basin is generally higher in the eastern and western regions than in the central area, with most values below  $2500 \text{ t}/(\text{km}^2 \cdot \text{a})$ . According to the soil erosion classification standard, soil erosion can be categorized into six levels: slight, moderate, severe, very severe, extremely severe, and extreme, with classification thresholds as shown in the legend. Most areas in the east and southwest fall into moderate erosion or higher, with extreme erosion observed in some regions. The results indicate that the headwaters of the Wakash River basin suffer from severe soil erosion, posing a significant threat to sediment deposition in mid- and downstream hydropower reservoirs.



**Figure 14.** Spatial distribution of soil erosion in the Vakhsh river basin.

### 3.2.4. Impact of Soil Erosion Changes on Hydropower Development

Soil erosion affects power generation efficiency primarily through loss of reservoir capacity and head reduction. Sediment deposition directly encroaches on the effective storage capacity of reservoirs, diminishing their water level regulation capabilities. Studies indicate that every cubic meter of sediment deposition reduces usable reservoir capacity by approximately  $0.7\text{--}1.0 \text{ m}^3$ .

Head height (water level difference) is a critical parameter for power generation efficiency; sedimentation lowers the  $H$  value, directly reducing electricity output. Additionally, sediment accumulation at turbine inlets and diversion channels creates flow resistance, increasing head loss (the mechanical energy lost per unit weight of water flow). The Three Gorges Dam case demonstrates that sediment-induced friction can decrease power generation by 5–15%.

Moreover, equipment wear escalates maintenance costs. Some hydropower stations report annual maintenance expenses rising by 30–50% due to sediment-related damage.

Based on the average soil erosion conditions in the watershed over multiple years, with 75% of sediment deposited along the river channel and hydropower reservoir, head loss is further calculated, followed by energy loss estimation. The results are shown in Table 4.

**Table 4.** The impact of climate fluctuations on the Rogun hydropower station (95% confidence level).

| Item                        | Reservoir Capacity<br>( $\times 10^8 \text{ m}^3$ ) | Accumulation of<br>Sediment ( $\text{m}^3$ ) | Head Loss (m)                             | Hydropower Energy<br>Loss (kWh/a) |
|-----------------------------|-----------------------------------------------------|----------------------------------------------|-------------------------------------------|-----------------------------------|
| $p$ (Multi-year average)    | 133                                                 | 329,818.71                                   | $(8.31 \pm 1.15) \times 10^{-3}$          | $(1.57 \pm 0.22) \times 10^6$     |
| $p \uparrow + 1 \text{ mm}$ |                                                     | $\uparrow 1978.91$                           | $\uparrow (4.98 \pm 0.69) \times 10^{-5}$ | $\uparrow 939.93 \pm 130.78$      |

In the table,  $\uparrow$  represents an increase.

### 3.2.5. Uncertainty Analysis

This study primarily employs the SRM (Snowmelt Runoff Model) and MUSLE (Modified Universal Soil Loss Equation) models to analyze changes in water resources and hydropower under climate change scenarios. The uncertainties in the study mainly stem from the following aspects:

- (1) Multi-source input data. The analysis utilizes DEM (Digital Elevation Model), remote sensing data, climate change scenario data, and observed data. Uncertainties arise from the resolution differences among these data sources and the impact of cloud cover on remote sensing data, with climate change scenario data exhibiting particularly prominent uncertainties.
- (2) Hydrological model parameters. The spatial distribution of three hydrological meteorological stations in the basin is insufficient to accurately represent spatial variations in the Vakhsh River Basin, which features complex topography. Although most model parameters are derived from field observations and information inversion, data errors and calibrated parameters still introduce significant uncertainties.
- (3) MUSLE factor parameters. In practical applications of the MUSLE model, many scholars adapt it to specific scenarios, leading to variations in thresholds across different contexts. Given the study area's elevation range from low to high altitudes, high-altitude regions are characterized by steeper terrain and significant differences in soil erosion due to snow and glacier influence. This study does not classify these variations, introducing uncertainties into the results.

To address these issues, subsequent research will implement the following improvements: integrate high-resolution remote sensing data; interpolate and fuse spatial hydrological meteorological data; and refine regional segmentation based on terrain and runoff generation conditions. These enhanced measures will establish more robust feedback mechanisms.

## 4. Conclusions

Hydropower resources are critically important for the development of clean energy in Tajikistan. To address potential fluctuations in hydropower resources caused by climate change, this study takes the Vakhsh River Basin as a case study. Using the SRM model and MUSLE model, along with hydrological observation data, DEM, and CMIP6 datasets, it analyzes the impacts of changes in runoff and soil erosion under RCP2.6, RCP4.5, and RCP8.5 scenarios on hydropower resources. The main findings are as follows:

The analysis of hydropower potential reveals that the basin's resources range from 5.2 to 537.4 billion kWh/a, with high-value zones concentrated in mountainous areas upstream of Nurex and low-value zones in downstream plains. Over 90% of the region exceeds 100 billion kWh/a. At reliability levels of  $75\% \leq p \leq 90\%$ , hydropower values at different sections surpass  $1.81 \times 10^{10}$  kWh/a, enabling feasible exploitation planning.

Snowmelt runoff modeling: Results show runoff increases with rising precipitation and temperatures, with a projected 68–226 m<sup>3</sup>/s rise over 50 years. The trend follows RCP4.5 < RCP2.6 < RCP8.5. Under all three scenarios, hydropower grows in tandem with runoff (162–328,108 kWh/a), spatially and temporally consistent with runoff changes.

Soil erosion analysis reveals that eastern and southwestern regions experience moderate to severe erosion, posing significant siltation threats to downstream reservoirs. A one-millimeter increase in precipitation causes 939.93 kWh/a of hydropower loss due to erosion.

To address the impact of water resource changes on hydropower generation under climate change and ensure the operational safety of water conservancy projects, it is

essential to establish a climate assessment system for hydropower stations, optimize site selection and scheduling plans, develop emergency response protocols, and scientifically adjust reservoir water levels to mitigate droughts or floods.

**Author Contributions:** H.L.: Conceptualization, Methodology, Formal analysis, Visualization, Writing—Original draft, Writing—review and editing. A.G.: Writing—review and editing. F.S.: Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Science and Technology Program of Xinjiang Uyghur Autonomous Region (2025E01030) and the Intergovernmental international scientific and technological innovation cooperation project (2025YFE0102800).

**Data Availability Statement:** Data used in this study are available under reasonable request from authors.

**Acknowledgments:** The authors would like to thank the Institute of Water Problems, Hydropower and Ecology of the National Academy of Sciences of Tajikistan for providing us the historical average daily water flow data.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Gilland, B. Population, economic growth, and energy demand, 1985–2020. *Popul. Dev. Rev.* **1988**, *14*, 233–244. [CrossRef]
2. Holdren, J.P. Population and the energy problem. *Popul. Env.* **1991**, *12*, 231–255. [CrossRef]
3. van Ruijven, B.J.; De Cian, E.; Wing, I.S. Amplification of future energy demand growth due to climate change. *Nat. Commun.* **2019**, *10*, 2762. [CrossRef]
4. Siri, R.; Mondal, S.R.; Das, S. Hydropower: A renewable energy resource for sustainability in terms of climate change environmental protection. In *Alternative Energy Resources; The handbook of environmental chemistry*; Pathak, P., Srivastava, R.R., Eds.; Springer: Cham, Switzerland, 2020; Volume 99. [CrossRef]
5. Andronova, I.V.; Kuzmin, V.V.; Tinkova, A.A. Global Hydropower as the main driver of sustainable development in the context of industry 4.0. In *Industry 4.0*; Zavyalova, E.B., Popkova, E.G., Eds.; Palgrave Macmillan: Cham, Switzerland, 2022. [CrossRef]
6. Othman, M.E.F.; Sidek, L.M.; Basri, H.; El-Shafie, A.; Ahmed, A.N. Climate challenges for sustainable hydropower development and operational resilience: A review. *Renew. Sustain. Energy Rev.* **2025**, *209*, 115108. [CrossRef]
7. Viers, J.H. Hydropower relicensing and climate change. *J. Am. Water Resour. Assoc.* **2011**, *47*, 655–661. [CrossRef]
8. Falchetta, G.; David, E.H.J.; Gernaat, J.H.; Sebastian, S. Hydropower dependency and climate change in sub-Saharan Africa: A nexus framework and evidence-based review. *J. Clean. Prod.* **2019**, *231*, 1399–1417. [CrossRef]
9. Neupane, P.; Khanal, A.; Kafle, M.; Bhattarai, S.; Khanal, A. Climate change influence on hydropower energy output variability. *J. Appl. Water Eng. Res.* **2025**, *13*, 251–268. [CrossRef]
10. Vliet, M.V.; Wiberg, D.; Leduc, S.; Riahi, K. Power-generation system vulnerability and adaptation to changes in climate and water resources. *Nat. Clim. Change* **2016**, *6*, 375–380. [CrossRef]
11. Shu, J.; Qu, J.J.; Motha, R.; Dong, D.F. Impacts of climate change on hydropower development and sustainability: A review. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *163*, 012126. [CrossRef]
12. Zhang, Q.; Kang, S.; Chen, F. Glacier variations in the Fedchenko basin, Tajikistan, 1992–2006: Insights from remote-sensing images. *Mt. Res. Dev.* **2014**, *34*, 56–65. [CrossRef]
13. Guay, C.; Minville, M.; Braun, M. A global portrait of hydrological changes at the 2050 horizon for the province of Québec. *Can. Water Resour. J. Rev. Can. Des Ressources Hydr.* **2015**, *40*, 285–302. [CrossRef]
14. Fu, X.; Wang, X.; Gong, Y.; Wang, Y.; Zhang, Y. Impact of snow weather on PV power generation and improvement of power forecasting. In Proceedings of the 2023 International Conference on Power Energy Systems and Applications (ICoPESA), Nanjing, China, 24–26 February 2023; pp. 448–453. [CrossRef]
15. Chen, F.; Chen, J.; Huang, W.; Chen, S.; Huang, X.; Jin, L.; Jia, J.; Zhang, X.; An, C.; Zhang, J.; et al. Westerlies Asia and monsoonal Asia: Spatiotemporal differences in climate change and possible mechanisms on decadal to sub-orbital timescales. *Earth-Sci. Rev.* **2019**, *192*, 337–354. [CrossRef]
16. Dilinuer, T.; Yao, J.-Q.; Chen, J.; Mao, W.-Y.; Yang, L.-M.; Yeernaer, H.; Chen, Y.-H. Regional drying and wetting trends over Central Asia based on Köppen climate classification in 1961–2015. *Adv. Clim. Change Res.* **2021**, *12*, 363–372. [CrossRef]

17. Woolway, R.I.; Kraemer, B.M.; Lenters, J.D.; Merchant, C.J.; O'Reilly, C.M.; Sharma, S. Global lake responses to climate change. *Nat. Rev. Earth Env.* **2020**, *1*, 388–403. [CrossRef]
18. Li, Z.; Li, Y.; Li, H.; Liu, Y.; Wang, C. Analysis of drought change and its impact in Central Asia. *Adv. Earth Sci.* **2022**, *37*, 37–50. [CrossRef]
19. Jalilov, S.M.; Keskinen, M.; Varis, O.; Amer, S.; Ward, F.A. Managing the water–energy–food nexus: Gains and losses from new water development in Amu Darya river basin. *J. Hydrol.* **2016**, *539*, 648–661. [CrossRef]
20. Yu, J.; Gao, B.; Li, M.; Xiao, P. Improving runoff modelling through strengthened snowmelt and glacier module enhances runoff attribution in a large watershed in Central Asia. *J. Hydrol.* **2025**, *660*, 133528. [CrossRef]
21. Tan, Z.; Liu, H.; Gao, F. Experimental research on reduction measures of sediment Deposition of the power plant port under the long period wave. *Procedia Eng.* **2015**, *116*, 229–236. [CrossRef]
22. Timmerman, J.G.; Koepfel, S.; Bernardini, F.; Buntzma, J.J. Adaptation to climate change: Challenges for transboundary water management. In *The Economic, Social and Political Elements of Climate Change*; Climate Change Management; Leal Filho, W., Ed.; Springer: Berlin/Heidelberg, Germany, 2011. [CrossRef]
23. Abuduwaili, J.; Issanova, G.; Saparov, G. Water resources in Tajikistan. In *Hydrology and Limnology of Central Asia*; Water Resources Development and Management; Springer: Singapore, 2019. [CrossRef]
24. Fallah, B.; Didovets, I.; Rostami, M.; Hamidi, M. Climate change impacts on Central Asia: Trends, extremes and future projections. *Int. J. Climatol.* **2024**, *44*, 3191–3213. [CrossRef]
25. Chen, M.; Zeng, S.; Yang, L.; Ren, Y.; Xia, J. Statistical insights into the water exchange process in the Yangtze-Poyang System: The YPWES model. *J. Hydrol. Reg. Stud.* **2022**, *40*, 101054. [CrossRef]
26. Bilal, H.; Siwar, C.; Mokhtar, M.B.; Lahlou, F.Z.; Kanniah, K.D.; Al-Ansari, T. Snow runoff modelling in the upper Indus River Basin and its implication to energy water food nexus. *Ecol. Model.* **2024**, *498*, 110871. [CrossRef]
27. Williams, J.R. Sediment-yield prediction with universal equation using runoff energy factor. In *Present and Prospective Technology for Predicting Sediment Yield and Sources*; US Department of Agriculture, Agriculture Research Service: Washington, DC, USA, 1975; pp. 244–252.
28. Sadeghi, S.H.R.; Gholami, L.; Darvishan, A.K.; Saeidi, P. A review of the application of the MUSLE model worldwide. *Hydrol. Sci. J.* **2014**, *59*, 365–375. [CrossRef]
29. Yang, L.; Gao, M.; Chen, J.; Shi, W.; Hou, C.; Liu, Z.; Luo, C.; Yu, J.; Yang, X.; Dong, J. Effects of climate variables and human activities on groundwater level fluctuations in unconsolidated sedimentary aquifers: A data-driven approach. *Hydrology* **2025**, *12*, 215. [CrossRef]
30. Peng, Q.; Wang, R.; Jiang, Y.; Zhang, W.; Liu, C.; Zhou, L. Soil erosion in Qilian Mountain National Park: Dynamics and driving mechanisms. *J. Hydrol. Reg. Stud.* **2022**, *42*, 101144. [CrossRef]
31. Wischmeier, W.H.; Mannering, J.V. Relation of soil properties to its erodibility. *Soil Sci. Soc. Am. Proc.* **1963**, *33*, 131–137. [CrossRef]
32. Williams, J.R.; Renard, K.G.; Dyke, P.T. A new method for assessing erosions effect on soil Productivity. *J. Soil Water Conserv.* **1983**, *38*, 38–383. [CrossRef]
33. Zhang, M.; Wei, J.C.; Yang, Q.K.; Baartman, J.E.M.; Gai, L.; Yang, X.; Li, S.Q.; Yu, J.; Ritsema, C.J.; Geissen, V. An improved method for calculating slope length( $\lambda$ ) and the LS parameters of the revised universal soil loss equation for large watersheds. *Geoderma* **2017**, *308*, 36–45. [CrossRef]
34. Cai, C.; Ding, S.; Shi, Z.; Huang, L.; Zhang, G.Y. Study of applying USLE and geographical information system IDRISI to predict soil erosion in small watershed. *J. Soil Water Conserv.* **2000**, *14*, 19–24.
35. Bhagwat, N.; Kumar, R.; Qureshi, M.; Nagisetty, R.M.; Zhou, X. The Status, Applications, and Modifications of the Snowmelt Runoff Model (SRM): A Comprehensive Review. *Hydrology* **2025**, *12*, 156. [CrossRef]
36. Phien, H.N.; Ajirajah, T.J. Applications of the log Pearson type-3 distribution in hydrology. *J. Hydrol.* **1984**, *73*, 359–372. [CrossRef]
37. Paul, J.; Pilon, K. Adamowski, Asymptotic variance of flood quantile in log Pearson Type III distribution with historical information. *J. Hydrol.* **1993**, *143*, 481–503. [CrossRef]
38. Asghari, F.B.; Mohammadi, A.A.; Dehghani, M.H.; Yousefi, M. Data on assessment of groundwater quality with application of ArcGIS in Zanjan, Iran. *Data Brief.* **2018**, *18*, 375–379. [CrossRef] [PubMed]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## Article

# Quantifying the Contribution of Global Precipitation Product Uncertainty to Ensemble Discharge Simulations and Projections: A Case Study in the Liujiang Catchment, Southwest China

Yong Chang <sup>1,\*</sup>, Nan Mu <sup>1</sup>, Yaoyong Qi <sup>2</sup> and Ling Liu <sup>3</sup>

<sup>1</sup> School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China; 231309080016@hhu.edu.cn

<sup>2</sup> HeNan Third Geology and Mineral Survey Institute Co., Ltd., Zhengzhou 450000, China; 211309010003@hhu.edu.cn

<sup>3</sup> Geological Survey of Jiangsu Province, Nanjing 210018, China; liuling\_jsddy@163.com

\* Correspondence: yongchang@hhu.edu.cn

**Abstract:** Reliable precipitation inputs are essential for hydrological modeling, yet global precipitation products often exhibit substantial discrepancies that introduce significant uncertainties into streamflow simulations and projections. In this study, we assessed the relative contribution of precipitation dataset uncertainty to discharge simulations and projections, in comparison with uncertainties from model structure, model parameters, and climate projections, in the Liujiang catchment, southwest China. Three widely used satellite-based products (CHIRPS, PERSIANN, and IMERG) and one reanalysis dataset (ERA5) were combined with three hydrological models of varying structural complexity to simulate streamflow. Using an ANOVA-based variance decomposition framework, we quantified the contributions of different uncertainty sources under both historical and future climate conditions. Results showed that precipitation input uncertainty dominates discharge simulations during the calibration period, contributing over 60% of total variance particularly at high flows, while interactions among precipitation, model structure, and parameters govern low-flow simulations. Under future climate scenarios, climate projection uncertainty overwhelmingly dominates discharge predictions with 50–80% of uncertainty contribution, yet precipitation products still contribute significantly across time scales. The compensation of precipitation biases by hydrological models can cause parameter values to deviate from their true physical meaning. This deviation may further amplify the differences in discharge projections driven by different precipitation products under future climate conditions and increase the overall uncertainty of streamflow projections. Overall, this study introduced an integrated approach to simultaneously assess precipitation uncertainty across flow regimes and future climate scenarios. These results emphasized the necessity of using ensemble approaches that incorporate multiple precipitation products in hydrological forecasting and impact studies, particularly in data-scarce regions reliant on global datasets.

**Keywords:** global precipitation product; precipitation uncertainty; streamflow projection; uncertainty analysis

## 1. Introduction

Precipitation, as the primary driver of the hydrological cycle, is essential for accurate hydrological modeling, flood forecasting, and water resources management [1–3].



However, in data-scarce or ungauged regions, obtaining reliable precipitation observations remains a big challenge. In such cases, the global precipitation datasets, particularly satellite and reanalysis precipitation products, provide an important alternative [4]. These datasets can offer consistent, near-global coverage with relatively high temporal and spatial resolution, enabling hydrological simulations even in remote, mountainous, and rural areas where ground-based observations are limited or unavailable. Nowadays, a number of global precipitation datasets have been released and available for research related to water resources [5,6].

Although global precipitation products were widely used in the streamflow simulation, they still contain substantial inherent uncertainties, and significant differences exist among different products [7–9]. These uncertainties mainly arise from sensor measurement errors, the frequency of sampling, retrieval algorithms and the representation of cloud physics [10–12]. In recent years, a number of studies have compared satellite and reanalysis-based precipitation products against gauge observations to assess their performance across different regions [13,14]. Overall, these studies demonstrate the performance of different precipitation products varies substantially across climatic and topographic regions, and no single product consistently outperforms others globally [11,15]. Furthermore, when these precipitation products are applied in the hydrological simulation, precipitation errors can be partly compensated through parameter calibration to some extent [16–19], leading to strong interactions between precipitation inputs and model structures. Such interactions further complicate the relationship between precipitation datasets and simulated streamflow, making it challenging to identify an optimal precipitation product for a specific region. Consequently, the use of global precipitation products for streamflow prediction in data-scarce or ungauged basins inevitably involves considerable uncertainty. A systematic understanding of how the uncertainty from the different global precipitation datasets contributes to streamflow simulation uncertainty is therefore essential for improving hydrological prediction and supporting effective water resources management.

Precipitation uncertainty represents a critical component of overall streamflow simulation and forecasting uncertainty, and its influence on hydrological model performance has been widely investigated. Typically, there are three primary approaches to measure or estimate precipitation: ground-based rain gauges, weather radar, and satellite-based sensors [14,20,21]. Rain gauges remain the most common source of observed precipitation data and are often used as the reference for validation. However, gauges only provide point-scale measurements, which may fail to adequately capture the spatial heterogeneity of rainfall fields, especially in regions with complex topography or sparse observation networks. As a result, numerous studies have examined the effects of gauge density and spatial distribution on streamflow simulations and highlighted their significant impacts [22–24], leading to efforts aimed at optimizing gauge network design to reduce representativeness errors. In contrast, radar-derived rainfall estimates are obtained remotely and indirectly, and are therefore subject to uncertainties arising from hardware limitations, signal processing, and retrieval algorithms [25,26]. Previous research has documented that radar rainfall uncertainties, including random errors, systematic biases, and spatiotemporal dependencies, can propagate through hydrological models and exert substantial impacts on simulated streamflow [21,27,28].

For satellite-based precipitation products, given the wide range of precipitation datasets now available, most studies have focused on intercomparison of different products by evaluating their ability to drive hydrological simulations and attempting to identify the most suitable product for specific catchments [10,29–31]. Only a limited number of studies have directly investigated the contribution of precipitation product uncertainty to overall

streamflow simulation uncertainty and compared it to other sources of uncertainty. Ma et al. (2018) [32] quantified the contributions of input uncertainty from three precipitation products in rainfall–runoff simulations and demonstrated that input uncertainty dominated the total predictive uncertainty in the distributed Coupled Routing and Excess Storage (CREST) model. Whereas for another lumped model, structure uncertainty was the leading source of total uncertainty. Qi et al. (2019) [33] evaluated six high-resolution precipitation products in a small basin in northern China and quantified the contributions of precipitation product and model structure uncertainty to discharge simulations. Their results showed that precipitation input was the dominant source of uncertainty in medium-magnitude flows, whereas for small and large discharges, hydrological models contribute most of the uncertainties. Although these studies provide valuable preliminary insights into the role of precipitation product uncertainty in streamflow simulations, how the precipitation dataset interacts with other uncertainty factors, e.g., model structure and parameter, and contribute to the uncertainty of discharge simulation is still needed further explored. Moreover, these studies mainly focused on discharge simulations under historical conditions. To the best of our knowledge, very few studies have explicitly examined the contribution of global precipitation product uncertainty to discharge prediction under future climate scenarios.

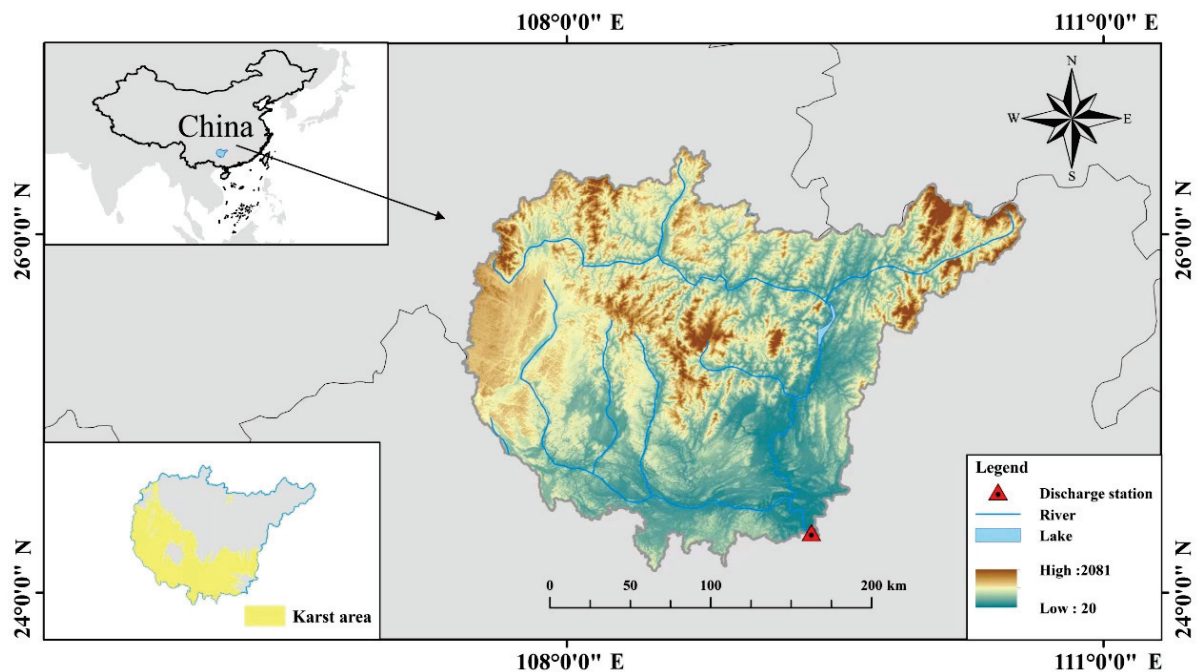
In this paper, the contribution of global precipitation dataset uncertainty to ensemble discharge simulation and projection uncertainty was investigated in the Liujiang catchment, located in southwest China. Three widely used global satellite precipitation products (PERSIANN, IMERG and CHIRPS) and a reanalysis precipitation dataset ERA5 are evaluated in three different hydrological models of different structures for the streamflow simulation and projections. Following the comparison of simulation results driven by different precipitation inputs, an analysis of variance (ANOVA)-based error decomposition framework is applied to quantify the relative contributions of precipitation datasets, model structures, model parameters, climate projections, and their interactions to the total uncertainty during both the historical calibration period and future climate scenarios. This work would enhance our understanding of how precipitation uncertainty propagates through hydrological models to influence ensemble discharge predictions and clarifies the role of global precipitation products as a source of uncertainty in streamflow simulation and projection, particularly in data-scarce or ungauged catchments.

## 2. Description of Study Site and Datasets

The Liuzhou catchment, located in the middle–upper reaches of the Pearl River Basin, is the second largest tributary of the Xijiang River. It originates from Dushan County, Guizhou Province, with a drainage area of approximately 46,000 km<sup>2</sup> (Figure 1). The basin exhibits an overall topographic gradient from high elevations in the northwest to lower elevations in the southeast. The upstream region, except for the river source area, is predominantly characterized by metamorphic rock–dominated gorges with elevations generally exceeding 1000 m. The middle and lower reaches are dominated by karst landforms, with a large proportion of mountainous and hilly terrain, well-developed surface karst features, and extensive underground river networks. The mean elevation ranges from 500 to 1000 m, with cultivated land concentrated in certain areas and a relatively dense population. Overall, karst terrain accounts for approximately 41.14% of the total basin area.

The basin is situated in a typical subtropical monsoon climate zone, with a mean annual temperature of 17–20 °C and annual precipitation ranging from 1200 to 1800 mm. Although the basin receives abundant rainfall, its temporal distribution is highly uneven, with approximately 70% of the annual precipitation occurring between April and September. To simulate the streamflow of Liujiang catchment, this study employed three widely

used satellite-based precipitation datasets, CHIRPS, PERSIANN, and IMERG, along with a reanalysis dataset ERA5.



**Figure 1.** Location of Liujiang catchment.

CHIRPS (Climate Hazards Group Infrared Precipitation with Stations) [34] is a quasi-global ( $50^{\circ}$  S– $50^{\circ}$  N), land-only precipitation product with a high spatial resolution of  $0.05^{\circ}$ . It used a novel interpolation procedure to blend satellite-derived cold cloud duration rainfall estimates with in situ gauge observations using a two-phase smart interpolation approach. CHIRPS provides daily to monthly precipitation data ranging from 1981 to near-present with relatively low latency ( $\sim 2$  days for preliminary products and  $\sim 3$  weeks for final products), and low bias, making it particularly useful in data-scarce regions.

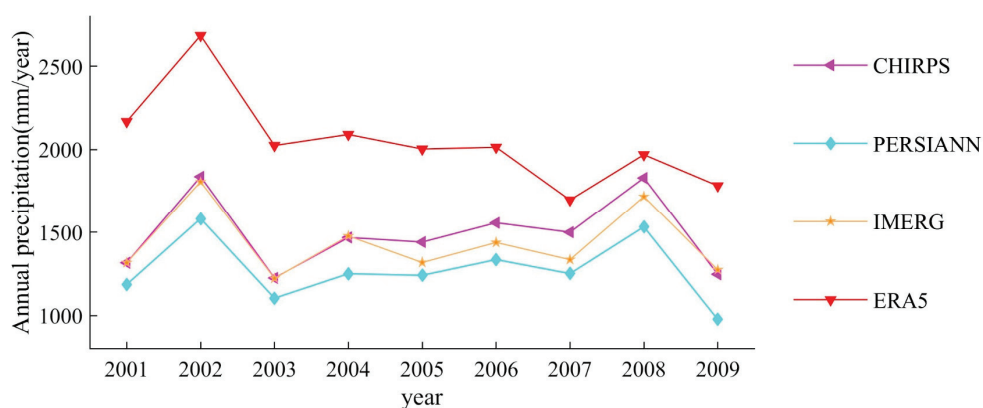
PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) [35] is a long-term, near-global ( $60^{\circ}$  S– $60^{\circ}$  N),  $0.25^{\circ}$  daily precipitation dataset generated using an artificial neural network applied to GridSat-B1 infrared satellite observations. A bias-adjustment procedure using GPCP monthly precipitation ensures temporal consistency and accuracy while retaining fine-scale spatial and temporal variability. This dataset maintains stable algorithm parameters throughout its record, allowing for consistent characterization of both convective and stratiform rainfall.

IMERG (Integrated Multi-satellite Retrievals for GPM) [36] is a NASA product estimating global surface precipitation rates at a high resolution of  $0.1^{\circ}$  every half-hour beginning 2000 under the GPM mission. It was generated from combining passive microwave (PMW) and infrared data of the GPM constellation satellites and calibrated by gauge analysis of the Global Precipitation Climatology Center (GPCC). The fine spatial and temporal resolution of IMERG data allows them to be accumulated to the scale of the application for increased skill, which can be used for global-scale applications as well as over regions with sparse or no reliable surface observations.

ERA5 [37] is the fifth-generation ECMWF atmospheric reanalysis, spanning 1950 to the present with a spatial resolution of 31 km, 137 vertical levels, and hourly output. It is produced using the IFS Cy41r2 forecast model with a 12 h 4D-Var assimilation system, complemented

by a 10-member Ensemble of Data Assimilations to quantify uncertainty. Compared with its predecessor ERA-Interim, ERA5 incorporates higher resolution, more variables, enhanced model physics, and an expanded range of reprocessed satellite and in situ observations.

The differences among these four datasets primarily stem from their different data sources, retrieval algorithms, and processing methodologies. CHIRPS relies mainly on infrared satellite observations blended with gauge data, while IMERG combines multi-satellite microwave and infrared estimates with gauge-based bias correction. PERSIANN estimates precipitation from infrared brightness temperatures via a neural network, followed by a bias-adjustment procedure using GPCP dataset. In contrast, ERA5 derives precipitation fields from a physically consistent numerical weather prediction model assimilating multi-source observations. The four precipitation datasets exhibit distinct temporal and spatial resolutions: CHIRPS ( $0.05^\circ$ , daily), PERSIANN ( $0.25^\circ$ , daily), IMERG ( $0.1^\circ$ , half-hourly), and ERA5 ( $0.25^\circ$ , hourly). These differences in resolution, along with variations in data sources (e.g., infrared vs. microwave sensors, gauge correction methods, and model assimilation techniques), represent potential sources of uncertainty in rainfall estimates, particularly in capturing the spatial heterogeneity and intensity distribution of precipitation, especially in complex terrain. For all datasets, the daily precipitation data were downloaded and converted to daily areal averages based on the catchment boundary. These processed data then served as the precipitation input for the hydrological models. Figure 2 presents the annual precipitation of the Liujiang catchment from 2001 to 2009 as estimated by four precipitation datasets. All datasets show the similar interannual variability of precipitation, but with notable differences in magnitude. ERA5 consistently produces the highest estimates throughout the period, and the overestimation is significant from 2001 to 2006. In contrast, CHIRPS and IMERG provide very similar annual precipitation. PERSIANN generally provides the lowest annual precipitation in most years compared to other datasets. These differences highlight the systematic biases among precipitation datasets, which may have important impact on streamflow simulations and projections.

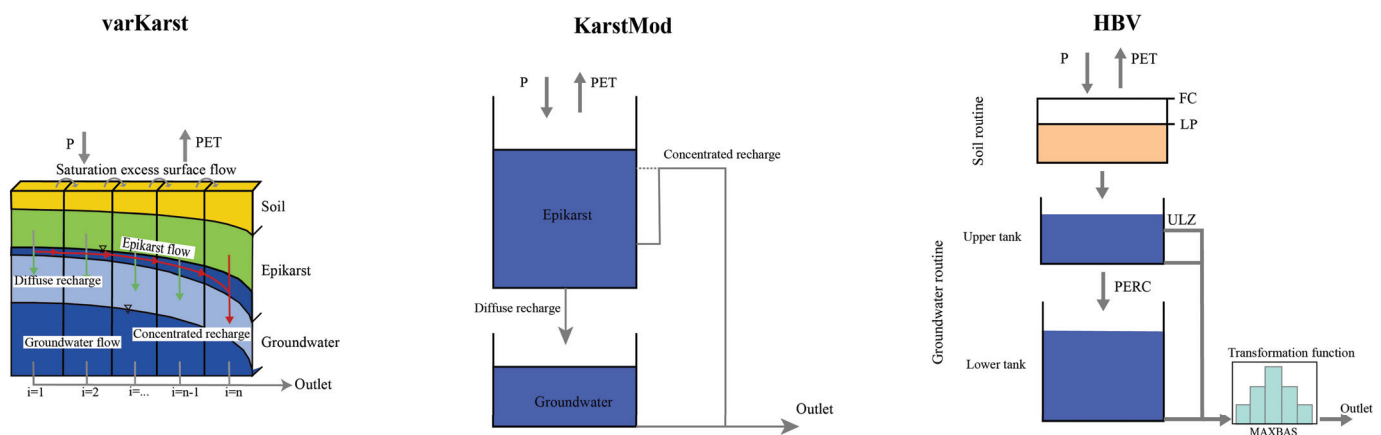


**Figure 2.** Annual precipitation comparison of different precipitation products from 2001 to 2009.

### 3. Methodology

#### 3.1. Three Different Hydrological Models

Currently, a wide range of hydrological models have been developed and applied for simulating catchment discharge. Given the extensive carbonate rock distribution in the study region and the complex hydrological processes characteristic of karst aquifers, two widely used karst-specific models, varKarst and KarstMod, were selected (Figure 3). In addition, the conceptual HBV model, originally developed for non-karst catchments, was also included to account for structural uncertainty in the ensemble analysis.



**Figure 3.** Structures of three karst hydrological models.

The varKarst model [38] is a semi-distributed karst hydrological model that partitions the aquifer into three compartments: the soil layer, the epikarst, and the saturated groundwater zone. The soil layer controls evapotranspiration, infiltration, and saturation overland flow generation when its maximum storage is exceeded. Infiltration and saturation excess water recharges the epikarst linear reservoir and the epikarst then transfers water to the saturated zone through two pathways: concentrated recharge to a conduit reservoir and diffuse recharge to a fissure reservoir. The total simulated discharge is the sum of flows from both reservoirs. A key feature of varKarst is the use of a Pareto function to represent the spatial heterogeneity of soil and epikarst thickness and hydraulic properties, enabling realistic simulation of aquifer heterogeneity. This model has been widely applied to karst spring and catchment studies.

KarstMod is a flexible karst modeling platform [39], with the structure applied here originally proposed by [40]. The model consists of two reservoirs: one for the soil/epikarst zone and another for the saturated zone. The epikarst reservoir receives precipitation input and accounts for evapotranspiration losses. It has two outlets: a fast-flow outlet and a slow recharge outlet to the lower reservoir. Fast flow is modeled using a nonlinear hysteretic function to capture changing hydrological connectivity under different saturation conditions. The simulated discharge is the sum of fast flow from the epikarst reservoir and flow from the saturated reservoir.

The HBV model comprises soil and groundwater routines. The soil routine estimates actual evapotranspiration (AET) and effective recharge. Groundwater is represented by two interconnected tanks, with the upper tank producing overflow when a threshold is exceeded, generating both fast and slow flow components. The combined outflows are routed through a triangular weighting function to produce catchment discharge.

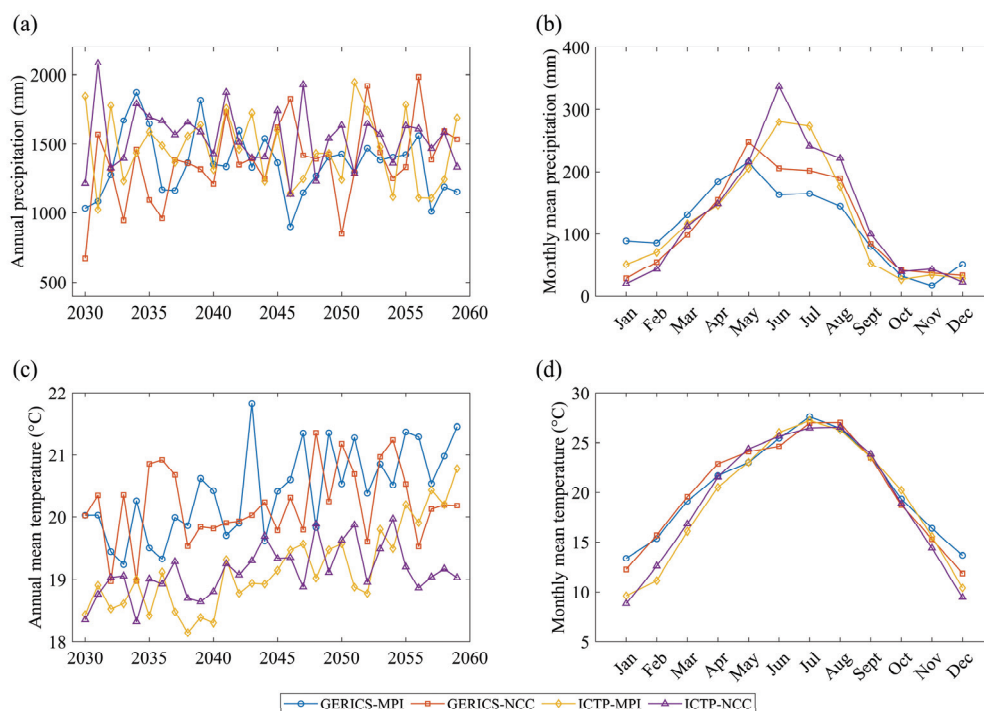
The three selected hydrological models have fundamentally different structural designs and representations of internal hydrological processes, which allows for a comprehensive assessment of model structural uncertainty. For each model, PET is estimated by a simplified Penman-Monteith equation proposed by Huang and Chen (2014) [41], which only needs daily minimum and maximum temperature. These temperature data during the simulation period are calculated from the meteorological stations within the catchment. For all models, the measured daily discharges from 1998 to 2009 were used for model calibration with the first three years of data as the warm-up period to eliminate the uncertainty from the initial conditions.



### 3.2. Future Climate Projections

To quantitatively assess the contribution of precipitation data uncertainty to the overall uncertainty in future discharge predictions, four sets of climate projections from the CORDEX–East Asia initiative (GERICS-MPI, GERICS-NCC, ICTP-MPI, and ICTP-NCC) were selected to represent the uncertainty from climate model projections. These datasets are generated by different institutions using distinct driving global climate models, thereby capturing the combined effects of institutional and model-related differences on future precipitation and temperature estimates. In this paper, the daily precipitation and temperature data under the RCP8.5 gas emission scenario from 2030 to 2059 were used to drive hydrological models for future discharge simulation and uncertainty decomposition.

Figure 4 presents the annual and monthly mean precipitation and temperature of the four climate projections during the future period. For annual precipitation, all products exhibit pronounced interannual variability and noticeable differences in magnitude. ICTP-MPI and ICTP-NCC generally project wetter conditions, whereas GERICS-MPI tends to give the lowest annual precipitation in multiple years. All products show a clear seasonal pattern in monthly precipitation, with peaks during the rainy season from May to August. These wet months also correspond to the period with the largest differences in monthly precipitation among different projections. In June, the largest difference between ICTP-NCC and GERICS-MPI is about 140 mm. Annual mean temperature projections indicate a consistent warming trend across all four products, but with differences in both the rate of increase and the baseline temperature values. GERICS-MPI and GERICS-NCC consistently simulate higher annual mean temperatures compared to ICTP-MPI and ICTP-NCC. For monthly mean temperature, their differences are most apparent during the dry season from November to March, with GERICS-MPI and GERICS-NCC projecting noticeably warmer winter conditions than GERICS-MPI and GERICS-NCC.



**Figure 4.** Variations in annual and average monthly precipitation in the future period (2030–2059), (a) Annual precipitation, (b) Monthly mean precipitation, (c) Annual mean temperature, (d) Monthly mean temperature.

### 3.3. ANOVA Method

The analysis of variance (ANOVA) was applied to quantify the relative contribution of different sources of uncertainty to discharge simulation and prediction [32,42]. In this study, we separately assessed the role of precipitation uncertainty during the calibration period and in future discharge projections. For the calibration period, three major sources of uncertainty were considered: (1) precipitation input (four datasets: CHIRPS, PERSIANN, IMERG, and ERA5); (2) model structural uncertainty (three models: varKarst, KarstMod, and HBV); (3) parameter uncertainty. For the discharge prediction under future climate conditions, the extra climate prediction uncertainty was further included which contains four climate projections (GERICS-MPI, GERICS-NCC, ICTP-MPI, and ICTP-NCC).

Model performance for each combination of factors was evaluated using the Kling–Gupta Efficiency (KGE) as the objective function, which jointly accounts for correlation, variability, and bias between simulated and observed discharge. Parameter uncertainty was addressed using a Monte Carlo sampling approach, in which 100,000 parameter sets were randomly sampled from the feasible range of each model. For each model, only the top 1000 parameter sets ranked by KGE were retained to represent the parameter uncertainty and to predict the discharge in the future projection.

Based on the ANOVA theory, total error variance can be decomposed into the contributions from each factor's main effect and their interactions. For illustration, considering three factors in the calibration period, total error variance (SST) of discharge simulations can be described as

$$SST = SS_A + SS_B + SS_C + \sum SS_{ABC} \quad (1)$$

where  $SS_A$ ,  $SS_B$  and  $SS_C$  are error contribution of precipitation uncertainty, model structure uncertainty and parameter uncertainty, respectively.  $\sum SS_{ABC}$  is the total error contribution of their interactions.

Each term in Equation (1) can be estimated as follows.

$$SST = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left( Y^{i,j,k} - Y^{0,0,0} \right)^2 \quad (2)$$

$$SS_A = J \cdot K \cdot \sum_{i=1}^I \left( Y^{i,0,0} - Y^{0,0,0} \right)^2 \quad (3)$$

$$SS_B = I \cdot K \cdot \sum_{j=1}^J \left( Y^{0,j,0} - Y^{0,0,0} \right)^2 \quad (4)$$

$$SS_C = I \cdot J \cdot \sum_{k=1}^K \left( Y^{0,0,k} - Y^{0,0,0} \right)^2 \quad (5)$$

$$\sum SS_{ABC} = SST - SS_A - SS_B - SS_C \quad (6)$$

where  $I$ ,  $J$  and  $K$  represent the number of precipitation datasets, models and parameter sets, respectively.  $Y^{0,0,0}$  is the average KGE among all the simulations. The symbol  $i$  indicates the average KGE over a particular index. Then, the uncertainty contribution ( $\eta$ ) of each factor is calculated by the following equations.

$$\eta_{precipitation}^2 = \frac{SS_A}{SST} \times 100\% \quad (7)$$

$$\eta_{model}^2 = \frac{SS_B}{SST} \times 100\% \quad (8)$$

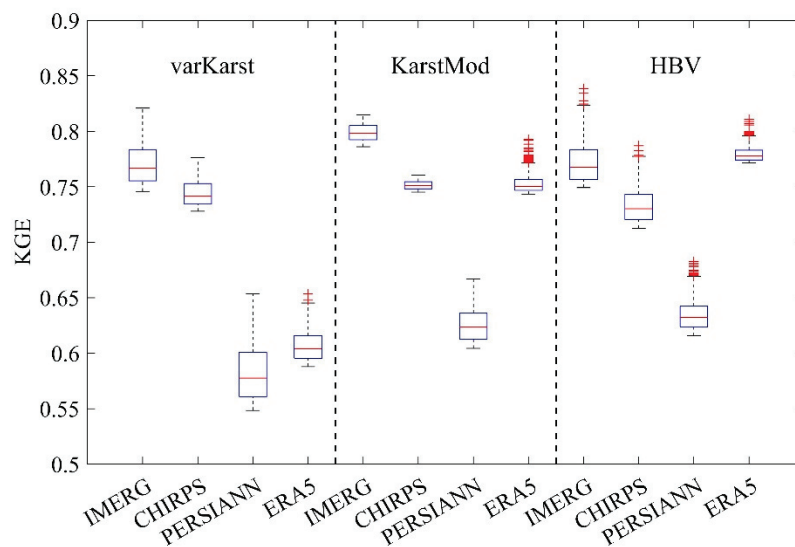
$$\eta_{parameter}^2 = \frac{SS_C}{SST} \times 100\% \quad (9)$$

$$\eta_{interaction}^2 = \frac{\sum SS_{ABC}}{SST} \times 100\% \quad (10)$$

## 4. Results

### 4.1. Calibration Results of Different Precipitation Products

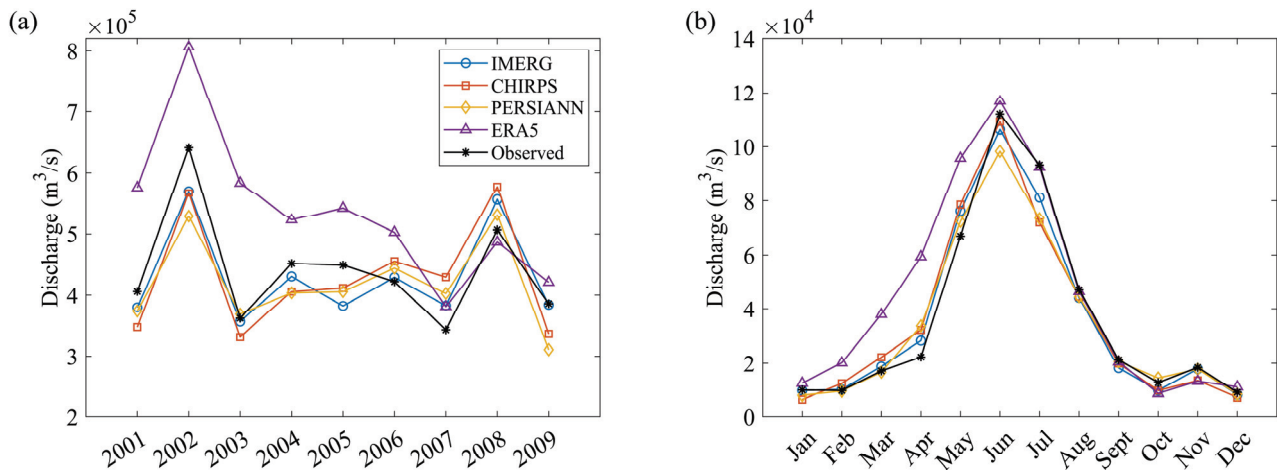
The calibration performance, evaluated using the objective KGE, of the three hydrological models driven by four precipitation datasets is presented in Figure 5. For each model, the top 1000 parameter realizations were retained to represent parameter uncertainty. The results show that the choice of precipitation dataset has a significant impact on discharge simulations and the same precipitation input can yield markedly different results across different models. For varKarst, IMERG provides the highest median KGE (about 0.77), while PERSIANN and ERA5 give the relatively bad simulation result with median values of KGE below 0.65. KarstMod achieves relatively high performance with IMERG, CHIRPS, and ERA5, but its efficiency drops substantially when driven by PERSIANN. In contrast, HBV exhibits intermediate performance among three models with a relatively smaller discrepancies among rainfall products. ERA5 slightly outperforms IMERG, whereas PERSIANN yields the lowest KGE values. Overall, IMERG delivers the most consistent and reliable performance across all three models, while PERSIANN consistently shows the worst results. The performance of ERA5 varies across different models highlighting the critical role of interactions between precipitation datasets and model structures. These findings underscore the importance of both precipitation dataset selection and model structural choice in achieving good discharge simulations.



**Figure 5.** Calibration results of three hydrological models driven by different precipitation datasets.

Figure 6 compares the average discharge simulations of the three hydrological models driven by four precipitation datasets at both interannual and monthly scales. At the interannual scale, IMERG, CHIRPS, and PERSIANN yield relatively consistent annual streamflow estimates that are close to the observations, whereas ERA5 produces substantially higher discharge during 2001–2006, likely due to its overestimation of annual precipitation compared with the other datasets (Figure 2). At the monthly scale, all products capture the pronounced streamflow seasonality, with peak flows from May to July and low flows during November to February. However, ERA5 systematically overestimates discharge from February to July, while IMERG, CHIRPS, and PERSIANN provide results that are more consistent with the observations. Compared with the calibration results in Figure 5,

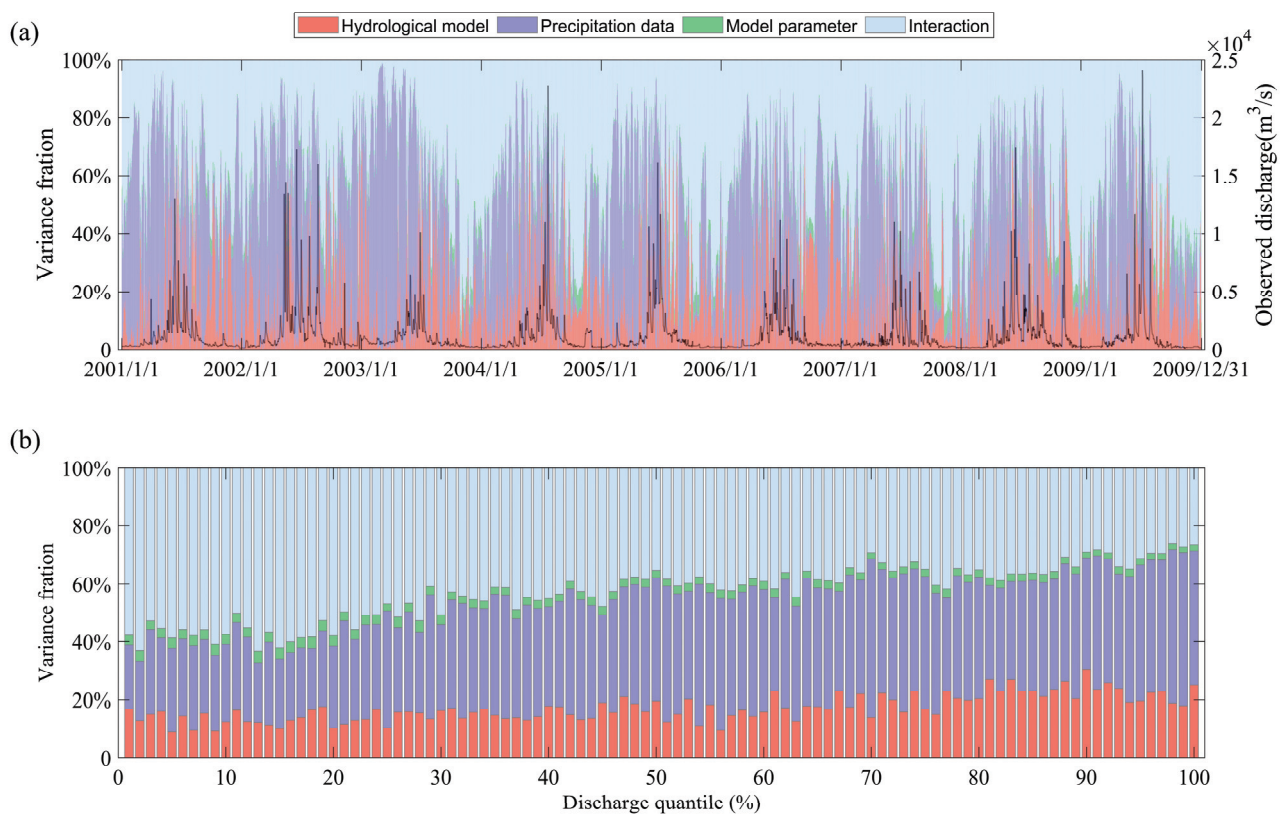
PERSIANN exhibits relatively low KGE values, indicating that its poor streamflow performance is mainly related to daily scale biases, although its long-term water balance is reasonably reproduced at annual and monthly scales. In contrast, ERA5 achieves relatively high KGE values in KarstMod and HBV (Figure 5), but it consistently overestimates discharge at both interannual and seasonal scales (Figure 6), suggesting that its high KGE is primarily attributable to relatively better reproduction of daily discharge dynamics.



**Figure 6.** Average simulation results of three different models driven by each precipitation data: (a) annual discharge; (b) monthly discharge.

#### 4.2. Uncertainty Analysis in the Calibration Period

The AVOVA method was applied to decompose the contributions of three uncertainty sources (precipitation data, model structure and model parameter) and their interaction to the daily discharge simulations during the calibration period. Among all uncertainty factors, precipitation data provide a major contribution to the total uncertainty, particularly in the early wet period when streamflow begins to rise slowly. During this period, the uncertainty contribution from precipitation data often exceeds 60% and can even surpass 80% at certain times. The low impact of precipitation data on the discharge uncertainty mainly happens in the low discharge period, whereas for the uncertainty from the model structure, it has a higher contribution mainly in high-flow periods and its proportion can hardly exceed 50%. In most periods, the uncertainty contribution from model structure is always smaller than that from precipitation data with only exception for some small daily discharge period. In contrast, the model parameter always plays only a minor role in the discharge uncertainty, generally accounting for less than 5% of the total uncertainty, with slightly higher contributions (up to about 10%) during low-flow conditions. Compared to their individual uncertainty contributions, their interactions give a pronounced contribution to the total uncertainty during low-flow period, exceeding 60% in some low discharges. Whereas for other times, the contribution from the interaction remains below 40%. Figure 7b further illustrates how the uncertainty contribution from different factors varies with discharge. As discharge increases, the contribution of precipitation data and model structure to the total uncertainty rises, while the uncertainty contribution of interactions gradually declines. Parameter uncertainty remains consistently small across all flow ranges. At very low flows (below the 25th quantile), interactions dominate the total uncertainty, accounting for more than 50% of the variance. Precipitation data gradually become the dominant uncertainty source with the increase in discharge and its contribution consistently outweighs model structure.

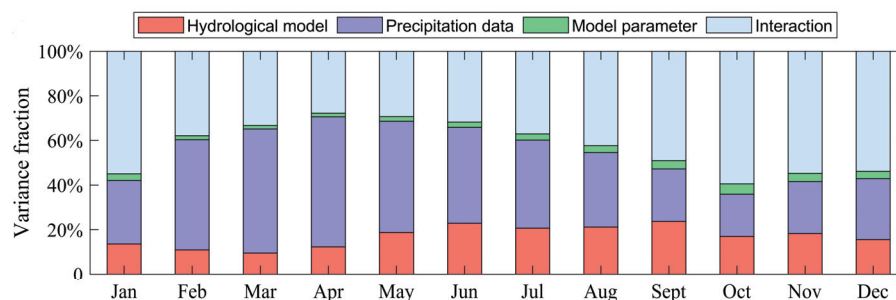


**Figure 7.** Contributions of uncertainty sources to (a) daily discharges and (b) discharge quantiles based on daily observations.

Figure 8 summarizes the monthly contributions of different uncertainty sources to discharge simulations during the calibration period. The results are consistent with the previous analysis. Precipitation data contribute most uncertainties of the simulated discharge during the transition from dry to wet months (February–June), with its maximum contribution reaching about 60% in April when runoff begins to rise rapidly. As streamflow decreases toward the dry months, the contribution of precipitation data gradually declines and reaches its minimum in low-flow months. In contrast, the contribution of interactions among uncertainty sources becomes most pronounced in dry months, exceeding 50% in October, whereas their influence is lowest in wet months. The effect of model structure is more evident during high-flow months, yet its uncertainty contribution remains much lower than that of parameter data. The precipitation data and model structure only show similar contributions in October. Model parameter uncertainty remains negligible throughout all months, always below 5%, with slightly higher values during dry months than in wet months. The dominance of precipitation uncertainty during the rising limb and high-flow periods can be attributed to the high sensitivity of runoff generation to rainfall intensity and spatial distribution during these phases. In contrast, during low-flow periods, streamflow is primarily governed by slower groundwater discharge and basin storage release processes, which are more strongly influenced by the complex interactions between model structures and parameters, hence the larger contribution from interaction uncertainty.

Overall, these results indicate that precipitation data dominate the simulation uncertainty during the discharge rising and wet-flow periods, whereas interactions are the primary source of uncertainty under low-flow conditions. Model structure exerts a secondary but less pronounced influence, while model parameters consistently contribute only marginally to the overall discharge uncertainty.

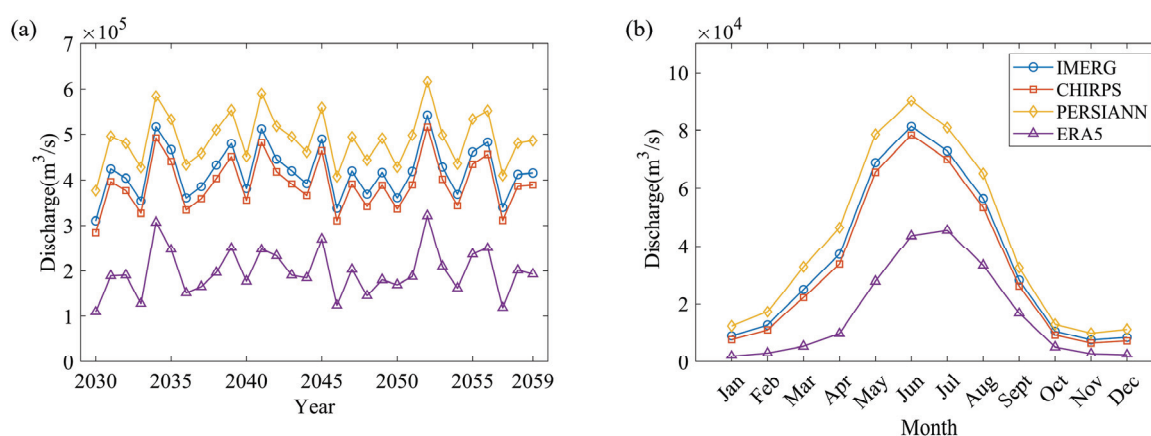




**Figure 8.** Contributions of different uncertainty sources to discharge simulations at the monthly scale during the calibration period.

#### 4.3. Uncertainty Analysis in the Future Period

Figure 9 presents the average prediction results of the three hydrological models calibrated with different precipitation datasets under future scenarios. In contrast to the simulation results during the calibration period (Figure 6), the ERA5-calibrated models predict substantially lower annual and monthly discharges than those calibrated with the other datasets, whereas the PERSIANN-calibrated models produce the highest values. The predictions of CHIRPS- and IMERG-calibrated models are relatively close to each other. These prediction differences from ERA5 and PERSIANN calibration models mainly arise from the compensating effect of parameter calibration with respect to precipitation biases. Specifically, among the four precipitation datasets, ERA5 generally has much higher precipitation estimations (Figure 2). To achieve satisfactory calibration performance, ERA5-driven models require parameter adjustments during the calibration to enhance evapotranspiration, thereby reducing effective infiltration to make the simulation closer to the observed streamflow. As a result, ERA5-calibrated models tend to simulate higher evapotranspiration and consequently much lower discharge under the same meteorological conditions compared to the model calibrated by other precipitation datasets. Conversely, PERSIANN underestimates precipitation relative to other precipitation datasets. To compensate, model calibration reduces evapotranspiration through parameter adjustment, which increases effective infiltration and leads to higher predicted discharge. This indicates that although model calibration can improve performance to some extent, it may also cause model parameters to deviate substantially from their true values. Such parameter deviations can in turn lead to significant biases in streamflow predictions under future climate scenarios.



**Figure 9.** Average discharge predictions of the three hydrological models calibrated with different precipitation datasets under future scenarios (2023–2059): (a) annual discharge and (b) monthly mean discharge.

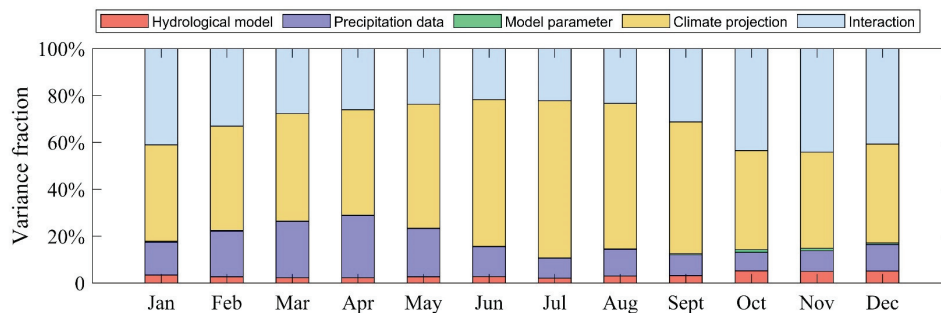
Figure 10 quantifies the sources of uncertainty in future discharge projections (2023–2059) using the ANOVA method. Climate prediction uncertainty dominates for most of the period, generally accounting for more than 50% of the variance and occasionally exceeding 80%, highlighting the critical role of climate projection spread in future streamflow predictions. Interactions among different sources provide the second-largest contribution, with higher fractions at low-flow conditions and smaller contributions at high flows. The influence of precipitation datasets is much weaker than that of climate predictions, typically remaining below 20%. The uncertainty attributable to model structure is even smaller, usually less than 10%, while model parameter uncertainty remains negligible, similar to the calibration period. Figure 10 further illustrates the variation in uncertainty contributions with discharge. As discharge increases, the uncertainty contribution from the climate projection rises steadily, exceeding 80% at the highest quantiles, whereas the relative impact of interactions declines. The variance fraction associated with precipitation data initially increases with discharge but decreases once the discharge quantile exceeds 60%. Structural and parameter uncertainties remain limited across all flow conditions, especially at high flows. Overall, climate predictions govern the majority of discharge prediction uncertainty, particularly at medium to high flows, while the interaction dominates under low-flow conditions. Precipitation datasets exert a tertiary influence which is much larger than model structure. The contribution of model parameters to the total uncertainty is consistently negligible compared to other factors.



**Figure 10.** Source analysis of discharge prediction uncertainty under the future scenarios.

The contributions of different uncertainty sources to the total uncertainty of monthly discharge predictions are shown in Figure 11. The results indicate that climate model uncertainty dominates throughout the year, typically accounting for 40–70% of the total variance, with particularly high contributions in the wet months (May–August). Interactions among uncertainty sources also contribute substantially, particularly in the dry months (October–January), when their maximum fraction reaches 40%. The influence of precipitation data is moderate, contributing around 10–20% throughout the year, with relatively higher values in the early wet season (April–June) which is similar to the calibration

period. In contrast, the uncertainties associated with hydrological model structure and model parameters remain very small, typically less than 10% in all months.



**Figure 11.** Contributions of different uncertainty sources to discharge prediction at the monthly scale under the future scenarios.

## 5. Discussion

Although the selected precipitation datasets have been bias-corrected against ground observations, differences remain due to variations in their input sources or data assimilation techniques. In this study, the four selected precipitation datasets selected also exhibit substantial discrepancies in the Liujiang catchment. When these datasets are applied for hydrological simulations, due to the adaption of parameter calibration for precipitation biases [16,43,44], some datasets have comparable streamflow simulations. However, our results also reveal that model performance differs considerably depending on both the precipitation dataset and the model structure. For example, ERA5-driven simulations perform much better with KarstMod and HBV than with varKarst, indicating that the ability of models to compensate for precipitation biases varies with structural design, which is consistent with the findings of Oudin et al. (2006) [16]. Such differences can be attributed to the representation of hydrological processes; models with more flexible structures may redistribute water fluxes more effectively to mask certain precipitation errors. In contrast, PERSIANN consistently produces poor results across all three models, suggesting that calibration cannot effectively compensate for its errors. Stephens et al. (2022) [18] investigated the types of precipitation errors that hydrological models can address and found that models are less capable of correcting errors in rainfall timing compared to errors in magnitude or variance. Thus, the poor performance of PERSIANN is most likely associated with its inadequate representation of rainfall timing, whereas the discrepancies in ERA5 are primarily attributable to its severe overestimation of rainfall magnitude.

Despite employing hydrological models with markedly different structures and internal process representations (karst-specific and non-karst models), the results indicate that the impact of precipitation dataset to discharge simulation uncertainty still outweighs both model structure and parameter uncertainty. Our finding that precipitation input uncertainty is the dominant source of uncertainty during high-flow periods aligns with Ma et al. (2018) [32] and Qi et al. (2019) [33]. This is mainly due to the substantial differences that remain among available precipitation products. As the primary driver of hydrological models, precipitation errors can directly propagate into runoff simulations, especially the systematic error and its error in rainfall occurrence [45]. Therefore, the large difference among precipitation datasets can inevitably exert a dominant impact on discharge simulation uncertainty. More specifically, the contributions of both precipitation dataset and model structure to discharge uncertainty increase with discharge magnitude, this suggests that high-flow simulations are particularly sensitive to accurate precipitation inputs and

appropriate model structures [15,46–48]. In contrast, interaction among different factors dominate the uncertainty of low-flow simulations and the uncertainty contribution from precipitation data is relatively low. This is probably due to the climate characteristic of Liujiang catchment with distinct dry and wet periods. The most low-flow periods generally occur during the dry months when rainfall is scarce and much lower than potential evapotranspiration. Under such conditions, most precipitation is consumed by evapotranspiration before infiltration, leading to limited effective infiltration and rainfall-driven variability in streamflow. Therefore, the streamflow variation is mainly related to catchment storage-release process, which is mainly controlled by the model structure, model parameter and their interaction together, rather than by direct precipitation input. Since model parameters are also closely related to precipitation data, their combined influence further amplifies the role of interactions. As a result, interactions among these factors make the largest contribution to uncertainty in low-flow simulations.

When extending the analysis to future climate scenarios, the results show that climate projection uncertainty overwhelmingly dominates discharge prediction uncertainty except for the low-flow periods. This finding is consistent with previous studies [49–52] which shows the larger uncertainty source from climate projections in the discharge prediction. The large spread among climate projections can be attributed to differences in climate model structures, internal climate variability, and regional downscaling methods [53]. Nevertheless, the precipitation dataset remains the important source of uncertainty which is much larger than the model structure and parameter. Importantly, the results also indicate that the calibration-induced parameter bias can further exacerbate predictive spread across datasets. For instance, ERA5 tends to overestimate rainfall relative to observations, which forces calibration to increase evapotranspiration-related parameters to match observed streamflow. As a consequence, in future projections, ERA5-driven models simulate much higher evapotranspiration and significantly lower discharge compared to models calibrated with other precipitation datasets. This suggests that while model adaptation to precipitation errors can enhance simulation performance during calibration, the calibrated parameters may deviate substantially from their true values [16,54], potentially amplifying biases in discharge predictions under future climate conditions. These findings underscore the critical need to account for calibration-induced parameter distortions when using reanalysis or satellite rainfall products in climate impact assessments. In addition, the interaction also gives important contributions to the total discharge projection uncertainty particularly in the low-flow period because of the climate characteristic of the study catchment and the complex interplay among precipitation input, model structure, and parameterization.

This study also has several limitations that need to be further consideration. First, in this paper, we only adopted lumped hydrological models to simulate catchment discharge without explicitly accounting for the spatial heterogeneity of precipitation inputs. In reality, different precipitation datasets not only differ in mean catchment rainfall but may also exhibit substantial differences in spatial variability, which may amplify discrepancies in simulated streamflow. Therefore, the contribution of precipitation input uncertainty to discharge simulations may be partly underestimated in this study. Second, besides precipitation, PET as another key model input also carries considerable uncertainties arising from both meteorological data and estimation methods [55,56]. Moreover, previous studies have indicated a compensatory relationship between precipitation and PET in streamflow simulations [19]. Future research could further incorporate PET uncertainty into the analysis to achieve a more accurate assessment of the sources of discharge simulation uncertainty.

## 6. Conclusions

Global precipitation datasets, such as satellite-based products and reanalysis data, are widely used in hydrological simulations. However, different datasets often exhibit substantial discrepancies, which in turn introduce large uncertainties into discharge simulations. In this study, three widely used satellite precipitation products (CHIRPS, PERSIANN, and IMERG) and one reanalysis dataset (ERA5) were employed to simulate streamflow in the Liujiang catchment. To account for structural uncertainty, three hydrological models with distinct internal structures and process representations were applied. Using the ANOVA method, we quantified the contributions of precipitation input uncertainty to streamflow simulations and projections, and compared them with uncertainties from model structure, model parameters, and climate projections. The results revealed significant differences among the four precipitation datasets and demonstrated that streamflow simulations were strongly influenced by both precipitation input and model selection. In some cases, hydrological models were able to partially compensate for precipitation biases, allowing datasets with large discrepancies to yield similar or even improved discharge simulations. During calibration, precipitation uncertainty contributed over 60% of total variance in rising limb periods, far exceeding model structure (<20%) and parameter (<5%) effects. Under future scenarios, while climate projections dominated (50–80%), precipitation datasets consistently accounted for 15–25% of uncertainty across flow conditions, significantly larger than structural uncertainties (<10%). Overall, this study highlights the critical and persistent role of global precipitation product uncertainty in hydrological simulations and climate change projections. The methodological framework developed here provides a robust approach for quantifying uncertainty contributions across different hydrological regimes and future scenarios. Our findings have direct implications for water resources management and climate adaptation planning. They emphasize the necessity of using ensemble approaches that incorporate multiple precipitation products in hydrological forecasting and impact studies, particularly in data-scarce regions reliant on global datasets. This is crucial for improving the reliability of flood risk assessments and water security projections under climate change. Future work should further integrate uncertainty from potential evapotranspiration (PET) estimates into the overall uncertainty budget. A comprehensive assessment that concurrently considers uncertainties from both precipitation and PET inputs is crucial for a more complete and realistic quantification of total uncertainty in hydrological projections.

**Author Contributions:** Y.C., Y.Q. and L.L. designed the experiments; Y.Q., Y.C. and N.M. performed the experiments, analyzed the data and wrote the original draft; all authors provided helpful suggestions revised the document and figures. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study has received funding from National Natural Science Foundation of China (No. 42572302).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Acknowledgments:** All precipitation datasets used in this study are publicly available from open-access websites. The observed discharge data for the Liujiang catchment can be provided upon reasonable request.



**Conflicts of Interest:** The authors declare no conflicts of interest. The HeNan Third Geology and Mineral Survey Institute Co., Ltd. had no role in the design of the study; in the collection, analyses, or interpretation of data; in writing of the manuscript, or in the decision to publish results.

## References

1. Khatakho, R.; Talchabhadel, R.; Thapa, B.R. Evaluation of Different Precipitation Inputs on Streamflow Simulation in Himalayan River Basin. *J. Hydrol.* **2021**, *599*, 126390. [CrossRef]
2. Strauch, M.; Bernhofer, C.; Koide, S.; Volk, M.; Lorz, C.; Makeschin, F. Using Precipitation Data Ensemble for Uncertainty Analysis in SWAT Streamflow Simulation. *J. Hydrol.* **2012**, *414–415*, 413–424. [CrossRef]
3. Tudaji, M.; Nan, Y.; Tian, F. Assessing the Value of High-Resolution Rainfall and Streamflow Data for Hydrological Modeling: An Analysis Based on 63 Catchments in Southeast China. *Hydrol. Earth Syst. Sci.* **2025**, *29*, 1919–1937. [CrossRef]
4. Kidd, C.; Huffman, G. Global Precipitation Measurement. *Meteorol. Appl.* **2011**, *18*, 334–353. [CrossRef]
5. Liu, S.; Wang, J.; Wang, H. Assessing 10 Satellite Precipitation Products in Capturing the July 2021 Extreme Heavy Rain in Henan, China. *J. Meteorol. Res.* **2022**, *36*, 798–808. [CrossRef]
6. Sun, Q.; Miao, C.; Duan, Q.; Ashouri, H.; Sorooshian, S.; Hsu, K.-L. A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Rev. Geophys.* **2018**, *56*, 79–107. [CrossRef]
7. Chang, Y.; Qi, Y.; Wang, Z. Comprehensive Evaluation of IMERG, ERA5-Land and Their Fusion Products in the Hydrological Simulation of Three Karst Catchments in Southwest China. *J. Hydrol. Reg. Stud.* **2024**, *52*, 101671. [CrossRef]
8. Shaowei, N.; Jie, W.; Juliang, J.; Xiaoyan, X.; Yuliang, Z.; Fan, S.; Linlin, Z. Comprehensive Evaluation of Satellite-Derived Precipitation Products Considering Spatial Distribution Difference of Daily Precipitation over Eastern China. *J. Hydrol. Reg. Stud.* **2022**, *44*, 101242. [CrossRef]
9. Taye, M.; Mengistu, D.; Sahlu, D. Performance Evaluation of Multiple Satellite Rainfall Data Sets in Central Highlands of Abbay Basin, Ethiopia. *Eur. J. Remote Sens.* **2023**, *56*, 2233686. [CrossRef]
10. Alazzy, A.A.; Lü, H.; Chen, R.; Ali, A.B.; Zhu, Y.; Su, J. Evaluation of Satellite Precipitation Products and Their Potential Influence on Hydrological Modeling over the Ganzi River Basin of the Tibetan Plateau. *Adv. Meteorol.* **2017**, *2017*, 3695285. [CrossRef]
11. Dembélé, M.; Schaeffli, B.; van de Giesen, N.; Mariéthoz, G. Suitability of 17 Gridded Rainfall and Temperature Datasets for Large-Scale Hydrological Modelling in West Africa. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 5379–5406. [CrossRef]
12. Laiti, L.; Mallucci, S.; Piccolroaz, S.; Bellin, A.; Zardi, D.; Fiori, A.; Nikulin, G.; Majone, B. Testing the Hydrological Coherence of High-Resolution Gridded Precipitation and Temperature Data Sets. *Water Resour. Res.* **2018**, *54*, 1999–2016. [CrossRef]
13. Jiang, Q.; Li, W.; Fan, Z.; He, X.; Sun, W.; Chen, S.; Wen, J.; Gao, J.; Wang, J. Evaluation of the ERA5 Reanalysis Precipitation Dataset over Chinese Mainland. *J. Hydrol.* **2021**, *595*, 125660. [CrossRef]
14. Zhang, Y.; Zheng, X.; Li, X.; Lyu, J.; Zhao, L. Evaluation of the GPM-IMERG V06 Final Run Products for Monthly / Annual Precipitation under the Complex Climatic and Topographic Conditions of China. *J. Appl. Meteorol. Climatol.* **2023**, *62*, 929–946. [CrossRef]
15. Gebrechorkos, S.H.; Leyland, J.; Dadson, S.J.; Cohen, S.; Slater, L.; Wortmann, M.; Ashworth, P.J.; Bennett, G.L.; Boothroyd, R.; Cloke, H.; et al. Global-Scale Evaluation of Precipitation Datasets for Hydrological Modelling. *Hydrol. Earth Syst. Sci.* **2024**, *28*, 3099–3118. [CrossRef]
16. Oudin, L.; Perrin, C.; Mathevet, T.; Andréassian, V.; Michel, C. Impact of Biased and Randomly Corrupted Inputs on the Efficiency and the Parameters of Watershed Models. *J. Hydrol.* **2006**, *320*, 62–83. [CrossRef]
17. Oudin, L.; Andréassian, V.; Perrin, C.; Anctil, F. Locating the Sources of Low-Pass Behavior within Rainfall-Runoff Models. *Water Resour. Res.* **2004**, *40*, 3291. [CrossRef]
18. Stephens, C.M.; Pham, H.T.; Marshall, L.A.; Johnson, F.M. Which Rainfall Errors Can Hydrologic Models Handle? Implications for Using Satellite-Derived Products in Sparsely Gauged Catchments. *Water Resour. Res.* **2022**, *58*, 29331. [CrossRef]
19. Wang, J.; Zhuo, L.; Rico-Ramirez, M.A.; Abdelhalim, A.; Han, D. Interacting Effects of Precipitation and Potential Evapotranspiration Biases on Hydrological Modeling. *Water Resour. Res.* **2023**, *59*, 33323. [CrossRef]
20. Emmanuel, I.; Payraastre, O.; Andrieu, H.; Zuber, F. A Method for Assessing the Influence of Rainfall Spatial Variability on Hydrograph Modeling. First Case Study in the Cevennes Region, Southern France. *J. Hydrol.* **2017**, *555*, 314–322. [CrossRef]
21. Velásquez, N.; Krajewski, W.F.; Seo, B.-C. Assessing the Impact of Radar-Rainfall Uncertainty on Streamflow Simulation. *J. Hydrometeorol.* **2025**, *26*, 169–184. [CrossRef]
22. Bárdossy, A.; Das, T. Influence of Rainfall Observation Network on Model Calibration and Application. *Hydrol. Earth Syst. Sci.* **2008**, *12*, 77–89. [CrossRef]
23. Xu, H.; Xu, C.-Y.; Chen, H.; Zhang, Z.; Li, L. Assessing the Influence of Rain Gauge Density and Distribution on Hydrological Model Performance in a Humid Region of China. *J. Hydrol.* **2013**, *505*, 1–12. [CrossRef]

24. Yin, Z.; Liao, W.; Lei, X.; Wang, H.; Wang, R. Comparing the Hydrological Responses of Conceptual and Process-Based Models with Varying Rain Gauge Density and Distribution. *Sustainability* **2018**, *10*, 3209. [CrossRef]
25. He, X.; Refsgaard, J.C.; Sonnenborg, T.O.; Vejen, F.; Jensen, K.H. Statistical Analysis of the Impact of Radar Rainfall Uncertainties on Water Resources Modeling. *Water Resour. Res.* **2011**, *47*, W09526. [CrossRef]
26. Vignal, B.; Galli, G.; Joss, J.; Germann, U. Three Methods to Determine Profiles of Reflectivity from Volumetric Radar Data to Correct Precipitation Estimates. *J. Appl. Meteorol.* **2000**, *39*, 1715–1726. [CrossRef]
27. HABIB, E.; ADUVALA, A.V.; MESELHE, E.A. Analysis of Radar-Rainfall Error Characteristics and Implications for Streamflow Simulation Uncertainty. *Hydrol. Sci. J.* **2008**, *53*, 568–587. [CrossRef]
28. Zhu, D.; Peng, D.Z.; Cluckie, I.D. Statistical Analysis of Error Propagation from Radar Rainfall to Hydrological Models. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 1445–1453. [CrossRef]
29. Fei, K.; Chen, M.; Zhou, Y.; Du, H.; Deng, S.; Gao, L. Streamflow and Surface Soil Moisture Simulation Capacity of High-Resolution Satellite-Derived Precipitation Estimate Datasets: A Case Study in Xijiang River Basin, China. *J. Hydrol. Reg. Stud.* **2022**, *42*, 101163. [CrossRef]
30. Ji, H.; Peng, D.; Gu, Y.; Liang, Y.; Luo, X. Evaluation of Multiple Satellite Precipitation Products and Their Potential Utilities in the Yarlung Zangbo River Basin. *Sci. Rep.* **2022**, *12*, 13334. [CrossRef]
31. Yan, Y.; Wang, G.; Nanding, N.; Chen, W. Hydrological Evaluation of Satellite-Based Precipitation Products in Hunan Province. *Remote Sens.* **2022**, *14*, 3127. [CrossRef]
32. Ma, Q.; Xiong, L.; Liu, D.; Xu, C.-Y.; Guo, S. Evaluating the Temporal Dynamics of Uncertainty Contribution from Satellite Precipitation Input in Rainfall-Runoff Modeling Using the Variance Decomposition Method. *Remote Sens.* **2018**, *10*, 1876. [CrossRef]
33. Qi, W.; Zhang, C.; Fu, G.; Sweetapple, C.; Zhou, H. Evaluation of Global Fine-Resolution Precipitation Products and Their Uncertainty Quantification in Ensemble Discharge Simulations. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 903–920. [CrossRef]
34. Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. The Climate Hazards Infrared Precipitation with Stations—A New Environmental Record for Monitoring Extremes. *Sci. Data* **2015**, *2*, 150066. [CrossRef] [PubMed]
35. Ashouri, H.; Hsu, K.-L.; Sorooshian, S.; Braithwaite, D.K.; Knapp, K.R.; Cecil, L.D.; Nelson, B.R.; Prat, O.P. PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. *Bull. Am. Meteorol. Soc.* **2015**, *96*, 69–83. [CrossRef]
36. Huffman, G.J.; Bolvin, D.T.; Braithwaite, D.; Hsu, K.-L.; Joyce, R.J.; Kidd, C.; Nelkin, E.J.; Sorooshian, S.; Stocker, E.F.; Tan, J. Integrated Multi-Satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG). In *Satellite Precipitation Measurement: Volume 1*; Springer: New York, NY, USA, 2020; pp. 343–353.
37. Soci, C.; Hersbach, H.; Simmons, A.; Poli, P.; Bell, B.; Berrisford, P.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Radu, R.; et al. The ERA5 Global Reanalysis from 1940 to 2022. *Q. J. R. Meteorol. Soc.* **2024**, *150*, 4014–4048. [CrossRef]
38. Hartmann, A.; Barberá, J.A.; Andreo, B. On the Value of Water Quality Data and Informative Flow States in Karst Modelling. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 5971–5985. [CrossRef]
39. Mazzilli, N.; Guinot, V.; Jourde, H.; Lecoq, N.; Labat, D.; Arfib, B.; Baudement, C.; Danquigny, C.; Dal Soglio, L.; Bertin, D. KarstMod: A Modelling Platform for Rainfall—Discharge Analysis and Modelling Dedicated to Karst Systems. *Environ. Model. Softw.* **2019**, *122*, 103927. [CrossRef]
40. Tritz, S.; Guinot, V.; Jourde, H. Modelling the Behaviour of a Karst System Catchment Using Non-Linear Hysteretic Conceptual Model. *J. Hydrol.* **2011**, *397*, 250–262. [CrossRef]
41. Huang, Q.; Chen, Z. Applicability of Evapotranspiration Equations for the Pearl River Basin. *Trop. Geogr.* **2014**, *34*, 737–745. (In Chinese)
42. Bosshard, T.; Carambia, M.; Goergen, K.; Kotlarski, S.; Krahe, P.; Zappa, M.; Schär, C. Quantifying Uncertainty Sources in an Ensemble of Hydrological Climate-Impact Projections. *Water Resour. Res.* **2013**, *49*, 1523–1536. [CrossRef]
43. Wang, J.; Zhuo, L.; Han, D.; Liu, Y.; Rico-Ramirez, M.A. Hydrological Model Adaptability to Rainfall Inputs of Varied Quality. *Water Resour. Res.* **2023**, *59*, 32484. [CrossRef]
44. Zeng, Q.; Chen, H.; Xu, C.-Y.; Jie, M.-X.; Chen, J.; Guo, S.-L.; Liu, J. The Effect of Rain Gauge Density and Distribution on Runoff Simulation Using a Lumped Hydrological Modelling Approach. *J. Hydrol.* **2018**, *563*, 106–122. [CrossRef]
45. Miao, C.; Gou, J.; Hu, J.; Duan, Q. Impacts of Different Satellite-Based Precipitation Signature Errors on Hydrological Modeling Performance Across China. *Earth's Future* **2024**, *12*, e2024EF004954. [CrossRef]
46. Bárdossy, A.; Anwar, F. Why Do Our Rainfall–Runoff Models Keep Underestimating the Peak Flows? *Hydrol. Earth Syst. Sci.* **2023**, *27*, 1987–2000. [CrossRef]

47. Dayal, D.; Pandey, A.; Gupta, P.K.; Kalura, P. Investigating the Utility of Satellite-Based Precipitation Products for Simulating Extreme Discharge Events: An Exhaustive Model-Driven Approach for a Tropical River Basin in India. *Environ. Monit. Assess.* **2024**, *196*, 608. [CrossRef] [PubMed]
48. van Kempen, G.; van der Wiel, K.; Melsen, L.A. The Impact of Hydrological Model Structure on the Simulation of Extreme Runoff Events. *Nat. Hazards Earth Syst. Sci.* **2021**, *21*, 961–976. [CrossRef]
49. Addor, N.; Rössler, O.; Köplin, N.; Huss, M.; Weingartner, R.; Seibert, J. Robust Changes and Sources of Uncertainty in the Projected Hydrological Regimes of Swiss Catchments. *Water Resour. Res.* **2014**, *50*, 7541–7562. [CrossRef]
50. Her, Y.; Yoo, S.-H.; Cho, J.; Hwang, S.; Jeong, J.; Seong, C. Uncertainty in Hydrological Analysis of Climate Change: Multi-Parameter vs. Multi-GCM Ensemble Predictions. *Sci. Rep.* **2019**, *9*, 4974. [CrossRef]
51. Kingston, D.G.; Thompson, J.R.; Kite, G. Uncertainty in Climate Change Projections of Discharge for the Mekong River Basin. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 1459–1471. [CrossRef]
52. Zhang, L.; Yuan, F.; Wang, B.; Ren, L.; Zhao, C.; Shi, J.; Liu, Y.; Jiang, S.; Yang, X.; Chen, T.; et al. Quantifying Uncertainty Sources in Extreme Flow Projections for Three Watersheds with Different Climate Features in China. *Atmospheric Res.* **2021**, *249*, 105331. [CrossRef]
53. Evin, G.; Hingray, B.; Thirel, G.; Ducharne, A.; Strohmenger, L.; Corre, L.; Trambay, Y.; Vidal, J.-P.; Bonneau, J.; Colleoni, F.; et al. Uncertainty Sources in a Large Ensemble of Hydrological Projections: Regional Climate Models and Internal Variability Matter. *EGUsphere* **2025**, *2025*, 1–43. [CrossRef]
54. Kabir, T.; Pokhrel, Y.; Felfelani, F. On the Precipitation-Induced Uncertainties in Process-Based Hydrological Modeling in the Mekong River Basin. *Water Resour. Res.* **2022**, *58*, e2021WR030828. [CrossRef]
55. Akhter, M.; Ahanger, M.A. Uncertainty in Evapotranspiration Inputs Impacts Hydrological Modeling. *Water Sci. Technol.* **2024**, *91*, 235–251. [CrossRef]
56. Jayatilake, D.I.; Smith, T. Assessing the Impact of PET Estimation Methods on Hydrologic Model Performance. *Hydrol. Res.* **2020**, *52*, 373–388. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## Article

# Assessing the Impact of Assimilated Remote Sensing Retrievals of Precipitation on Nowcasting a Rainfall Event in Attica, Greece

Aikaterini Pappa <sup>1</sup>, John Kalogiros <sup>2</sup>, Maria Tombrou <sup>3</sup>, Christos Spyrou <sup>4</sup>, Marios N. Anagnostou <sup>5,\*</sup>, George Varlas <sup>6</sup>, Christine Kalogeri <sup>1</sup> and Petros Katsafados <sup>1</sup>

<sup>1</sup> Department of Geography, Harokopio University of Athens, El. Venizelou Str. 70, 17671 Athens, Greece; apappa@hua.gr (A.P.); ckalogeri@hua.gr (C.K.); pkatsaf@hua.gr (P.K.)

<sup>2</sup> Institute for Environmental Research and Sustainable Development, National Observatory of Athens, 15236 Palaia Penteli, Greece; jkalog@noa.gr

<sup>3</sup> Department of Physics, National and Kapodistrian University of Athens, 15771 Athens, Greece; mtombrou@phys.uoa.gr

<sup>4</sup> Research Centre for Atmospheric Physics and Climatology, Academy of Athens, 10679 Athens, Greece; cspyrou@academyofathens.gr

<sup>5</sup> Department of Informatics, Ionian University, 49100 Corfu, Greece

<sup>6</sup> Institute of Marine Biological Resources and Inland Waters, Hellenic Centre for Marine Research, 46.7 km Athens-Sounio, 19013 Anavissos, Greece; gvarlas@hcmr.gr

\* Correspondence: managnost@ionio.gr

**Abstract:** Accurate short-term rainfall forecasting, an essential component of the broader framework of nowcasting, is crucial for managing extreme weather events. Traditional forecasting approaches, whether radar-based or satellite-based, often struggle with limited spatial coverage or temporal accuracy, reducing their effectiveness. This study tackles these challenges by implementing the Local Analysis and Prediction System (LAPS) enhanced with a forward advection nowcasting module, integrating multiple remote sensing rainfall datasets. Specifically, we combine weather radar data with three different satellite-derived rainfall products (H-SAF, GPM, and TRMM) to assess their impact on nowcasting performance for a rainfall event in Attica, Greece (29–30 September 2018). The results demonstrate that combining high-resolution radar data with the broader coverage and high temporal frequency of satellite retrievals, particularly H-SAF, leads to more accurate predictions with lower uncertainty. The assimilation of H-SAF with radar rainfall retrievals (HX experiment) substantially improved forecast skill, reducing the unbiased Root Mean Square Error by almost 60% compared to the control experiment for the 60 min rainfall nowcast and 55% for the 90 min rainfall nowcast. This work validates the effectiveness of the specific LAPS/advection configuration and underscores the importance of multi-source data assimilation for weather prediction.

**Keywords:** LAPS; seamless prediction; nowcasting; data assimilation; weather radar; remote sensing data; satellite retrievals; forward advection

## 1. Introduction

Short-term rainfall forecasts have been challenging over the last few decades. Weather forecasts that focus on a short time period of the next few minutes up to six hours ahead and have high temporal and spatial resolution are referred to using the term nowcasting [1–4]. Nowcasting refers to a detailed analysis and description of current weather conditions and their evolution over very short periods, typically up to six hours ahead, as defined

by the World Meteorological Organization [5]. Accurate short-term rainfall forecasting, which is a component of nowcasting, is essential for managing high-impact weather events. Nowcasting is connected with data assimilation in meteorology to improve forecasts for short lead times. The data assimilation cycle entails a number of processes that are repeated, including the collection of observations, the short-range forecast, the comparison of the forecast and observations, and the development of an analysis that combines the forecast and observations. An essential setting that corresponds to the dynamical scales addressed by the observation system and the numerical model is the cycle length in sequential data-assimilating forecasting systems [6]. The cycle's duration influences how frequently new observations are included in the model. Modern nowcasting often relies on radar and satellite observations, which have complementary strengths: radar offers high resolution, making it possible to monitor rainfall dynamics in detail [7,8], while satellites provide broader coverage [9]. However, both sources have limitations: radar may suffer limited range, and satellite retrievals may have coarser resolution or longer latency. Moreover, numerical models exhibit a "spin-up" problem at small scales, making pure model forecasts unreliable at less than 3 h lead times. To overcome these challenges, blending multiple data sources via data assimilation has become a promising approach.

Previous studies have demonstrated that assimilating radar or satellite data can improve short-term rainfall forecasts [10]. For example, the assimilation of radar reflectivity into mesoscale models can improve heavy rainfall prediction [11,12], and radar-derived water vapor retrievals significantly improve rainfall predictions [13–15]. Similarly, incorporating satellite-derived rainfall fields into models has shown skill improvements [16–18]. Several studies have investigated the assimilation of both radar and satellite rainfall data to improve quantitative rainfall forecasting, demonstrating that combining these observational systems can significantly enhance forecast performance [10,19–22].

This study aims to build on these findings by investigating the sensitivity of the rainfall nowcasts regarding the assimilation of weather radar data and satellite retrievals. The primary goal is to evaluate how the characteristics of satellite-derived rainfall estimates, such as spatial and temporal resolution, impact the quality of short-term forecasts. The approach involves implementing the Local Analysis and Prediction System (LAPS), configured with a recently developed nowcasting scheme. It is a forward advection scheme which relies on assimilating radar and satellite-derived rainfall retrievals to provide a more complete description of the atmospheric state at a given time. For the purpose of this study, we used the rainfall event that affected the Athens metropolitan area in Greece on 29 and 30 September 2018. Specifically, how the assimilation of satellite data combined with radar data impacts the accuracy of rainfall nowcasts has been assessed. The choice of Attica is motivated by its complex topography and coastal influences, which can lead to localized and rapidly evolving rainfall events, making accurate nowcasting particularly challenging and valuable for the region. The availability of high-quality, high-resolution XPOL radar data in this region provides a unique opportunity to assess the impact of integrating such data with different satellite products. Although this study focuses on the Athens metropolitan area as a case study, the methodology and findings are not limited to this location and can be applied to other regions.

While high-resolution radar data alone could support nowcasting through extrapolation techniques such as pySTEPS [23], these approaches are limited in coverage and cannot represent regions outside the radar range. The integration of satellite data ensures broader spatial coverage and enables the dynamically consistent initialization of the atmospheric state. By assimilating both data sources within the LAPS, our approach aims to improve rainfall nowcasting.



The main motivation of this study is to overcome the challenges and limitations of current rainfall nowcasting systems, particularly in achieving accurate and reliable short-term forecasts. Traditional numerical weather prediction (NWP) models suffer from the “spin-up” problem, which occurs during the first few hours of simulations. Additionally, conventional systems often rely on meteorological data from surface weather stations with coarse spatial coverage. These limitations reduce their ability to effectively capture localized and rapidly evolving rainfall events. Thus, the specific contributions of this study are as follows:

- Implementation of a forward advection nowcasting module within the LAPS.
- Evaluation of three assimilation configurations (CTRL, GTX, HX) to assess the impact of radar and satellite retrievals on nowcasting.
- Significant error reduction by the implementation of X-band weather radar estimates blended with H-SAF satellite retrievals.

The method and the description of the data utilized for this study are provided in Section 2. The study area and the synoptic analysis of the event are described in Section 3. Section 4 provides the results and a thorough discussion, and Section 5 summarizes the entire outcomes.

## 2. Methods and Data

Short-term weather forecasting remains a challenge due to the rapid evolution of rainfall events and the limitations of conventional prediction models. This study explores the following question: how does integrating high-resolution radar and satellite rainfall retrievals improve nowcasting accuracy compared to conventional methods? To address this, we designed three assimilation experiments:

- CTRL (Baseline): Relies only on Global Forecast System (GFS) background forecasts combined with METAR and SYNOP in situ observations.
- GTX Experiment: Incorporates XPOL radar data alongside satellite retrievals from GPM IMERG and TRMM and METAR and SYNOP in situ observations.
- HX Experiment: Combines XPOL radar data and METAR and SYNOP in situ observations with H-SAF rainfall retrievals for improved temporal resolution.

Unlike traditional methods that depend solely on radar or satellite data, our approach dynamically integrates multiple sources within the LAPS, allowing for real-time adjustments in rainfall estimation. The novelty of this approach lies not only in its ability to balance fine-detail radar observations with broad satellite coverage but also in demonstrating that implementing a forward advection scheme improves forecast accuracy compared to the control experiment, highlighting the benefits of combining satellite and radar data. Although a direct comparison with external nowcasting systems was not conducted, the performance observed in this study is consistent trends reported in other nowcasting studies [24].

The nowcasting method relies on the deployment of a mesoscale 3D-Var data assimilation system, the Local Analysis and Prediction System (LAPS) developed by NOAA. The impact of various remote sensing retrievals on the nowcasting skill of a rainfall event that affected the Athens metropolitan area in Greece on 29 and 30 September 2018 is analyzed. The LAPS is a tool used for the assimilation of mesoscale meteorological data, developed to assimilate data from all available sources, including radar, soundings, satellite, aircraft, and meteorological networks, and produce a three-dimensional depiction of atmospheric processes that evolve over time and are spatially distributed [25,26]. The Local Analysis and Prediction System (LAPS) was selected for this study due to its ability to provide a

simple and effective framework for assimilating a variety of observational data sources, including radar and satellite retrievals, in real-time nowcasting applications. The LAPS offers a computationally efficient approach that is particularly suitable for short-term rainfall forecasting, where rapid updates and high temporal resolution are critical [27,28]. While more complex data assimilation systems exist, the results of this study demonstrate that even a simplified system can substantially improve short-term rainfall forecasts.

Before being assimilated into the LAPS, the remote sensing data undergo preprocessing to ensure compatibility. The radar (XPOL) data are processed using the Self-Consistent Optimal Parameterization–Microphysics Estimation (SCOPE-ME) algorithm [29] and are regridded to the LAPS spatial domain, covering  $110 \times 110$  grid points over Attica, with a spatial resolution of 1 km. Satellite rainfall products from Global Precipitation Measurement (GPM), Tropical Rainfall Measuring Mission (TRMM), and EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF) are processed using Integrated Multi-satellite Retrievals for GPM (IMERG) [30], TRMM 3B42 [31], and blending MW/IR [32] algorithms, respectively. These data are reformatted to LAPS-compatible formats and used as surface rainfall observations (gauge observations).

The LAPS employs a three-dimensional variational (3DVAR) assimilation minimizing the difference between observed and model-predicted fields [33], as described by the cost function in Equation (1).

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2}(H(x) - y)^T O^{-1} (H(x) - y) \quad (1)$$

where  $B$  refers to the model error covariance matrix,  $O$  denotes the observational error matrix,  $x$  corresponds to the control variable,  $x_b$  represents the background field,  $y$  represents the observation values, and  $H$  denotes the observation operator. The observation operator  $H$  is implemented as a bilinear interpolation operator, which is the link between the observation and the model, the background error covariance is derived from multiple iterations of a recursive filter [34], and the observational error is a diagonal matrix.

The assimilation process begins with background first guess fields (forecasts near-to-analysis) derived from the Global Forecast System (GFS) which has a horizontal resolution of  $0.25^\circ \times 0.25^\circ$  and is interpolated to the LAPS grid as initial first guess fields. The Global Forecast System (GFS), a widely used numerical weather prediction model, serves as the primary source of background fields in our configuration. Then, multiple data sources, including radar, satellite, and in situ observations, are integrated into the LAPS to produce high-resolution analysis fields.

The accumulation analysis for rainfall is based on the Barnes objective analysis, which is designed to blend different data sources effectively. In this study, satellite precipitation products are reformatted and ingested into the LAPS as surface observations. Then, these satellite-derived observations are blended with the gridded XPOL radar data using the successive correction method of the Barnes multi-pass iteration to perform accumulation analysis for rainfall [27,28]. The process unfolds over two main steps. First, linear regression is conducted on valid radar–gauge observation pairs to correct biases in the radar estimates across all grid points. Following the bias correction, the successive correction method of the Barnes multi-pass iteration is used to blend and weight the observations with the background fields. This process continues until the consecutive corrections become insignificant. During each iteration, the differences between the observations and the background, along with their corresponding Root Mean Square (RMS) errors, are computed. The weights that are assigned to observations are based on representativeness errors and

errors of the instruments, while the weight of background points is chosen to accurately reflect the background error [28].

To predict the movement of rainfall, an advection scheme was integrated into the LAPS for a 1–3 hours' time horizon. This scheme models the spatial evolution of rainfall by solving the two-dimensional advection equation using wind fields derived from LAPS-assimilated data. The process begins with the rainfall fields produced by the LAPS, which is then advected in time using the horizontal wind components ( $u$  and  $v$ ) at each grid point. This scheme relies on the advection equation in the  $x$  and  $y$  directions, which is a first-order equation:

$$\frac{\partial R}{\partial t} + u \frac{\partial R}{\partial x} + v \frac{\partial R}{\partial y} = 0 \quad (2)$$

where  $R$  represents the rainfall rate,  $t$  is time,  $u$  and  $v$  are in the  $x$  and  $y$  directions, respectively,  $x$  and  $y$  are the spatial coordinates.

To implement this numerically, Equation (2) is discretized as follows:

$$\frac{1}{\Delta t} (R_{i,j}^{t+1} - R_{i,j}^t) + \frac{u}{\Delta x} (R_{i,j}^t - R_{i-1,j}^t) + \frac{v}{\Delta y} (R_{i,j}^t - R_{i,j-1}^t) = 0 \quad (3)$$

Here,  $\Delta t$  represents the time step,  $R$  denotes the rainfall rate,  $(i, j)$  refers to the LAPS domain's grid point,  $u$  corresponds to the  $x$  direction wind component,  $\Delta x$  refers to the  $x$  direction grid increment,  $v$  represents the  $y$  direction wind component, and  $\Delta y$  denotes the  $y$  direction grid increment. The rainfall rate at each grid point is updated using wind components at each time step.

The Courant numbers are calculated in order to see if the time step that is selected assures the stability of Equation (3):

$$CU = u \frac{\Delta t}{\Delta x} \geq 0 \quad (4)$$

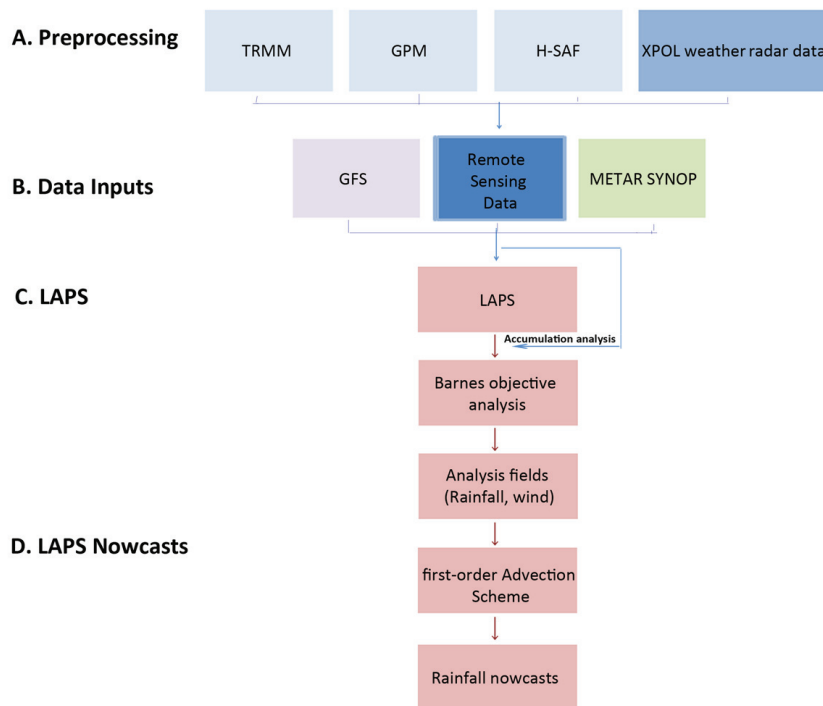
$$CV = v \frac{\Delta t}{\Delta y} \geq 0 \quad (5)$$

where  $CU$  represents the  $x$  direction Courant number and  $CV$  represents the  $y$  direction Courant number.

As a stability criterion, the restriction  $CU + CV \leq 1$  is used, which is known as the Courant–Friedrichs–Lewy (CFL) condition. With this criterion, it is ensured that the time step used in the simulation is small enough to accurately capture changes in the system and prevent numerical instabilities. The Courant number is calculated at each time step to determine whether the time step used in the simulation ensures the stability of the formulation. Additionally, Equations (3) through (5) were solved in order to estimate the  $R$  variable, which represents the rainfall, at time  $t + 1$  [35]. The advection scheme is developed to transfer rainfall over time, solving the transport equation in the assimilation system's structure (Figure 1).

Although rainfall is the result of various physical processes, the advection equation has been successfully applied to rainfall nowcasting for short lead times up to 90 min [36–39]. Nowcasting using the advection–diffusion equation is made possible by studies like Ha and Park [36], which assumed that fluid dynamics govern the motion field of rainfall. Furthermore, the approach avoids numerical instability issues (spin-up effects) common in mesoscale models, making it advantageous for nowcasting. By using a forward advection scheme in a 2D spatial domain with a short lead time (up to 3 h), this model estimates the approximate movement of rainfall fields without incorporating complex physical

processes [38]. This approach demonstrates the effectiveness of the advection equation for nowcasting applications [36–39].



**Figure 1.** Methodology system diagram for nowcasting.

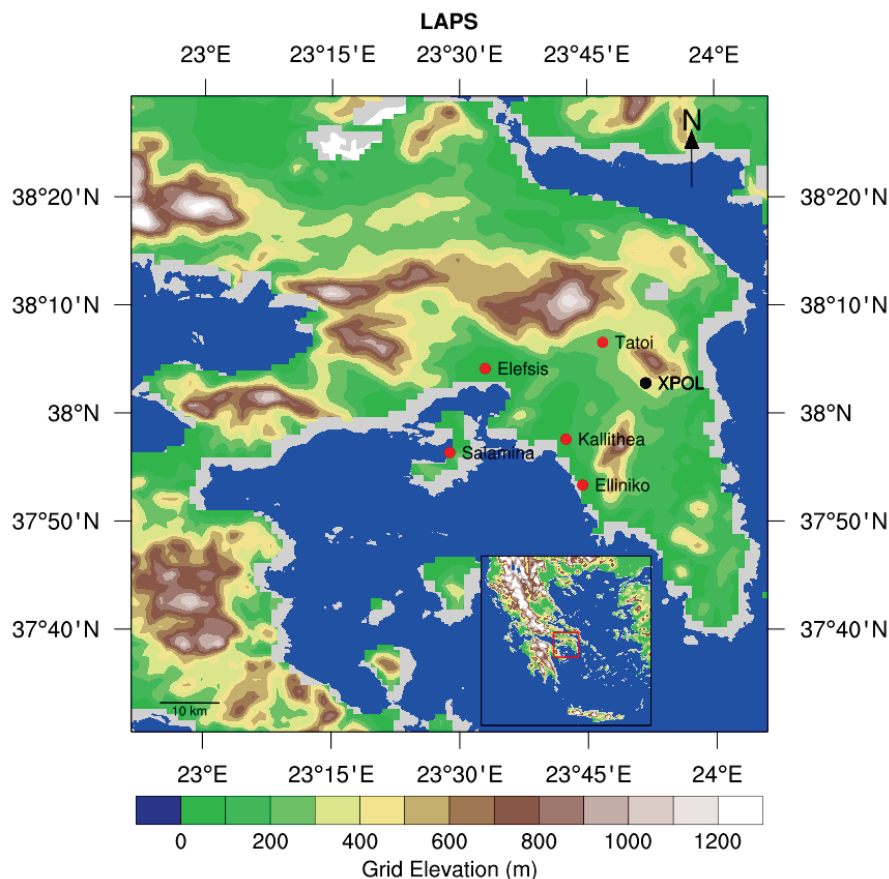
The advection scheme that is embedded in the LAPS avoids the “spin-up” period, which is a critical issue in the numerical weather prediction mesoscale models, and it is necessary to ensure that the model physics and dynamics become more consistent and reasonable. The scheme’s disadvantage is that it could not be used for forecasting beyond a 3 h time frame because no further physical parameterizations were added. The dynamical processes can no longer be disregarded after the forecast horizon.

Table 1 compares the proposed LAPS-based assimilation approach relative to other general nowcasting methodologies. These represent internal variations in the proposed assimilation scheme rather than comparisons with external nowcasting systems. This comparison highlights the primary advantage of the proposed approach: improved short-term predictability by dynamically fusing radar and satellite data, balancing precision with coverage.

**Table 1.** Characteristics of the proposed LAPS-based approach in the context of other nowcasting methodologies.

| Methodology                      | Data Sources                    | Forecast Lead Time | Spatial Resolution | Limitations                                                              |
|----------------------------------|---------------------------------|--------------------|--------------------|--------------------------------------------------------------------------|
| Traditional NWP                  | Numerical models                | 3–6 h              | 10–50 km           | Computationally demanding, limited short-term accuracy                   |
| Radar-Based Nowcasting           | Weather radar only              | Up to 3 h          | 1–5 km             | High accuracy but limited to radar coverage areas                        |
| Satellite-Based Nowcasting       | Satellite retrievals            | 1–6 h              | 10–25 km           | Broad coverage but lower resolution                                      |
| Proposed LAPS-Based Assimilation | Radar with Satellite retrievals | Up to 3 h          | 1 km               | Optimized integration of high-resolution radar with broad satellite data |

For the needs of this study, the LAPS was configured over a domain that covered Attica (Figure 2), and its details are presented in Table 2. Global Forecast System (GFS) 0.25-degree resolution data from the 29 and 30 September 2018 at the 0000 UTC cycle were used to generate 3D background fields.



**Figure 2.** Covering Attica, the LAPS model domain, and its topography. The location of the XPOL radar is indicated by the black dot, and the independent surface meteorological stations Salamina, Kallithea, Elliniko, Tatoi, and Elefsis are indicated by the red dots. The inset map provides a regional locator for the study domain. A north arrow and a scale bar are included for geographic reference.

**Table 2.** LAPS configuration.

| Horizontal Grid Increment | 1 km                           |
|---------------------------|--------------------------------|
| Grid points               | 110 × 110                      |
| Vertical levels           | 41                             |
| Background fields         | GFS forecasts on 0.25° × 0.25° |
| Assimilated observations  | METAR, SYNOP                   |

Three assimilation experiments have been performed, each using the GFS near to analysis forecasts as background fields.

The experiments are as follows:

- The control experiment—CTRL, for which only the GFS forecasts and METAR and SYNOP in situ observations (GFS + METAR + SYNOP) are used.
- The GTX experiment, where the GFS forecasts and METAR and SYNOP in situ observations are combined with XPOL radar data and Global Precipitation Measurement (GPM) and Tropical Rainfall Measuring Mission (TRMM) retrievals (GFS + METAR + SYNOP + GPM + TRMM + XPOL).



- (c) The HX experiment, where the GFS forecasts and METAR and SYNOP in situ observations are combined with XPOL radar data and EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF) retrievals (GFS + METAR + SYNOP + H-SAF + XPOL).

The three assimilation experiments were designed to evaluate the impact of different data sources on rainfall nowcasting. The CTRL experiment serves as the baseline, using only Global Forecast System (GFS) forecasts combined with METAR and SYNOP in situ observations. This experiment provides a reference for assessing the improvements brought about by the assimilation of additional remote sensing data. The GTX experiment incorporates GFS forecasts, XPOL radar data, and satellite retrievals from the Global Precipitation Measurement (GPM) and Tropical Rainfall Measuring Mission (TRMM). This experiment was designed to evaluate the impact of combining radar data with satellite-derived rainfall estimates from GPM and TRMM, which have different spatial and temporal resolutions (30 min and 3 h, respectively). The HX experiment combines GFS forecasts, XPOL radar data, and satellite retrievals from the EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF). This experiment was included to assess the added value of H-SAF products, which offer higher temporal resolution (15 min). The division of satellite data into two experiments (GTX and HX) allows for a direct comparison of the performance of different satellite products, highlighting the different contributions of each dataset to the accuracy of the nowcasting.

The assimilated fields and the experiments are listed in Table 3. The three assimilation experiments have been performed twice. The first set of experiments was conducted on 29 September 2018. The LAPS was run at 15:00 UTC, producing analysis fields and advected rainfall fields every 10 min for a total of 90 min. The second set of experiments was conducted on 30 September 2018. The LAPS ran at 03:00 UTC, producing analysis fields and advected rainfall fields similarly every 10 min for a total of 90 min.

**Table 3.** LAPS experiments.

| Experiment | Assimilated Fields                      |
|------------|-----------------------------------------|
| CTRL       | GFS + METAR + SYNOP                     |
| GTX        | GFS + METAR + SYNOP + GPM + TRMM + XPOL |
| HX         | GFS + METAR + SYNOP + H-SAF + XPOL      |

To evaluate nowcasting accuracy, two statistical metrics are analyzed: the BIAS (6) and the unbiased Root Mean Square Error (ubRMSE) (8). These metrics serve distinct purposes: the BIAS reveals consistent systematic over- or underestimation across nowcasts, while ubRMSE measures the magnitude of forecast errors by excluding the mean bias component, thus quantifying errors independent of systematic bias. Both metrics are calculated by comparing the predicted rainfall values ( $F_i$ ) with the observed rainfall values ( $O_i$ ) at each grid point over the LAPS domain for each forecast lead time. Here,  $N$  represents the total number of data points,  $F_i$  is the predicted value at the  $i$ th data point, and  $O_i$  is the true value at the  $i$ th data point. For BIAS and ubRMSE evaluation, the empirical 95% confidence intervals based on the bootstrap technique [40] are computed. Specifically, for each lead time, the forecast–observation pairs are resampled with replacement 1000 times to obtain the 95% confidence level. By computing the 97.5th and 2.5th percentiles in the bootstrap distribution, the 95% confidence level is obtained.

$$BIAS = \frac{1}{N} \sum_{i=1}^N (F_i - O_i) \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2} \quad (7)$$

$$ubRMSE = \sqrt{RMSE^2 - BIAS^2} \quad (8)$$

Another statistical evaluation is performed using contingency Table 4, which is appropriate for discrete variables, with emphasis on identifying false detections and the accuracy of the experimental outcomes. Two detection metrics (POFD and accuracy) are estimated for four rainfall thresholds (0.2, 0.5, 1, and 2 mm) using rainfall derived from LAPS nowcasts, validated through XPOL radar data. The evaluation sample size, representing the matched LAPS–XPOL pairs across all thresholds, ranges from 9245 to 9259.

$$POFD = \frac{False\ Alarms}{Correct\ Negatives + False\ Alarms} \quad (9)$$

$$ACCURACY = \frac{Hits + Correct\ Negatives}{Total} \quad (10)$$

**Table 4.** Contingency table.

| Model Forecasts | Observations |                   |
|-----------------|--------------|-------------------|
|                 | YES          | NO                |
| YES             | Hits         | False Alarms      |
| NO              | Misses       | Correct Negatives |

### 2.1. X-Band Dual-Polarization (XPOL) Weather Radar Data

The X-band dual-polarization (XPOL) Doppler weather radar collects data on both the horizontal and vertical components of the electromagnetic waves that are reflected back to the radar from rainfall particles in the atmosphere, and its data allow for more detailed insights into the orientation, size, and composition of rainfall particles. The National Observatory of Athens operates the XPOL ground-based radar that is located on Penteli mountain (latitude 38.0491°, longitude 23.8647°) at a height of 500 m above sea level, at the northeast part of the Athens Basin, and it operates the XPOL ground-based radar. Plan position indicator (PPI) mode is used by the radar, capturing observations in a 180° wide sector scan (south–west–north) at elevation sweeps of 1°, 2°, and 3° with a 150 m range resolution (maximum range of 120 km) during rainfall events. The antenna rotation rate was 6 degs/s, and less than 2 min was the duration of a full-volume scan. To retrieve rainfall values, the Self-Consistent Optimal Parameterization–Microphysics Estimation (SCOPE-ME) X-band dual-polarization algorithm was used [29,41].

### 2.2. TRMM

The objective of the Tropical Rainfall Measuring Mission (TRMM) was to study the variability and spatial distribution of rainfall [42]. It was a collaborative mission conducted by the Japan Aerospace Exploration Agency (JAXA) and the National Aeronautics and Space Administration (NASA). Five instruments have been installed on TRMM, and four of them are used for rainfall: a nine-channel passive microwave imager (TMI), a Precipitation Radar (PR), a five-channel Visible and Infrared Scanner (VIRS), and a Lightning Imaging Sensor (LSI) [43–45]. The purpose of the TMI is to quantify rainfall, water vapor, and sea surface temperature. Its swath width is 760 km. The PR provides radiance measurements, and its swath width is 215 km [43]. The VIRS provides measurements of rainfall intensity, distribution,

and type, and it has a 0.6 to 12  $\mu\text{m}$  wavelength and a 2 km horizontal resolution and a swath width of 720 km. The LSI provides a link between lighting occurrence and rainfall events [46].

Among the TRMM rainfall products, 3B42 and 3B43 are highly popular. These products have high temporal and spatial resolutions and are widely used in the monitoring and forecasting of extreme events. For the needs of this study, TRMM 3B42 products are used that produce high spatial (25 km) and temporal resolution (3 h) rainfall data. The rainfall data have a 0.25 degrees spatial resolution in terms of both latitude and longitude, covering the area between 50 S and 50 N. TRMM was in operation from 1997 to 2015. TRMM Multi-satellite Precipitation Analysis (TMPA) products continued to be produced into the GPM until the end of 2019.

### 2.3. GPM

The Global Precipitation Measurement (GPM) program succeeded the TRMM [47,48]. The goals of this program are to enhance hydrological modeling, climate modeling, and weather forecasts and to improve the knowledge of freshwater availability and water cycle variability [48,49]. Comparing TRMM and GPM, GPM has better resolution than TRMM as the GPM has 10 km spatial and 30 min temporal resolution [50].

The GPM mission comprised a core observatory satellite that carries a Dual-frequency Precipitation Radar (DPR) to provide three-dimensional rainfall data and a passive GPM microwave imager (GMI) to estimate rain rate and nine constellation satellites. The DPR includes two radars, a KaPR and a KuPR, with 35.5 GHz and 13.6 GHz frequencies, respectively [30,49,51]. A Ku-band radar had already been used by the TRMM (Tropical Rainfall Measuring Mission) that measures moderate to heavy rain in tropical and mid-latitude regions, while a Ka-band measures weak rain at higher latitudes as well [49]. The passive GPM microwave imager (GMI) is a conical-scanning, high-resolution, multi-channel microwave radiometer. The GMI provides a reference standard to rainfall estimates and has thirteen microwave channels with a frequency ranging from 10.65 to 183.31 GHz [49,51].

To estimate rainfall retrievals, the algorithm used is the Integrated Multi-satellitE Retrievals for GPM (IMERG) [52]. Moreover, IMERG Early, IMERG Late, and IMERG Final are the three types of IMERG products [53]. The difference among these types is in terms of data availability as IMERG Early has a 4 h delay (is well-suited for real time applications, such as predicting flash flood), IMERG Late has a 12 h to 14 h delay (is well-suited for water resource management), and IMERG Final has a 2.5–3.5 months delay (is well-suited for research). The spatial resolution of the IMERG rainfall data is  $0.1^\circ \times 0.1^\circ$ , with a temporal resolution reaching up to 30 min [52,54]. GPM IMERG EARLY half hourly products are chosen for this study.

### 2.4. H-SAF

The EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF) is a program that focuses on parameters related to soil moisture, rainfall, and snow. The H-SAF products are derived from measurements collected by microwave and infrared satellite sensors located on geostationary and polar-orbiting satellites [55,56].

The H03B is the H-SAF Operational Products that are used in this study. The spatial resolution of these products ranges from 3 km, near the sub-satellite, to 8 km over Europe, with a high temporal resolution of 15 min [56,57]. The algorithm for product H03B is based on the technique blended satellite “Rapid Update” that was initially created at the Naval Research Laboratory (NRL) [57–60]. This product measures the ground rainfall rate blending operational geostationary IR imagers and LEO MW sensors [57]. This technique

combines the brightness temperatures from the IR geostationary sensors with the rainfall rates from passive microwave (PMW) satellite sensors [61].

### 3. Synoptic Analysis

Due to its position in the eastern Mediterranean Sea, Greece is vulnerable to medicanes (Mediterranean hurricanes), which refer to a rare and powerful type of cyclone that forms over the Mediterranean Sea [62,63]. Medicanes are usually associated with strong winds, heavy rain, and storms, and they can display characteristics that are similar to both hurricanes and mid-latitude cyclones. On 27 September 2018, a notable extreme event was formed in Libya around 12:00 UTC [64]. The system developed into three phases [65]. The first phase took place between 27/00 UTC and 28/00 UTC with the formation of an extratropical cyclone [65]. Between 28/06 UTC and 28/12 UTC, the second phase took place when an extratropical cyclone transitioned into medicane Zorbas. Between 28/18 UTC and 29/18 UTC, the third phase took place with the transformation of medicane to a strong mid-latitude cyclone. Four days after its formation, Zorbas decayed in the Aegean Sea. The medicane significantly damaged Southern Greece, including Crete, Peloponnese, Evia, and Athens, with severe winds, torrential rainfall, and flooding.

The development of medicane Zorbas began on 27 September 2018, with the emergence of a low-pressure system at 1007 hPa (Figure 3a) in the central Mediterranean Sea. Concurrently, a warm front and a trough began to take shape over the Mediterranean Sea between Sicily, South Greece, and Libya, which indicates atmospheric instability that there was a chance for showers. Meanwhile, a weather system associated with a cold anticyclone in Central Europe gradually moved eastward towards Greece. This favored the increase in baroclinicity in the lower layers on the coastal areas of eastern Libya. The system was deepening further due to the cold anomaly (Figure 3b) in the Mediterranean Sea. On 29 September 00UTC, the system approached mainland Greece with a low pressure at 997 hPa (Figure 3c) accompanied by a front over the Ionian Sea and north Greece. Furthermore, the detached low pressure at the 500 hPa level had moved further northeast, affecting the western parts of Greece with the geopotential height reaching to 576 gpdam (Figure 3d). On 30 September 2018 at 00UTC, the medicane continued to affect Greece and Turkey. The center of the low pressure directed to the northeast had a rise in pressure at 1004 hPa (Figure 3e) and was accompanied by an occluded front. Eventually, the medicane moved to the north–north east direction and reached Turkey and gradually dispersed.

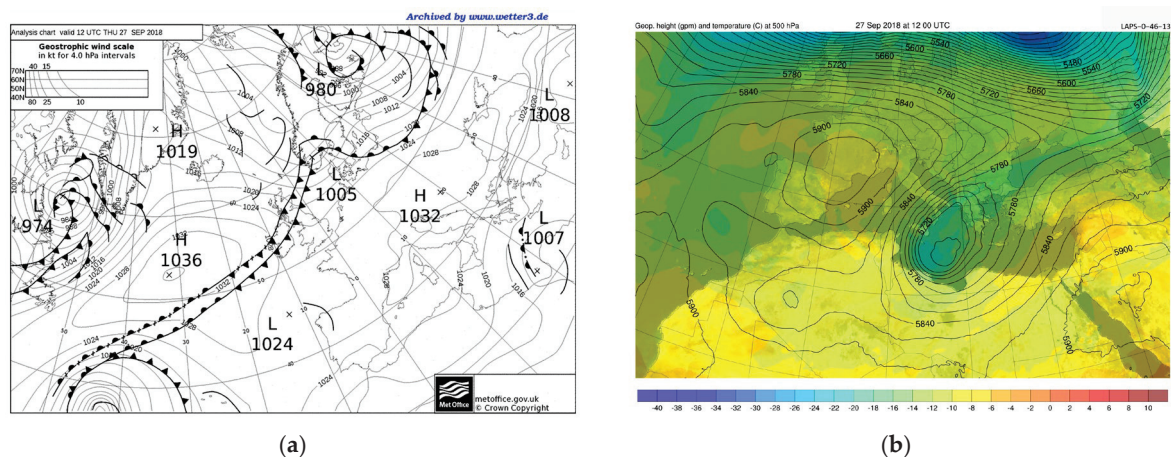
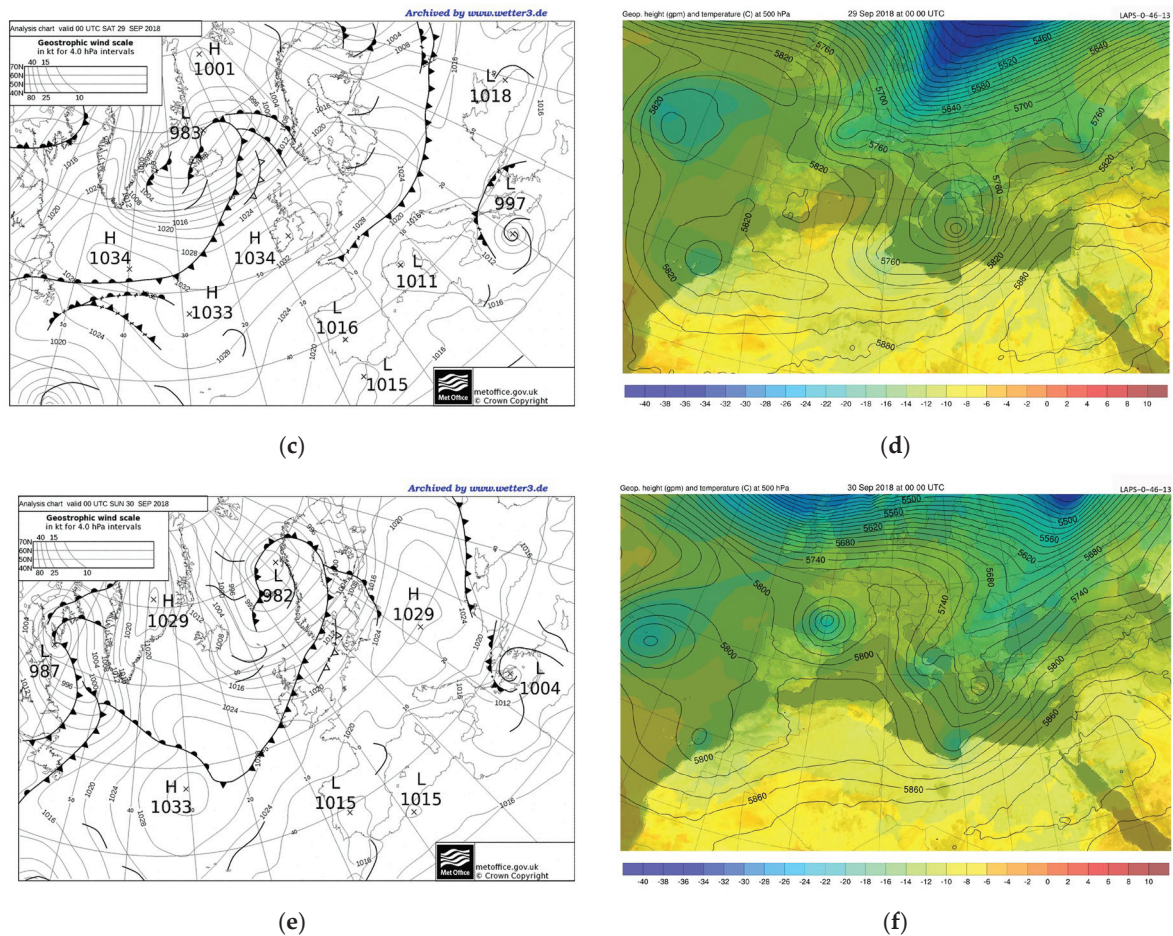


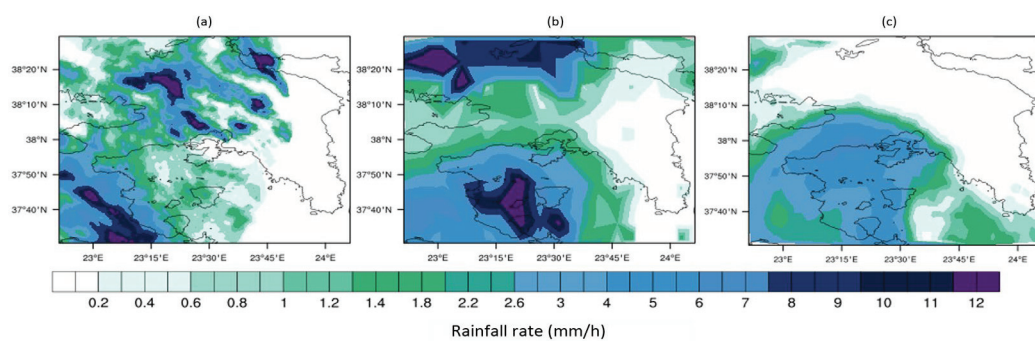
Figure 3. Cont.





**Figure 3.** Left column: Analysis map of mean sea level pressure (hPa) for (a) 27 September 2018 at 12:00 UTC; (c) 29 September 2018 at 00:00 UTC; (e) 30 September 2018 at 00:00 UTC, provided by the UK Met Office. The map is based on the wetter3.de archive ([https://www.wetter3.de/archiv\\_ukmet\\_en.html](https://www.wetter3.de/archiv_ukmet_en.html)) (retrieved on 4 March 2024)); right column: temperature is represented by color-shaded areas in °C and geopotential height is represented by black contours in gpm at 500 hPa for (b) 27 September 2018 at 12:00 UTC; (d) 29 September 2018 at 00:00 UTC; (f) 30 September 2018 at 00:00 UTC. The map is derived from analysis data provided by the GFS (Global Forecast System).

In addition to the analysis maps, Figure 4 provides the rainfall characteristics at the analysis time that initiated the first experiment, offering insights into the spatial distribution and intensity of rainfall for 29 September 2018 at 15:00 UTC. Specifically, Figure 4 illustrates (a) radar-derived rainfall rates, (b) GPM + TRMM, rainfall retrievals, and (c) H-SAF rainfall retrievals on 29 September 2018 at 15:00 UTC.



**Figure 4.** (a) Radar-derived rainfall rates, (b) combined GPM and TRMM, rainfall retrievals, and (c) H-SAF rainfall retrievals on 29 September 2018 at 15:00 UTC.

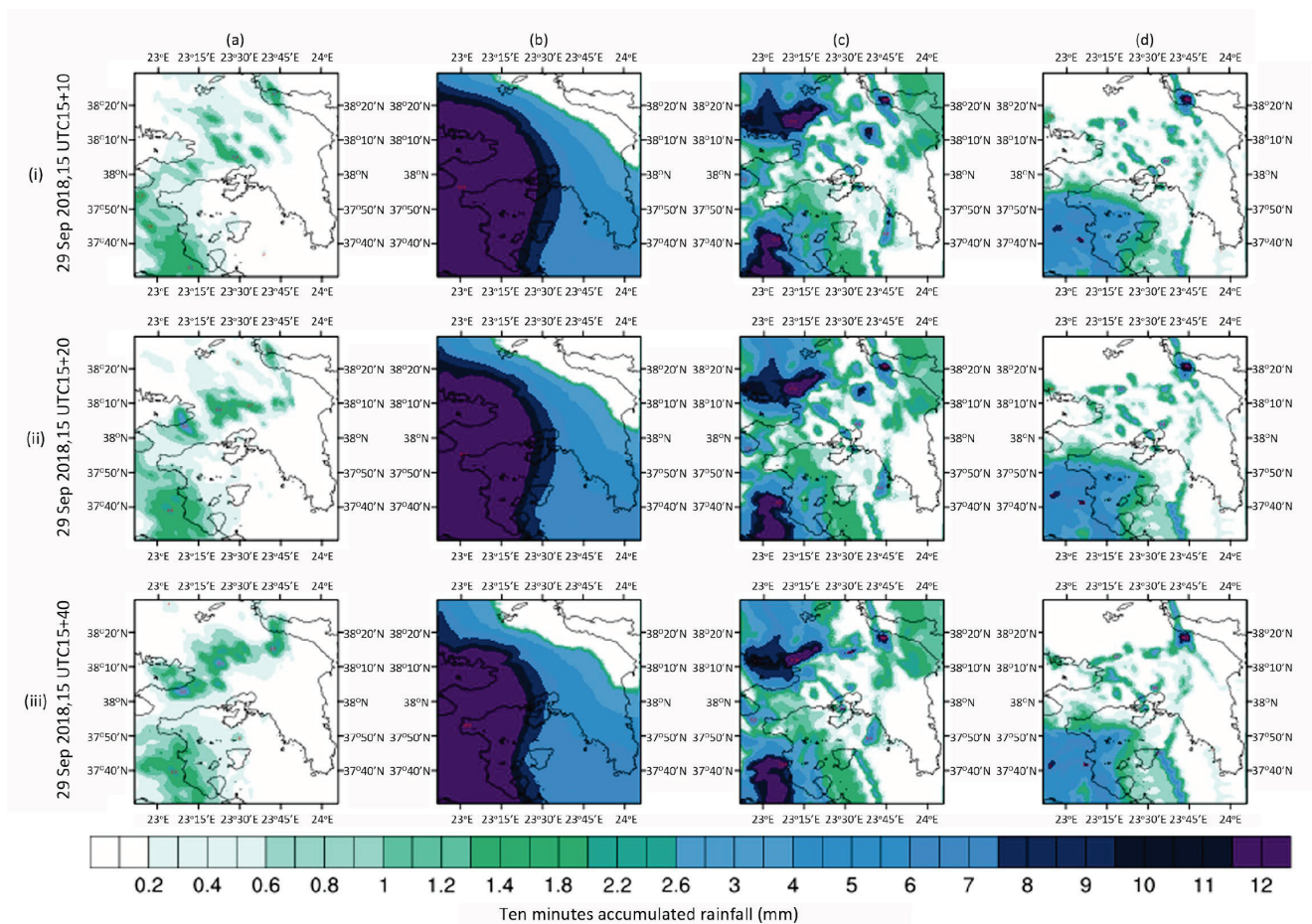


## 4. Results and Discussion

Two sets of experiments were conducted to assess the forward advection scheme's effectiveness in predicting a rainfall event in the Athens metropolitan area on 29 and 30 September 2018 with different rainfall intensities. The insights into the spatial distribution and variability of rainfall are presented in the panel plots for these experiments. The XPOL radar data are processed by the Self-Consistent Optimal Parameterization–Microphysics Estimation (SCOPE-ME) algorithm. This algorithm estimates rainfall rate from polarimetric radar variables and accumulates it over time. This process allows for consistent, independent radar-based rainfall accumulation estimations [29].

### 4.1. Spatial Pattern Comparison

Figure 5 illustrates the 10 min total rainfall in mm for radar data and the relevant LAPS nowcasts at 15:10, 15:20, and 15:40 UTC on 29 September 2018. Thus, each nowcast corresponds to 10, 20, and 40 min after the LAPS analysis on 29 September at 15:00 UTC. The panel plot (Figure 5) consists of twelve panel figures arranged in a  $3 \times 4$  grid. The leftmost panel is the XPOL radar data at the forecasting times, which are not assimilated into the LAPS, and the other three columns are the nowcasts that are produced using three different experiments: the control experiment (CTRL), the GTX experiment, and the HX experiment.



**Figure 5.** Ten minutes accumulated rainfall (mm) on 29 September 2018 with analysis time 15:00 UTC. The columns represent the following: (a) XPOL radar data (reference), (b) control experiment-CTRL, (c) GTX experiment, and (d) HX experiment, respectively. Each row of figures corresponds to (i) 10 min, (ii) 20 min, and (ii) 40 min ahead of the analysis time.

Figure 5 provides a comprehensive view of rainfall patterns. In the top panels (15:10 UTC), Figure 5a(i), which is the reference rainfall field from the radar data in its scanning sector, shows moderate rainfall concentrated in the center of the simulation domain, as well as in areas outside the basin, particularly to the southwest and west. The rainfall pattern from the HX experiment (Figure 5d(i)) is similar to the radar data, but heavier rainfall is presented in areas outside the basin, especially at the southwest part of the domain, while the Athens Basin has more rain cells. In contrast, the control experiment (CTRL) (Figure 5b(i)) depicts heavy rain across almost the entire domain, excluding a small northeastern part with no rainfall. The GTX experiment (Figure 5c(i)) indicates intensified rainfall throughout the area west of the Athens Basin with limited rain at the southeast, which is an intermediate behavior between the HX and control experiment—CTRL.

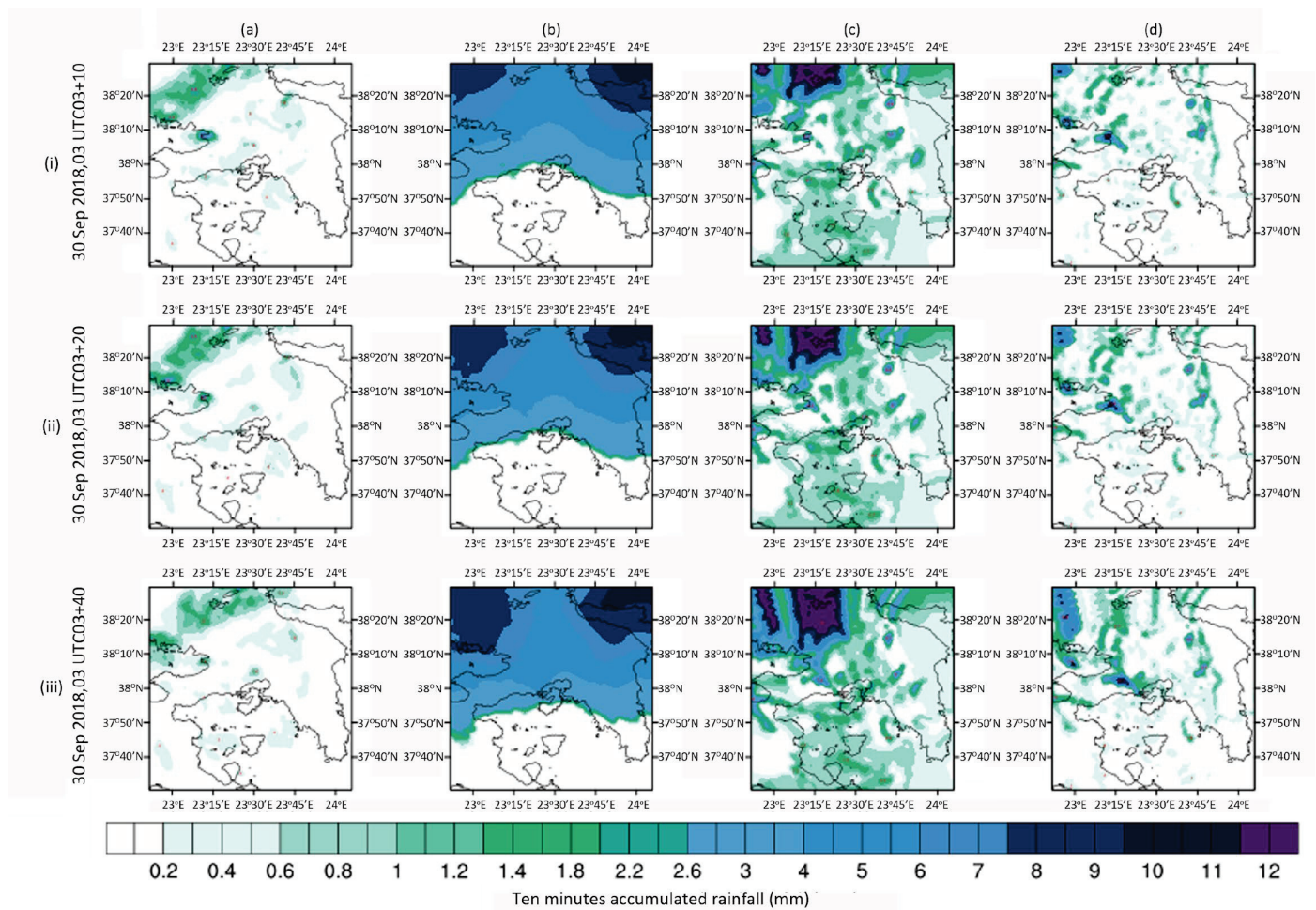
In the next lead times, we can see in the radar data that the rain cells at the northwest part of the domain are getting larger and intensifying. The GTX and HF experiments show a similar evolution with convergence of the rain cells in this area. The rain at the southwest part shows a small movement to the northwest and a decrease in its extension in the radar data, and all experiments are relative to the first lead time (top row of Figure 5, Figure 5(i)).

The middle panels present the next 10 min of accumulated rainfall. The radar data (Figure 5a(ii)) illustrate intensified rainfall compared to the first lead time (Figure 5a(i)), especially in the northwest part of the domain. Similarly, the HX experiment (Figure 5d(ii)) shows an increase in rainfall within the Athens Basin and the areas outside to the southwest and west, in comparison to the first lead time (Figure 5d(i)). However, the control experiment—CTRL (Figure 5b(ii))—shows a slightly smaller area of intensified rainfall than the first lead time (Figure 5b(i)). Moreover, the GTX experiment (Figure 5c(ii)) shows an increase in heavy rainfall outside the Athens Basin, particularly in the northwest and southwest, compared to the first lead time (Figure 5c(i)).

In the bottom panels (15:40 UTC), the radar data (Figure 5a(iii)) show increased rainfall compared to the previous left panels. The control experiment—CTRL (Figure 5b(iii))—shows a smaller area of intensified rainfall than the previous figures with a larger area in the northeast that has no rainfall. The GTX experiment (Figure 5c(iii)) depicts heavy rainfall, apart from a larger area in the south part of the domain. The HX experiment (Figure 5d(iii)) is similar to the radar data (Figure 5a(iii)) but displays different patterns with more scattered rainfall distribution. Throughout all figures, there is a noticeable northwestward movement of rainfall.

The second experiment, with lighter rainfall, was performed on 30 September 2018 at 03:10, 03:20, and 03:40 UTC, as shown in Figure 6 (each nowcast is for 10, 20, and 40 min after the LAPS analysis on 30 September at 03:00 UTC). On this day, the rainfall is weaker due to a less intense low-pressure system compared to the more intense event of the previous day. Similarly to Figure 5, the following panel plot shows the 10 min total rainfall in mm for radar data and the LAPS nowcasts.

In the top panels (03:10 UTC), the radar data (Figure 6a(i)) and the HX experiment (Figure 6d(i)) depict dispersed light to moderate rainfall occurrences at the Athens Basin, characterized by varying intensities. The HX experiment produces more rain cells than seen in the radar data, similarly with the result from the previous event shown in Figure 5. Moderate rainfall is observed in areas outside the Athens Basin particularly to the north of the Athens Basin. The control experiment—CTRL (Figure 6b(i))—shows a localized concentration of rainfall in the whole area north of the domain but no rain at the southern part. Conversely, the GTX experiment (Figure 6c(i)) illustrates the prevalence of widespread rainfall across the domain with intensified rainfall observed north–northwest of the Athens Basin.



**Figure 6.** Ten minutes accumulated rainfall (mm) on 30 September 2018 with analysis time 03:00 UTC. The columns represent the following: (a) XPOL radar data (reference), (b) control experiment (CTRL), (c) GTX experiment, and (d) HX experiment, respectively. Each row of figures corresponds to (i) 10 min, (ii) 20 min, and (iii) 40 min ahead of the analysis time.

In the next lead times, the radar data show that the rainfall field is moving slowly to the south with small changes. The same movement of the rainfall field can be seen in all the experiments with a small change in rain intensity. Thus, intense rain at the edge of the domain in the CTRL and GTX experiments is brought to its center, which is not observed in radar data where the rain pattern seems to move more slowly. Thus, the advection of the rainfall field is not exactly the same as the wind field indicates.

In the middle panels (03:20 UTC), the radar data (Figure 6a(ii)) show a slightly broader area of rainfall compared to the first lead time (Figure 6a(i)). Similarly to the radar data (Figure 6a(ii)), the HX experiment (Figure 6d(ii)) shows rainfall that appears to be more widespread than the first lead time (Figure 6d(i)). On the other hand, the control experiment—CTRL (Figure 6b(ii))—shows increased rainfall, suggesting a more intense rainfall event compared to the first lead time (Figure 6b(i)). Additionally, the GTX experiment (Figure 6c(ii)) shows increased rainfall intensity in the north–northwest of the Athens Basin compared to the first lead time (Figure 6c(i)).

In the bottom panels (03:40 UTC), the radar data (Figure 6a(iii)) show more rainfall in the northwest of the Athens Basin. In contrast, the control experiment—CTRL (Figure 6b(iii))—exhibits a significant area of rainfall with a heavier rainfall event, suggest-



ing a more intense event than observed in previous figures. Moreover, the GTX experiment (Figure 6c(iii)) displays a focused area of heavy rainfall in the northwest of the Athens Basin. Finally, the HX experiment (Figure 6d(iii)) shows scattered rainfall similar to the radar data (Figure 6a(iii)) but with distinct patterns. Throughout all figures, rainfall is observed moving in a south direction.

When comparing the two sets of experiments, on 29 September with an analysis time 15:00 UTC, rainfall is more concentrated, intense, and localized in the western and central parts of the domain. In contrast, on 30 September, with analysis time 03:00 UTC, the rainfall patterns show a shift with lighter and more dispersed rainfall, particularly in the northern and central part of the domain.

In the radar data, Figure 5a displays moderate rainfall more concentrated in the southwest and central areas of the domain, while Figure 6a shows rainfall lighter and more dispersed across the domain. In the control experiment (CTRL), Figure 5b shows widespread rainfall across almost the entire domain, excluding the northeastern part. In contrast, Figure 6b shows a weaker system with heavier rainfall in the northern area of the domain, while the southern area remains relatively dry. The GTX experiment in Figure 5c shows intense rainfall in the west and limited rain in the southeast. However, in Figure 6c, rainfall is still widespread across the domain but less intense overall, with a heavier concentration in the northern part of the domain. Lastly, in the HX experiment, Figure 5d shows intense rainfall and central parts of the domain, while in Figure 6d the rainfall is more scattered with reduced intensity, especially towards the southern part of the domain.

#### 4.2. Validation Against Ground Stations

Next, the independent rainfall rate observations from two meteorological stations in the Attica region (Salamina and Kallithea) operated by Harokopio University are compared to the 10 min rainfall nowcasts for both sets of experiments. Tables 5 and 6 present the results from an independent validation of radar-based nowcasts at the locations of ground weather stations. The comparison verifies that the nowcasts perform significantly better in the HX experiment when H-SAF satellite products are assimilated with XPOL radar data, as already concluded from Figures 5 and 6. It is important to mention that at both meteorological stations, the systematic overprediction of the rainfall in the CTRL experiment was significantly improved in the HX experiment (Tables 5 and 6).

**Table 5.** Comparison of the ten-minute LAPS rainfall nowcasts (mm) on 29 September 2018 with analysis time 15:00 UTC from 15:10 to 15:40 UTC and 30 September 2018 with analysis time 03:00 UTC from 3:10 to 3:40 UTC with implementing control experiment (CTRL), GTX experiment, and HX experiment and the recorded rainfall (mm) at Salamina station.

| Date              | Time (UTC) | Control<br>Experiment-CTRL | GTX Experiment | HX Experiment | Salamina |
|-------------------|------------|----------------------------|----------------|---------------|----------|
| 29 September 2018 | 15:10      | 10.9                       | 0.2            | 0.0           | 0.0      |
|                   | 15:20      | 9.8                        | 0.0            | 0.0           | 0.0      |
|                   | 15:30      | 8.8                        | 0.4            | 0.0           | 0.0      |
|                   | 15:40      | 8.0                        | 0.3            | 0.0           | 0.0      |
|                   | 03:10      | 0.0                        | 0.4            | 0.0           | 0.0      |
| 30 September 2018 | 03:20      | 0.0                        | 0.7            | 0.1           | 0.0      |
|                   | 03:30      | 0.3                        | 0.3            | 0.0           | 0.0      |
|                   | 03:40      | 2.9                        | 0.0            | 0.0           | 0.0      |
|                   |            |                            |                |               |          |

In the GTX experiment there is a systematic overestimation of recorded rainfall. On the other hand, in the HX experiment there is a slight tendency towards overestimating recorded rainfall. Specifically, at Salamina station on 29 September 2018, the HX experiment nowcasts no rainfall in every time slice. However, at the same station on 30 September 2018, there is a slight overestimation at 03:20 UTC with 0.1 mm/10 min. At Kallithea station, the HX experiment generally overestimates rainfall except for two time slices on 29 September 2018 at 15:40 UTC and on 30 September 2018 at 03:10 UTC, where there is an underestimation.

**Table 6.** As shown in Table 5, this refers to the recorded rainfall data from Kallithea station.

| Date              | Time (UTC) | Control Experiment—CTRL | GTX Experiment | HX Experiment | Kallithea |
|-------------------|------------|-------------------------|----------------|---------------|-----------|
| 29 September 2018 | 15:10      | 5.9                     | 0.0            | 0.0           | 0.0       |
|                   | 15:20      | 5.7                     | 0.0            | 0.3           | 0.2       |
|                   | 15:30      | 5.6                     | 0.4            | 0.3           | 0.0       |
|                   | 15:40      | 5.4                     | 0.8            | 0.0           | 0.2       |
|                   | 03:10      | 2.8                     | 0.5            | 0.0           | 0.2       |
| 30 September 2018 | 03:20      | 2.8                     | 0.7            | 0.2           | 0.0       |
|                   | 03:30      | 3.0                     | 0.3            | 0.2           | 0.2       |
|                   | 03:40      | 3.0                     | 0.3            | 0.3           | 0.0       |

The primary rainfall core, however, was situated away from these two stations. Consequently, an additional comparison is conducted with three other independent meteorological stations within the Hellenic National Meteorological Service’s network. The results between the nowcasts at 12:00 UTC for 29 and 30 September 2018 are presented in Tables 7 and 8. The nowcasts correspond to 6 h ahead from the analysis time at 06:00 UTC.

**Table 7.** Comparison of the LAPS rainfall nowcasts in mm for 12:00 UTC on 29 September 2018, implementing the control experiment (CTRL), GTX experiment, and HX experiment and the recorded rainfall in mm at Elefsis, Elliniko, and Tatoi stations (the station measurements retrieved from the ECMWF).

| Time (UTC) | Station  | Control Experiment—CTRL | GTX Experiment | HX Experiment | Past 6h Rainfall |
|------------|----------|-------------------------|----------------|---------------|------------------|
| 12:00      | Elefsis  | 0.0                     | 0.1            | 0.1           | 7.0              |
| 12:00      | Elliniko | 0.0                     | 0.0            | 0.0           | 11.0             |
| 12:00      | Tatoi    | 0.0                     | 0.2            | 0.1           | 8.0              |

**Table 8.** As in Table 7, but on 30 September 2018 at 12:00 UTC.

| Time (UTC) | Station  | Control Experiment—CTRL | GTX Experiment | HX Experiment | Past 6h Rainfall |
|------------|----------|-------------------------|----------------|---------------|------------------|
| 12:00      | Elefsis  | 32.3                    | 2.6            | 2.3           | 17.0             |
| 12:00      | Elliniko | 31.6                    | 2.4            | 1.8           | 0.9              |
| 12:00      | Tatoi    | 31.8                    | 13.1           | 12.3          | 9.0              |

On 29 September 2018, the control experiment’s nowcasts show no rainfall at any of these stations. There is a significant underestimation from all experiments compared to the observed in situ data across all stations. At Elliniko station, in particular, the nowcasts of all experiments predict no rainfall despite a recorded value of 11.0 mm/6 h.

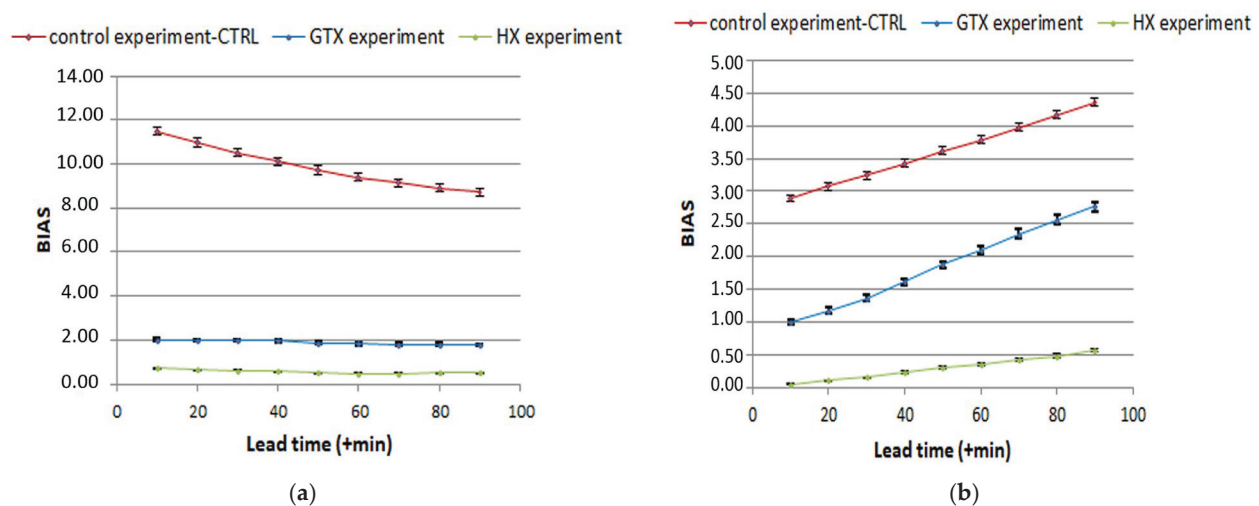


On the other hand, on 30 September 2018, the nowcasts of the control experiment had an overestimation of rainfall at all stations. Specifically, rainfall is overestimated in comparison to the in situ data at the Tatoi and Elliniko stations. With the exception of the Elefsis station, where the nowcast is 2.3 mm/6 h rather than the recorded value of 17 mm/6 h, the nowcasts of the HX experiment are actually closer to the values of the independent stations.

Recognizing the limitation of comparing with two initial rain gauge stations (Salamina and Kallithea), additional comparisons are conducted using another three independent meteorological stations (Elliniko, Tatoi, and Elefsis). While these additional stations provided further validation, the overall number of ground weather stations remains insufficient to derive reliable conclusions about the overall performance of the nowcasts.

The surface-based validation presented here relies on a small number of meteorological stations, which limits the spatial representativeness of the results. Some stations also recorded near-zero rainfall, which further complicates the interpretation. Therefore, the conclusions drawn from these point-based comparisons should be treated in conjunction with the more reliable statistical scores against the XPOL radar estimates presented in Section 4.3. However, a denser observation network would be necessary to evaluate system performance more comprehensively.

An important point in the validation results is the underestimation observed in the 6 h nowcasts on 29 September (Table 7), which seems to contradict the general overestimation bias that is identified for shorter lead times in the point-to-point analysis of Section 4.3 (Figure 7a). This discrepancy likely arises from the limitations of the forward advection scheme at longer lead times. As the forecast horizon extends beyond three hours, dynamic processes not represented by the advection model begin to dominate rainfall development. As a result, the model's ability to track rainfall diminishes, often leading to underestimation.



**Figure 7.** BIAS score (mm) and corresponding 95% confidence intervals derived from bootstrap sampling for rainfall forecasts on (a) 29 September 2018 with analysis time 15:00 UTC; (b) 30 September 2018 with analysis time 03:00 UTC.

Given the sparse station density and the presence of near-zero rainfall values, these station-based results are best considered a supplement to the more spatially comprehensive grid-to-grid validation against XPOL radar data in the sections that follow.

#### 4.3. Point-Wise Statistics

To quantitatively evaluate forecast accuracy, a point-to-point statistical evaluation is employed between the LAPS-nowcasts and the XPOL radar data (in the forecast time period) that are not assimilated in the LAPS. Specifically, XPOL radar data at analysis time  $T$  (e.g., 15:00 UTC) are assimilated into the LAPS. The nowcasts at  $T + 10$  to  $T + 90$  min are then validated against independent XPOL radar observations at those times. This radar data used for validation are not assimilated into the system, ensuring an independent assessment. The nowcasts are applied with lead times of 10 to 90 min and analysis time 15:00 UTC on 29 September 2018 and 03:00 UTC on 30 September 2018. The calculated BIAS and BIAS scores, as presented in Figure 7a,b, indicate significant differences among these experiments. More specifically, the HX experiment consistently exhibits the lowest BIAS scores, while the control experiment (CTRL) shows the highest. Across all experiments, there is a tendency to overestimate recorded rainfall, as BIAS scores are positive.

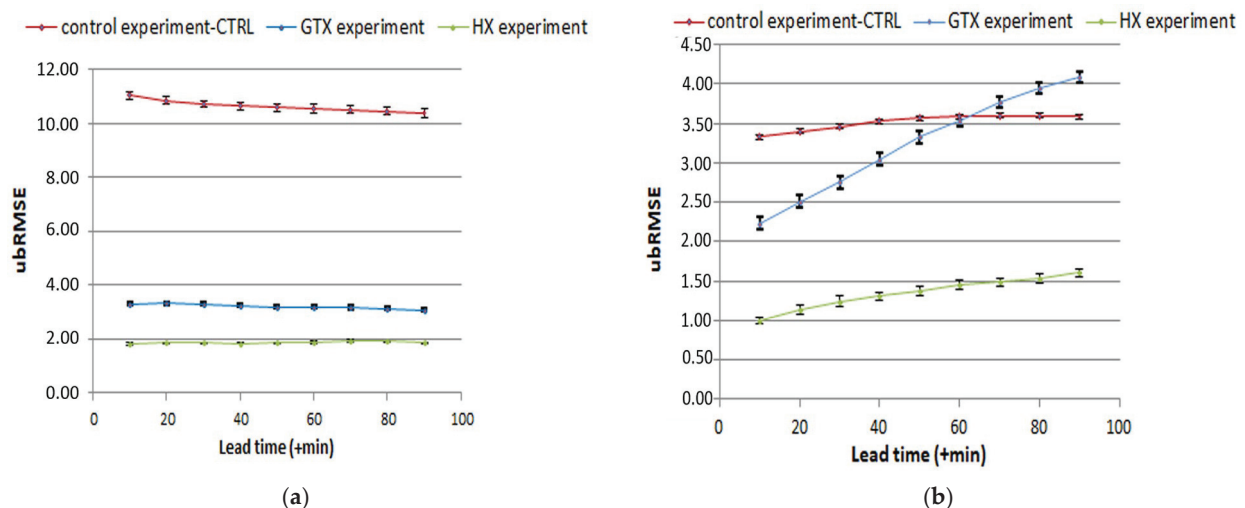
On 29 September 2018, with analysis time 15:00 UTC (Figure 7a), there is a trend of decreasing BIAS scores as lead time increases from 10 to 90 min in all experiments. The HX experiment consistently achieves the lowest BIAS scores across all lead times, demonstrating superior performance in terms of rainfall prediction accuracy. The BIAS for the HX experiment decreased steadily from 0.71 mm at a 10 min lead time to 0.49 mm at 70 min, with slight variations thereafter. In contrast, the GTX experiment shows significantly higher BIAS scores, ranging from 2.02 mm to 1.77 mm, while the control consistently exhibits the largest biases, starting at 11.48 mm and declining to 8.71 mm. Notably, the confidence intervals for HX remains narrow throughout, underscoring the robustness of its performance.

Although this reduction in BIAS might initially suggest improved forecast skill, it is important to emphasize that this apparent improvement can be misleading. Forecast quality generally degrades with increasing lead time; thus, the declining BIAS likely reflects a statistical artifact rather than a genuine gain in predictive accuracy. Two key mechanisms contribute to this behavior: (1) the advection of intense rainfall cells out of the verification domain, which mechanically reduces the domain-averaged rainfall and hence lowers the positive BIAS; (2) the positive–negative error cancelation, where overpredicted and underpredicted rainfall regions offset each other, resulting in an artificial reduction in overall bias. This balancing effect of positive and negative errors with the advection of the mostly scattered (rain cells) rainfall field could be misleading to an apparent improvement in nowcasting accuracy. Initially, the nowcasts may show high bias due to intensified rainfall in certain areas, but as the system moves, the differences between the observed and forecasted rainfall on average diminish. For the control experiment (CTRL), the incoming lighter rainfall from the southeast and outgoing higher rainfall from the northwest reduce the significant differences initially observed in the control experiment (CTRL), which have very high values and decreased BIAS scores.

On 30 September 2018 (analysis time 03:00 UTC), the BIAS scores from the HX experiment gradually increases from 0.05 mm to 0.56 mm as lead time extends from 10 to 90 min (Figure 7b). This trend shows a decrease in the accuracy of the predictions over time. At the shortest lead time (+10 min), the HX experiment exhibits the lowest bias of 0.05 mm (95% CI: 0.03–0.07), compared to 1.00 mm (95% CI: 0.96–1.03) for GTX and 2.89 mm (95% CI: 2.84–2.94) for the control experiment (CTRL). As the lead time increases to +90 min, the BIAS for HX is 0.56 mm (95% CI: 0.53–0.59), yet it remains substantially lower than GTX (2.76 mm) and the control experiment (CTRL) (4.36 mm). The steady, low-bias scores of HX underscore its accuracy in terms of quantitative rainfall nowcasting, significantly outperforming both the baseline and the other experiment. Overall, the HX experiment

demonstrates a noteworthy enhancement in nowcast accuracy over time compared to the other two experiments.

Moreover, the unbiased Root Mean Square Error (ubRMSE) scores that are presented in Figure 8a,b illustrate the performance differences among the experiments in terms of random errors. On 29 September 2018 (analysis time 15:00 UTC), the HX experiment maintained the lowest ubRMSE across all lead times, although the values vary in a non-monotonic manner (Figure 8a). At 10 min lead time, the HX experiment has the lowest ubRMSE score of 1.83 mm (95% CI: 1.79–1.87 mm), stabilizing near 1.85 mm until 50 min before rising to 1.92 mm at 70 min, indicating a substantial decrease in prediction accuracy. In contrast, the control experiment (CTRL) shows little-decreasing ubRMSE scores as lead time increases. The GTX experiment, on the other hand, shows increasing ubRMSE scores for the first 20 min followed by a decrease, but with overall larger values compared to the HX experiment. The GTX experiment shows the lowest ubRMSE score of 3.08 mm at 90 min lead time and the highest score of 3.32 mm at 20 min lead time. As mentioned previously, this may be due to the advection of the rainfall field. Overall, the HX experiment consistently performs better and exhibits lower errors in terms of ubRMSE scores.



**Figure 8.** Unbiased RMSE (ubRMSE) score (mm) and corresponding 95% confidence intervals derived from bootstrap sampling for rainfall forecasts on (a) 29 September 2018 with analysis time 15:00 UTC; (b) 30 September 2018 with analysis time 03:00 UTC.

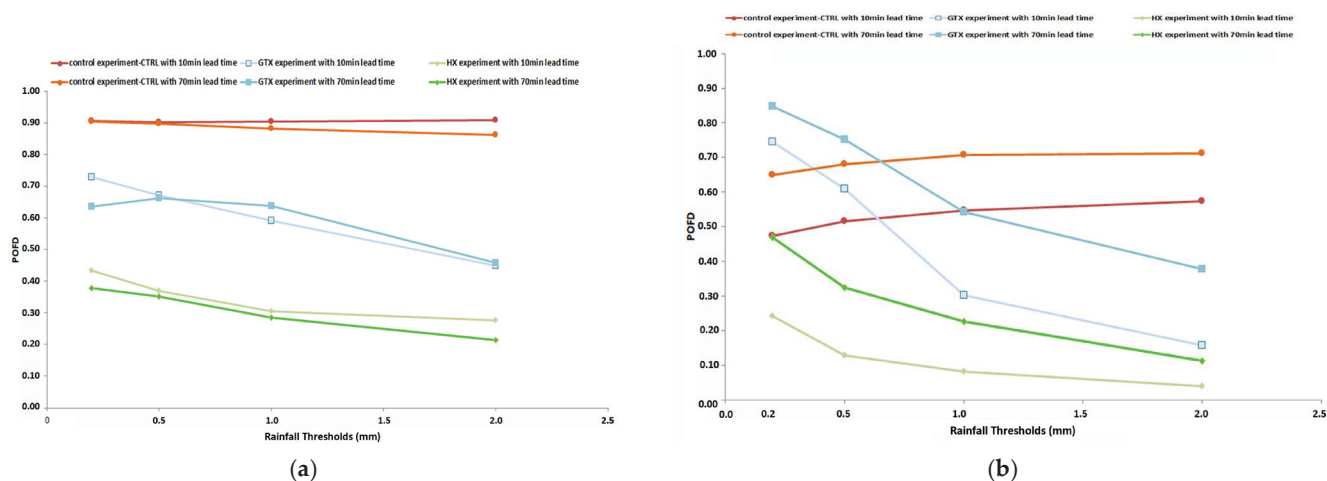
For the 30 September event (analysis: 03:00 UTC), the HX experiment exhibits a gradual increase in ubRMSE with lead time, rising from 1.00 mm (10 min forecast, 95% CI: 0.95–1.04 mm) to 1.60 mm (90 min forecast), reflecting diminishing precision as the forecast horizon extends (Figure 8b). Despite this decline, HX maintains superior accuracy compared to CTRL and GTX, both of which show ubRMSE values exceeding 2.0 mm, indicative of systematic overestimation. The GTX experiment, while outperforming CTRL, still demonstrates a ubRMSE score from 2.23 mm (10 min) to 4.08 mm (90 min), with wider confidence intervals than HX. From all of the above, it is concluded that the HX experiment performs better than the GTX and the CTRL experiments in maintaining prediction accuracy over time. While all experiments show a tendency to overestimate rainfall, the HX experiment demonstrates the most accurate nowcasts, making it the most reliable experiment. However, as the lead time increases, the nowcasting performance decreases.

The analysis of the rainfall event on 29 and 30 September 2018 reveals significant improvements in forecast lead time when assimilating radar and satellite data into the nowcasting system. For the first experiment (29 September), characterized by intense

rainfall, the HX experiment provides the most accurate nowcasts. The HX experiment consistently exhibits the lowest BIAS and unbiased RMSE scores, demonstrating a reduction in overestimation and improved spatial accuracy compared to the control experiment (CTRL). For the second experiment (30 September), which featured lighter and more dispersed rainfall, the HX experiment again outperforms the other two experiments, particularly at shorter lead times, where it achieves the highest accuracy and lowest BIAS scores. However, as the lead time increases, the performance of all experiments decreases, with the HX experiment maintaining a relative advantage. These findings indicate that the assimilation of high-resolution satellite data (H-SAF) with radar observations significantly enhances short-term rainfall forecasts. The improvements are more pronounced in intense rainfall events, where the higher temporal and spatial resolution of H-SAF data provides a more accurate representation of rainfall dynamics. This suggests that the integration of advanced remote sensing data can extend the effective forecast lead time for nowcasting systems, particularly in high-impact weather scenarios.

#### 4.4. Detection Skill: Probability of False Detection (POFD) and Accuracy

On 29 September 2018 (analysis time 15:00 UTC), the Probability of False Detection (POFD) varied significantly across experiments and rainfall thresholds (0.2 mm to 2 mm) at both 10 min and 70 min lead times, as shown in Figure 9a. The control experiment (CTRL) exhibits a stable POFD score ranging between 0.86 and 0.91 across all rainfall thresholds, indicating limited skill in distinguishing between true and false rainfall detections. In contrast, the GTX and the HX experiments show reduced POFD scores, especially at higher thresholds, highlighting an improved ability to detect false alarms as rainfall intensity increases. The HX experiment stands out with the lowest POFD values, demonstrating superior capability to correctly identify significant rainfall. While there is a slight increase in POFD score from the 10 min to the 70 min lead time, HX consistently outperformed both GTX and CTRL in minimizing false detections across all thresholds.

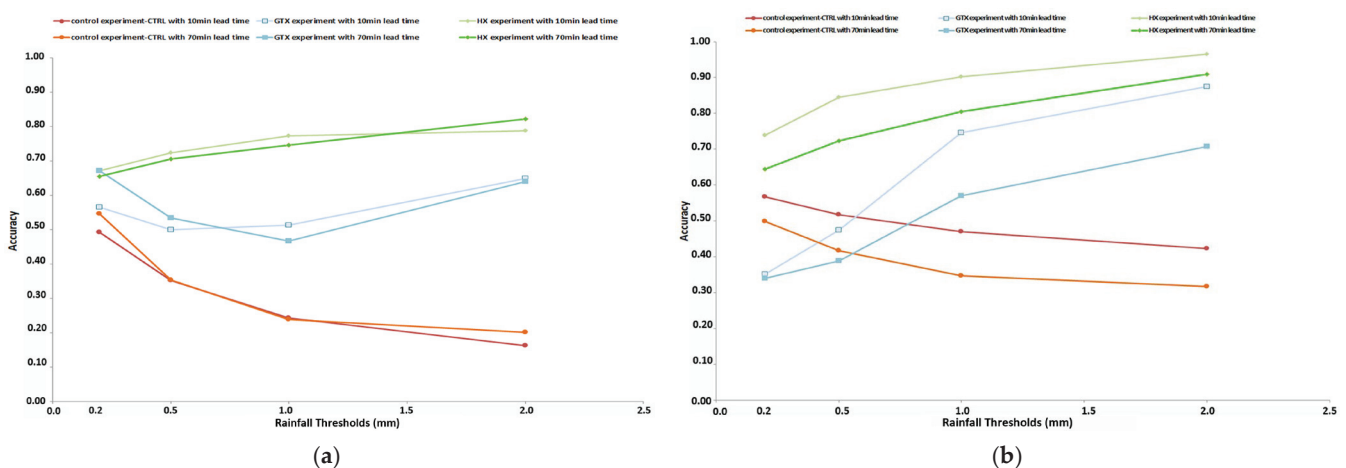


**Figure 9.** POFD metrics for different thresholds of control experiment (CTRL), GTX experiment, and HX experiment at 10 min and 70 min lead times on (a) 29 September 2018 with analysis time 15:00 UTC; (b) 30 September 2018 with analysis time 03:00 UTC.

On 30 September 2018 (analysis time 03:00 UTC), Figure 9b shows the POFD scores at both 10 min and 70 min lead times. The control experiment (CTRL) exhibits a stable range between 0.47 and 0.71 across all rainfall thresholds, indicating consistency. At the 10 min lead time, the control experiment (CTRL) exhibits a moderate POFD range (0.47–0.57), demonstrating a lack of differentiation in prediction accuracy across thresholds. In contrast,

the GTX experiment exhibits a high POFD score of 0.75 at the 0.2 mm threshold but shows a steep decline to 0.16 at 2 mm, suggesting an over-detection tendency for lighter rainfall and improved discrimination for heavier events. The HX experiment, however, demonstrated the most favorable scores—starting at 0.24 and dropping sharply to 0.04 at 2 mm—indicating the strong dictation of false alarms. At a 70 min lead time, all experiments showed increased POFD values, signaling a general decline in forecast accuracy over time. The control experiment (CTRL) ranges from 0.65 to 0.71, and the GTX experiment ranges from 0.85 at the lowest threshold down to 0.38 at 2 mm. Meanwhile, the HX experiment sustained its advantage with values between 0.47 and 0.11, confirming its ability to maintain lower false alarm rates for higher rainfall thresholds, even as lead time increased. Overall, both figures underscore the superior performance of the HX experiment in minimizing false alarms across all thresholds and lead times. The control experiment (CTRL) maintains a stable behavior across rainfall thresholds, while the GTX experiment improves with rainfall intensity.

Furthermore, on 29 September 2018 (analysis time 15:00 UTC), Figure 10a shows the accuracy scores at 10 min and 70 min lead times. The HX experiment demonstrates a significant increase in accuracy scores from light to moderate rainfall thresholds ranging between 0.65 and 0.82. The accuracy scores decrease as lead time increases, except for the 2.00 mm threshold, where the accuracy score increases a little with lead time. At the 10 min lead time, the HX experiment exhibits the highest accuracy scores, increasing from 0.67 at 0.2 mm to 0.79 at 2 mm, indicating strong capability to detect heavier rainfall events. In contrast, the GTX experiment exhibits a wider range of accuracy scores varying between 0.57 and 0.65. The control experiment (CTRL) shows much lower and decreasing accuracy values, from 0.49 to 0.16, suggesting reduced effectiveness in identifying higher-intensity rainfall. At the 70 min lead time, the HX experiment maintains superior performance, with accuracy scores from 0.65 to 0.82, showing a continued increase with rainfall intensity. The CTRL experiment remained relatively weak and unchanged from the 10 min lead time, with values ranging from 0.55 at 0.2 mm to 0.20 at 2 mm. The GTX experiment, while showing some improvement at lower thresholds, is still behind the HX experiment.



**Figure 10.** Accuracy metrics for different thresholds of control experiment (CTRL), GTX experiment, and HX experiment at 10 min and 70 min lead times on (a) 29 September 2018 with analysis time 15:00 UTC; (b) 30 September 2018 with analysis time 03:00 UTC.

Similarly, Figure 10b shows the accuracy scores for 10 min and 70 min lead times on 30 September 2018 with analysis time 03:00 UTC. Both the GTX and the HX experiments show a significant increase in accuracy scores from light to moderate rainfall thresholds.



At the 10 min lead time, the GTX and HX experiments show a gradual improvement as the threshold increases, with the HX experiment achieving the highest performance across all threshold levels. In contrast, the control experiment (CTRL) shows a gradual decrease in performance as the threshold increases, suggesting reduced effectiveness. At the 70 min lead time, the GTX and HX experiments continue to improve, although at lower values compared to the 10 min lead time. Specifically, the HX experiment maintains the highest performance. Meanwhile, the control experiment (CTRL) displays a more stable behavior varying between 0.32 and 0.57, across the precipitate thresholds, as was seen in POFD. In conclusion, both figures highlight the HX experiment's exceptional ability to identify rainfall events at all thresholds and lead times.

In conclusion, the HX experiment demonstrates the most accurate nowcasts regarding both nowcasts and warnings, making it the most reliable experiment. However, as the lead time increases, the nowcasting performance decreases.

#### 4.5. Summary and Interpretation

This study demonstrates that the forward advection scheme implemented in the Local Analysis and Prediction System (LAPS) incorporated with satellite-derived rainfall retrievals and XPOL radar data significantly improves nowcasting accuracy.

Across all analysis dimensions—spatial structure, ground validation, bias, error metrics, and detection skill—the HX experiment consistently outperformed the other configurations. The lower BIAS in the HX experiment indicates a reduced systematic overestimation of rainfall, a common issue in the control experiment. This improvement suggests that the assimilation of remote sensing data, particularly H-SAF, helps to constrain the model's tendency towards overprediction. The HX experiment led to a significant improvement in short-term rainfall forecasts, showing a 60% reduction in ubRMSE error at a 60 min lead time and a 55% reduction in ubRMSE error at a 90 min lead time compared to the control experiment—CTRL. The lower ubRMSE signifies a reduction in random errors, pointing to a more spatially accurate and reliable forecast. Moreover, the better POFD and accuracy scores of the HX experiment, especially at higher rainfall thresholds, indicate an enhanced ability to correctly detect significant rainfall events and minimize false alarms, which are crucial for effective early warning systems in flood-prone areas like Attica.

The comparison between the two sets of experiments (29 and 30 September) reveals that the benefits of assimilating high-resolution satellite data are evident for both intense and lighter rainfall events, although the magnitude of improvement may vary. This suggests that the sensitivity to remote sensing data is not limited to specific types of rainfall events but has a broader positive impact on nowcasting accuracy.

While the methodology employed a forward advection scheme within the LAPS, the results demonstrate that the choice of assimilated data significantly influences the scheme's performance. The optimized integration of high-resolution radar with the broad and frequent coverage of H-SAF satellite data proves to be a key factor in enhancing short-term predictability in this region. The methodology developed in this study is adaptable to different geographical regions and data availability, highlighting its potential for widespread application in nowcasting systems.

Although this study does not include a direct benchmark against external nowcasting systems such as pySTEPS or optical flow-based techniques, it provides the relative value of different data assimilation configurations within the LAPS framework. While focused on a medicane-type event, the methodology should be tested across a broader range of rainfall patterns, such as frontal systems. Future work will aim to validate the approach under different meteorological scenarios to better assess its operational generalizability.

## 5. Conclusions

Assimilating remote sensing in NWP models can improve the accuracy of rainfall estimates. In this study, a rainfall event that affected the Athens metropolitan area in Greece on 29 and 30 September 2018 is analyzed through two set of experiments, each focusing on different intensities; this was used as a case study for the implementation of a forward advection scheme in the LAPS as a system for early warnings. In nowcasting mode, several remote sensing measurements are assimilated by the LAPS, including weather radar data to produce objective analyses. Applying the methodology presented in Section 2, the produced fields were advected and the outputs were presented in Section 4. It was found that assimilating H-SAF with radar rainfall retrievals (HX experiment) substantially improved the short-term predictability of rainfall, offering a substantial improvement of almost 60% against the control experiment for the 60 min and 55% for the 90 min unbiased RMSE of the nowcasted rainfall. Indeed, the percentage increases by 10 points at elevations less than or equal to 300 m compared to higher elevations. This is an encouraging outcome as lower elevations are the most flood-prone and vulnerable populated areas. Entire control nowcasts with GFS near to the analysis forecasts as background fields with METAR and SYNOP observations systematically overestimated the rainfall rate (control experiment—CTRL). Moreover, the assimilation of TRMM and IMERG products with radar data (GTX experiment) also gives nowcasts with a significant systematic overestimation of the rainfall rate.

This study demonstrates that integrating satellite-derived rainfall retrievals with radar data significantly enhances short-term rainfall forecasts, outperforming traditional NWP models and single-source assimilation methods. The LAPS-based forward advection scheme effectively reduces bias and improves accuracy in nowcasting weather events. Among the tested configurations, the HX experiment—incorporating H-SAF satellite data—delivered the most reliable results, highlighting the advantages of higher-temporal-resolution satellite retrievals. These findings emphasize the importance of multi-source data assimilation for real-time forecasting, particularly in regions with complex rainfall dynamics. Although this case study was conducted in the Attica region, the results underscore the potential applicability of the system to other geographies. The LAPS is modular and capable of ingesting various radar and satellite inputs. Beyond this case study, our approach has broader applications in operational weather forecasting and disaster management, providing a scalable framework for improving short-term rainfall predictions.

Assimilating H-SAF rainfall estimates with radar data substantially improved the short-term predictability of rainfall compared with independent measurements. In most of the runs, the predictability of the nowcasts decreased with the lead time. Combining radar data with H-SAF and IMERG-TRMM products, rainfall nowcasts were improved due to their higher resolution in both space and time. H-SAF products have 15 min temporal resolution, while IMERG and TRMM have a 30 min and 3 h temporal resolution, respectively. Moreover, HSAF products have 8 km spatial resolution over Europe, while IMERG and TRMM have 10 km and 25 km spatial resolution, respectively.

In general, using a forward advection scheme in the LAPS which incorporated H-SAF and radar data had benefits such as more precise and timely rainfall forecasts. The scheme's adaptability, which allowed for nowcasts to be provided using any available data, was an advantage. The radar data, usable in real time, were the most significant source of rainfall information. In the case of the absence of the radar data, the LAPS with background fields would also provide early warnings. This scheme could predict rainfall's path and potential impact, allowing the authorities and individuals to take the necessary precautions to protect themselves and minimize damage. People could be better prepared

and make better choices to mitigate potential risks and minimize negative impacts by using nowcasting as a system of early warnings.

It should be noted that the results of this study are based on a case study of a single meteorological event including different rainfall patterns, one with intense, localized rainfall and another with more scattered rainfall. While this provides some variability, future work will expand this testing across more diverse meteorological scenarios. In order to completely assess the system's capabilities, additional high-impact events have to be tested. Future work will focus on extending this methodology to additional storm events, refining computational efficiency, and integrating machine learning techniques to further enhance forecasting capabilities. Additionally, the advection scheme will be enhanced by incorporating the effects of orography on rainfall distribution.

**Author Contributions:** A.P. contributed to the methodology, formal analysis investigation, visualization, and writing—original draft preparation, M.T., J.K., C.S., M.N.A., C.K. and G.V. contributed to the methodology, review, and editing; P.K. contributed to the conceptualization, methodology, supervision, investigation, review, and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** The National Centers for Environmental Prediction (NCEP) is gratefully acknowledged for the provision of the Global Forecasting System (GFS) operational analyses. The National Aeronautics and Space Administration (NASA) for the provision of GPM and TRMM data and the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) for the provision of H-SAF data. The European Centre for Medium-Range Weather Forecasts (ECMWF) for the provision of the station measurements.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Wilson, J.; Crook, N.; Mueller, C.; Sun, J.; Dixon, M. Nowcasting thunderstorms: A status report. *Bull. Am. Meteorol. Soc.* **1998**, *79*, 2079–2100.
2. Wilson, J.; Ebert, E.; Saxen, T.; Roberts, R.; Mueller, C.; Sleigh, M.; Pierce, C.; Seed, A. Sydney 2000 forecast demonstration project: Convective storm nowcasting. *Weather Forecast.* **2004**, *19*, 131–150. [CrossRef]
3. Liguori, S.; Rico-Ramirez, M. Quantitative assessment of short-term rainfall forecasts from radar nowcasts and MM5 forecasts. *Hydrol. Process.* **2012**, *26*, 3842–3857. [CrossRef]
4. Liu, Z.; Dai, Q.; Zhuo, L. Relationship between Rainfall Variability and the Predictability of Radar Rainfall Nowcasting Models. *Atmosphere* **2019**, *10*, 458. [CrossRef]
5. Schmid, F.; Wang, Y.; Harou, A. Nowcasting Guidelines—A Summary. *Bulletin* **2019**, *68*, 2.
6. Sandery, P. Data assimilation cycle length and observation impact in mesoscale ocean forecasting. *Geosci. Model Dev.* **2018**, *11*, 4011–4019. [CrossRef]
7. Antonini, A.; Melani, S.; Corongiu, M.; Romanelli, S.; Mazza, A.; Ortolani, A.; Gozzini, B. On the Implementation of a Regional X-Band Weather Radar Network. *Atmosphere* **2017**, *8*, 25. [CrossRef]
8. Sokol, Z.; Szturc, J.; Orellana-Alvear, J.; Popová, J.; Jurczyk, A.; Céleri, R. The Role of Weather Radar in Rainfall Estimation and Its Application in Meteorological and Hydrological Modelling—A Review. *Remote Sens.* **2021**, *13*, 351. [CrossRef]
9. Ombadi, M.; Nguyen, P.; Sorooshian, S.; Hsu, K. How much information on precipitation is contained in satellite infrared imagery? *Atmos. Res.* **2021**, *256*, 105578. [CrossRef]

10. Sokol, Z. Effects of an assimilation of radar and satellite data on a very-short range forecast of heavy convective rainfalls. *Atmos. Res.* **2009**, *93*, 188–206. [CrossRef]
11. Bae, J.-H.; Min, K.-H. Forecast Characteristics of Radar Data Assimilation Based on the Scales of Precipitation Systems. *Remote Sens.* **2022**, *14*, 605. [CrossRef]
12. Sokol, Z.; Zacharov, P. Nowcasting of precipitation by an NWP model using assimilation of extrapolated radar reflectivity. *Q. J. R. Meteorol. Soc.* **2012**, *138*, 1072–1082. [CrossRef]
13. He, X.; Zhang, Y.; Wang, H. Application of Radar Data Assimilation on Convective Precipitation Forecasts Based on Water Vapor Retrieval. *Adv. Atmos. Sci.* **2021**, *38*, 1457–1468. [CrossRef]
14. Liu, P.; Yang, Z.; Wang, X.; Qiu, X.; Yang, Y. Assimilation of the Pseudo-Water Vapor Derived from Extrapolated Radar Reflectivity to Improve the Forecasts of Convective Events. *Atmospheric Research* **2022**, *279*, 106386. [CrossRef]
15. Lai, C.; Ban, J.; Sun, J. Assimilation of Radar Data, Pseudo Water Vapor, and Potential Temperature in a 3DVAR Framework for Improving Precipitation Forecast of Severe Weather Events. *Atmos. Res.* **2020**, *11*, 182. [CrossRef]
16. Kumar, P.; Kishtawal, C.; Pal, P. Impact of Satellite Rainfall Assimilation on Weather Research and Forecasting Model Predictions over the Indian Region. *J. Geophys. Res.* **2014**, *119*, 2017–2031. [CrossRef]
17. Torcasio, R.C.; Papa, M.; Del Frate, F.; Mascitelli, A.; Dietrich, S.; Panegrossi, G.; Federico, S. Data Assimilation of Satellite-Derived Rain Rates Estimated by Neural Network in Convective Environments: A Study over Italy. *Remote Sens.* **2024**, *16*, 1769. [CrossRef]
18. Wang, K.; Kong, L.; Yang, Z.; Singh, P.; Guo, F.; Xu, Y.; Tang, X.; Hao, J. GPM Annual and Daily Precipitation Data for Real-Time Short-Term Nowcasting: A Pilot Study for a Way Forward in Data Assimilation. *Water* **2021**, *13*, 1422. [CrossRef]
19. Pan, S.; Gao, J.; Jones, T.A.; Wang, Y.; Wang, X.; Li, J. The Impact of Assimilating Satellite-Derived Layered Precipitable Water, Cloud Water Path, and Radar Data on Short-Range Thunderstorm Forecasts. *Mon. Weather Rev.* **2021**, *149*, 1359–1380. [CrossRef]
20. Pan, S.; Gao, J.; Stensrud, D.J.; Wang, X.; Jones, T.A. Assimilation of Radar Radial Velocity and Reflectivity, Satellite Cloud Water Path, and Total Precipitable Water for Convective-Scale NWP in OSSEs. *J. Atmos. Ocean. Technol.* **2017**, *35*, 67–89. [CrossRef]
21. Jones, T.A.; Stensrud, D.J.; Wicker, L.; Minnis, P.; Palikonda, R. Simultaneous radar and satellite storm-scale data assimilation using an ensemble Kalman filter approach for 24 May 2011. *Mon. Weather Rev.* **2015**, *143*, 165–194. [CrossRef]
22. Jones, T.A.; Knopfmeier, K.; Wheatley, D.; Creager, G.; Minnis, P.; Palikonda, R. Storm-scale data assimilation and ensemble forecasting with the NSSL experimental warn-on-forecast system. Part II: Combined radar and satellite data experiments. *Weather Forecast.* **2016**, *31*, 297–327. [CrossRef]
23. Pulkkinen, S.; Nerini, D.; Pérez Hortal, A.A.; Velasco-Forero, C.; Germann, U.; Seed, A.; Foresti, L. Pysteps: An Open-Source Python Library for Probabilistic Precipitation Nowcasting (v1.0). *Geosci. Model Dev.* **2019**, *12*, 4185–4219. [CrossRef]
24. An, S.; Oh, T.J.; Sohn, E.; Kim, D. Deep Learning for Precipitation Nowcasting: A Survey from the Perspective of Time Series Forecasting. *Expert Syst. Appl.* **2025**, *268*, 126301. [CrossRef]
25. Albers, S.C. The LAPS Wind Analysis. *Weather Forecast.* **1995**, *10*, 342–352. [CrossRef]
26. Birkenheuer, D. The Effect of Using Digital Satellite Imagery in the LAPS Moisture Analysis. *Weather Forecast.* **1999**, *14*, 782–788. [CrossRef]
27. Jian, G.-J.; Shieh, S.-L.; McGinley, J.A. Precipitation Simulation Associated with Typhoon Sinlaku (2002) in Taiwan Area Using the LAPS Diabatic Initialization for MM5. *Terr. Atmos. Ocean. Sci.* **2003**, *14*, 261–288. [CrossRef] [PubMed]
28. Hiemstra, C.A.; Liston, G.E.; Pielke, R.A.; Birkenheuer, D.L.; Albers, S.C. Comparing Local Analysis and Prediction System (LAPS) Assimilations with Independent Observations. *Weather Forecast.* **2006**, *21*, 1024–1040. [CrossRef]
29. Anagnostou, M.; Kalogiros, J.; Marzano, F.; Anagnostou, E.; Montopoli, M.; Picciotti, E. Performance evaluation of a new dual-polarization microphysical algorithm based on long-term X-band radar and disdrometer observations. *J. Hydrometeor.* **2013**, *14*, 560–576. [CrossRef]
30. Gebregiorgis, A.; Kirstetter, P.; Hong, Y.; Gourley, J.; Huffman, G.; Petersen, W.; Xue, X.; Schwaller, M. To What Extent Is the Day 1 GPM IMERG Satellite Precipitation Estimate Improved as Compared to TRMM TMPA-RT? *J. Geophys. Res. Atmos.* **2018**, *123*, 1694–1707. [CrossRef]
31. Chang, L.T.C.; Cheung, K.K.W.; McAneney, J. Case Study of Trmm Satellite Rainfall Estimation for Landfalling Tropical Cyclones: Issues and Challenges. *Trop. Cyclone Res. Rev.* **2013**, *2*, 109–123.
32. Lábó, E. Validation studies of precipitation estimates from different satellite sensors over hungary—Analysis of new satellite-derived rain rate products for hydrological purposes. *J. Hydrol.* **2012**, *468–469*, 173–187. [CrossRef]
33. Katsafados, P.; Mavromatidis, E.; Spyrou, C. *Numerical Weather Prediction and Data Assimilation*, 1st ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2020; ISBN 978-1-78630-141-3.
34. He, Z.; Xie, Y.; Li, W.; Li, D.; Han, G.; Liu, K.; Ma, J. Application of the Sequential Three-Dimensional Variational Method to Assimilating SST in a Global Ocean Model. *J. Atmos. Ocean. Technol.* **2008**, *25*, 1018–1033. [CrossRef]



35. Spyrou, C.; Varlas, G.; Pappa, A.; Mentzafou, A.; Katsafados, P.; Papadopoulos, A.; Anagnostou, M.N.; Kalogiros, J. Implementation of a Nowcasting Hydrometeorological System for Studying Flash Flood Events: The Case of Mandra, Greece. *Remote Sens.* **2020**, *12*, 2784. [CrossRef]
36. Ha, J.; Park, J. Advancing very short-term rainfall prediction with blended U-Net and partial differential approaches. *Front. Earth Sci.* **2024**, *11*, 1301523. [CrossRef]
37. Ha, J. Application of Advection-Diffusion Equation for Nonlinearly Evolving Precipitation Field. *Nonlinear Process. Geophys.* **2024**, *2024*, 1–17. [CrossRef]
38. Ryu, S.; Lyu, G.; Do, Y.; Lee, G. Improved rainfall nowcasting using Burgers' equation. *J. Hydrol.* **2020**, *581*, 124140. [CrossRef]
39. Das, P.; Posch, A.; Barber, N.; Hicks, M.; Duffy, K.; Vandal, T.; Singh, D.; Werkhoven, K.; Ganguly, A.R. Hybrid physics-AI outperforms numerical weather prediction for extreme precipitation nowcasting. *NPJ Clim Atmos Sci.* **2024**, *7*, 282. [CrossRef]
40. Efron, B.; Hastie, T. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*; Cambridge University Press: Cambridge, MA, USA, 2016.
41. Kalogiros, J.; Anagnostou, M.; Anagnostou, E.; Montopoli, M.; Picciotti, E.; Marzano, F. Evaluation of a new polarimetric algorithm for rain-path attenuation correction of X-band radar observations against disdrometer. *Geosci. Remote Sens. IEEE Trans.* **2014**, *52*, 1369–1380. [CrossRef]
42. Simpson, J.; Kummerow, C.; Tao, W.; Adler, R. On the Tropical Rainfall Measuring Mission (TRMM). *Meteorol. Atmos. Phys.* **1996**, *60*, 19–36. [CrossRef]
43. Kummerow, C.; Barnes, W.; Kozu, T.; Shiue, J.; Simpson, J. The Tropical Rainfall Measuring Mission (TRMM) Sensor Package. *J. Atmos. Ocean. Technol.* **1998**, *15*, 809–817. [CrossRef]
44. Kummerow, C.; Simpson, J.; Thiele, O.; Barnes, W.; Chang, A.; Stocker, E.; Adler, R.; Hou, A.; Kakar, R.; Wentz, F.; et al. The Status of the Tropical Rainfall Measuring Mission (TRMM) after Two Years in Orbit. *J. Appl. Meteorol.* **2000**, *39*, 1965–1982. [CrossRef]
45. Liu, Z.; Ostrenga, D.; Teng, W.; Kempler, S. Tropical Rainfall Measuring Mission (TRMM) Precipitation Data and Services for Research and Applications. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1317–1325. [CrossRef]
46. Morsy, M.; Scholten, T.; Michaelides, S.; Borg, E.; Sherief, Y.; Dietrich, P. Comparative Analysis of TMPA and IMERG Precipitation Datasets in the Arid Environment of El-Qaa Plain, Sinai. *Remote Sens.* **2021**, *13*, 588. [CrossRef]
47. Sharifi, E.; Steinacker, R.; Sagharian, B. Assessment of GPM-IMERG and Other Precipitation Products against Gauge Data under Different Topographic and Climatic Conditions in Iran: Preliminary Results. *Remote Sens.* **2016**, *8*, 135. [CrossRef]
48. Global Precipitation Measurement Mission (GPM). Mission Overview\_Precipitation Measurement Missions. Available online: <https://pmm.nasa.gov/GPM> (accessed on 1 April 2023).
49. Hou, A.; Kakar, R.; Neeck, S.; Azarbarzin, A.; Kummerow, C.; Kojima, M.; Oki, R.; Nakamura, K.; Iguchi, T. The global precipitation measurement mission. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 701–722. [CrossRef]
50. Retalis, A.; Katsanos, D.; Tymvios, F.; Michaelides, S. Comparison of GPM IMERG and TRMM 3B43 Products over Cyprus. *Remote Sens.* **2020**, *12*, 3212. [CrossRef]
51. Hossan, A.; Jones, W.L. Ku- and Ka-Band Ocean Surface Radar Backscatter Model Functions at Low-Incidence Angles Using Full-Swath GPM DPR Data. *Remote Sens.* **2021**, *13*, 1569. [CrossRef]
52. Huffman, G.; Bolvin, D.; Nelkin, E. Integrated Multi-satellite Retrievals for GPM (IMERG) Technical Documentation. 2015; NASA/GSFC Code. Available online: [https://gpm.nasa.gov/sites/default/files/document\\_files/IMERG\\_doc.pdf](https://gpm.nasa.gov/sites/default/files/document_files/IMERG_doc.pdf) (accessed on 13 July 2025).
53. Li, Y.; Zhang, K.; Bardossy, A.; Shen, X.; Cheng, Y. Evaluation and Error Decomposition of IMERG Product Based on Multiple Satellite Sensors. *Remote Sens.* **2023**, *15*, 1710. [CrossRef]
54. Sungmin, O.; Foelsche, U.; Kirchengast, G.; Fuchsberger, J.; Tan, J.; Petersen, W. Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 6559–6572. [CrossRef]
55. Rinollo, A.; Vulpiani, G.; Puca, S.; Pagliara, P.; Kaňák, J.; Lábó, E.; Okon, L.; Roulin, E.; Baguis, P.; Cattani, E.; et al. Definition and impact of a quality index for radar-based reference measurements in the H-SAF precipitation product validation. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2695–2705. [CrossRef]
56. Mugnai, A.; Casella, D.; Cattani, E.; Dietrich, S.; Laviola, S.; Levizzani, V.; Panegrossi, G.; Petracca, M.; Sa, P.; Di Paola, F.; et al. Precipitation products from the hydrology SAF. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 1959–1981. [CrossRef]
57. Puca, S.; Porcu, F.; Rinollo, A.; Vulpiani, G.; Baguis, P.; Balabanova, S.; Campione, E.; Erturk, A.; Gabellani, S.; Iwanski, R.; et al. The validation service of the hydrological SAF geostationary and polar satellite precipitation products. *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 871–889. [CrossRef]
58. Turk, J.; Hawkins, J.; Smith, E.; Marzano, F.; Mugnai, A.; Levizzani, V. Combining SSM/I, TRMM and infrared geostationary satellite data in a near-realtime fashion for rapid precipitation updates: Advantages and limitations. In Proceedings of the 2000 EUMETSAT Meteorological Satellite Data Users' Conference, Bologna, Italy, 29 May–2 June 2000; Volume 2, pp. 705–707.



59. Turk, F.J.; Miller, S.D. Toward improved characterization of remotely sensed precipitation regimes with MODIS/AMSR-E blended data techniques. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1059–1069. [CrossRef]
60. Torricella, F.; Levizzani, V.; Turk, J. Application of a Blended MW-IR Rainfall Algorithm to the Mediterranean. In *Measuring Precipitation From Space. Advances In Global Change Research*; Levizzani, V., Bauer, P., Turk, F.J., Eds.; Springer: Dordrecht, The Netherlands, 2007; Volume 28.
61. H SAF. Product Detail. Available online: <https://hsaf.meteoam.it/Products/Detail?prod=h03B> (accessed on 2 April 2023).
62. Diakakis, M.; Mavroulis, S.; Filis, C.; Lozios, S.; Vassilakis, E.; Naoum, G.; Soukis, K.; Konsolaki, A.; Kotsi, E.; Theodorakidou, D.; et al. Impacts of Medicanes on Geomorphology and Infrastructure in the Eastern Mediterranean, the Case of Medicane Ianos and the Ionian Islands in Western Greece. *Water* **2023**, *15*, 1026. [CrossRef]
63. Lagouvardos, K.; Karagiannidis, A.; Dafis, S.; Kalimeris, A.; Kotroni, V. Ianos—A Hurricane in the Mediterranean. *Bull. Am. Meteorol. Soc.* **2022**, *103*, E1621–E1636. [CrossRef]
64. Portmann, R.; González-Alemán, J.; Sprenger, M.; Wernli, H. Medicane Zorbas: Origin and impact of an uncertain potential vorticity streamer. *Weather Clim. Dyn. Discuss.* **2019**, 1–30. [CrossRef]
65. Kouroutzoglou, J.; Samos, I.; Flocas, H.A.; Hatzaki, M.; Lamiris, C.; Mamara, A.; Emmannouil, A. Analysis of the Transition of an Explosive Cyclone to a Mediterranean Tropical-like Cyclone. *Atmosphere* **2021**, *12*, 1438. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

# From Climate to Cloud: Advancing Fog Detection Through Satellite Imagery

Andrés Gabriel Arguedas Chaverri, Rogério Hartung Toppa and Kelly Cristina Tonello \*

Department of Environmental Sciences, Universidade Federal de São Carlos, Campus Sorocaba, Sorocaba 18052-780, SP, Brazil; achaverri@estudante.ufscar.br (A.G.A.C.); toppa@ufscar.br (R.H.T.)

\* Correspondence: kellytonello@ufscar.br

**Abstract:** The broad spatiotemporal coverage provided by satellite remote sensing is fundamental for monitoring fog events, a phenomenon that impacts transportation, agriculture, and ecosystem functioning. Despite advances in remote sensing technology, significant knowledge gaps remain regarding the application of these techniques to fog detection, especially over terrestrial ecosystems. This scoping review synthesizes the trends in methods used for fog detection by analyzing 38 papers retrieved from Scopus and Web of Science. Only studies that utilized satellite imagery to analyze the spatiotemporal dynamics of fog were included. Articles that employed non-satellite methodologies or focused on processes other than the detection, formation, or identification of fog events were excluded. In addition to a term co-occurrence analysis of abstracts using VOSviewer, this study examines key parameters of the detection methods—including sensor type, spectral bands, temporal resolution, and algorithmic approaches (e.g., threshold methods and deep learning techniques)—to evaluate their evolution and current limitations. Our results reveal that while approximately 53% of studies rely on geostationary satellite data (95% CI: 36.7–68.5%), favored for their high temporal resolution, the remaining 47% employ polar-orbiting sensors (95% CI: 31.5–63.2%) that offer superior spatial resolution. Notably, most research has concentrated on maritime fog detection, with few studies extending these techniques to complex terrestrial environments. The review highlights critical gaps in current approaches and proposes an integrated framework that combines traditional brightness temperature difference methods with emerging machine learning techniques, which could advance fog detection in diverse settings.

**Keywords:** satellite images; term co-occurrence; PRISMA; fog detection; VOSviewer

## 1. Introduction

Fog is a meteorological phenomenon defined as a collection of water droplets suspended near the Earth's surface, which reduces horizontal visibility to less than 1 km [1]. For decades, fog has captured the attention of the scientific community due to its direct and indirect impacts on human activities [2]. Beyond its visual and atmospheric characteristics, fog poses significant challenges to transportation systems—aviation, marine, and land—resulting in considerable economic and social consequences [3,4]. For instance, during the winter of 1998–1999, persistent fog in eastern India and northeastern Pakistan affected over 2200 commercial flights, including 925 cancellations, disrupted rail services, and caused numerous vehicular accidents, illustrating the severity of fog-related impacts on human mobility and safety [5]. Additionally, fog has been linked to psychological

effects, the spread of epidemics, increased atmospheric pollution, and public health crises, as historically documented [6] in the London Smog in December 1952 [7].

While fog is widely recognized for its negative consequences—such as reduced visibility and increased accident risks—it also plays a vital ecological role by contributing to regional water cycles through fog precipitation. Also known as horizontal or occult precipitation, this process can substantially contribute to water budgets, particularly in arid and semi-arid regions [8,9]. Model-based studies indicate that cloud water interception (CWI) can account for less than 5% of total precipitation in humid regions and exceed 75% in low-rainfall areas [2]. In these environments, fog functions as a subtle yet persistent natural reservoir, sustaining ecosystems and supporting local communities where conventional rainfall is limited [10,11].

Despite the relevance of ground-based fog observations for local monitoring, their spatial and temporal limitations constrain broader assessments. In contrast, satellite remote sensing offers a synoptic and continuous view of fog dynamics, much like an aerial survey that captures the temporal evolution of fog across extensive regions [12,13]. Since the 1970s, when satellite imagery first began to be used for fog detection [14], technological advances have introduced both polar-orbiting and geostationary sensors capable of providing high-temporal-resolution data across multiple spectral channels [15,16]. This capability can be likened to creating a time-lapse sequence of fog development, movement, and dissipation—essential for both operational forecasting and the analysis of climatological patterns.

Nevertheless, despite technological progress, significant knowledge gaps persist regarding the spatiotemporal dynamics of fog, limiting our ability to understand its life cycle and accurately map its occurrence across different regions [17–20]. Filling these gaps is particularly critical for informing public policies and operational decisions in sectors such as transportation, agriculture, and industry [13]. Although advances in fog detection have been notable in Europe and North America, other parts of the world remain underrepresented in the scientific literature [12]. These disparities underscore the need to adapt and refine remote detection methodologies to better accommodate diverse climatic and geographic contexts. Recent efforts to simulate fog in urban areas have further revealed spatial mismatches and resolution limitations between satellite observations and numerical models, underscoring the need for methodological integration [21]. This challenge can be compared to refining a universal translation tool that must account for linguistic nuances—requiring both precision and adaptability.

Historically, the development of satellite-based fog detection began in the 1970s, when visible and infrared imagery were first employed to assess the radiative properties of clouds and fog, laying the physical groundwork for remote detection [22]. Early techniques using AVHRR (Advanced Very High Resolution Radiometer) imagery leveraged brightness temperature differences between thermal channels to detect nighttime fog [23,24]. These methods were progressively refined and adapted to different geographic contexts over the following decades [25,26], relying primarily on threshold-based analyses and manual interpretation.

Further advancements were achieved with the use of multispectral infrared imagery from GOES satellites, which enhanced nighttime fog detection [27], along with the development of microphysical retrieval techniques validated by in situ balloon measurements [28]. Subsequent improvements in satellite instrumentation, such as the GOES I-M series, expanded the capabilities for monitoring boundary-layer cloud properties [29]. Comprehensive reviews published in the early 2000s synthesized these advancements, highlighting key physical principles and methods for fog detection from space [4,30,31]. Despite the limitations of early sensors and algorithms, these pioneering contributions remain foundational to the evolution of current satellite-based fog detection methodologies.

Given the ecological, economic, and social relevance of fog, combined with the historical development and ongoing methodological challenges discussed above, it becomes essential to understand how scientific research in this field has evolved. In this context, the present study aims to map, through a bibliometric analysis, the main trends, methods, and gaps related to satellite-based fog detection, providing a critical overview of the current state of knowledge and outlining pathways for the development of more robust and applicable methodologies across diverse geographic and climatic contexts.

This study is structured around two primary research questions:

What are the specific methods employed in satellite-based fog detection, and what are their advantages and limitations?

Why is it urgent to refine these methods for terrestrial applications, and how can these improvements advance our understanding of fog dynamics and their implications for environmental management and societal needs?

By addressing these questions, this review aims to bridge knowledge gaps and propose a roadmap for future research. Ultimately, improving fog detection methodologies will contribute to more accurate forecasting and support informed decision-making in sectors that are highly sensitive to fog-related risks and opportunities, such as governmental planning, industrial operations, transportation logistics, and agricultural management.

## 2. Material and Methods

### *Literature Review and Data Extraction*

A scoping review was conducted following the PRISMA guidelines [32] to identify papers that apply satellite remote sensing for fog detection. Searches were performed in the Scopus and Web of Science databases, which are among the most comprehensive sources for publication metadata and impact indicators due to their extensive coverage of diverse topics and multidisciplinary content [33]. The search was conducted on 8 February 2024, ensuring the inclusion of the most recent publications relevant to the topic. The keywords “fog frequency”, “remote sensing”, and their variants were searched to analyze papers that were associated with fog event frequency detected through remote sensing (Table 1). The search was unrestricted by publication year and confined to scientific papers, thereby excluding review articles, books, book chapters, and publications in languages other than English. To manage the literature, the Rayyan (<https://www.rayyan.ai/> (accessed on 18 November 2024)) interface was employed, which facilitated duplicate detection and efficient tracking of included and excluded records during the identification stage. The resulting papers were evaluated in the first part of the screening stage by title and abstract, according to the established inclusion and exclusion criteria listed below. In the initial screening stage, abstracts and titles were reviewed based on predetermined inclusion and exclusion criteria. Specifically, only studies that utilized satellite imagery to analyze the spatiotemporal dynamics of fog were included. Articles that employed non-satellite methodologies or focused on processes other than the detection, formation, or identification of fog events were excluded. Furthermore, our search was ultimately limited to peer-reviewed original research and review papers. Any duplicate entries from the search results were removed prior to the detailed review process.

The studies were categorized based on fog detection techniques, sensor type, and research domain (terrestrial vs. maritime). Once the papers to be included in this review were selected, VOSviewer software version 1.6.19 was used to create a term co-occurrence map based on text data extracted from the title and abstract of each paper. We applied binary counting to capture only the presence or absence of terms in each document, rather than their frequency. Terms with fewer than two co-occurrences were excluded. Common

monosyllabic words and irrelevant terms were excluded from the final selection. Clusters were automatically generated by the software's default algorithm, which is based on cosine similarity. No manual adjustments were made to the allocation of terms into clusters.

**Table 1.** Query parameters used in the Scopus and Web of Science databases.

| Database        | Keywords                                                                                                                                                                                                                                                                                                                                                           |
|-----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Scopus:         | TITLE-ABS-KEY ("fog water" OR "fog interception" OR "fog characteristic" OR "fog intensity" OR "fog event" OR "fog occurrence" OR "fog formation" OR "fog frequency") AND TITLE-ABS-KEY ("remote sensing" OR "satellite imagery" OR "remote-sensing" OR "hyperspectral" OR "GIS" OR "image enhancement" OR "remote sensing images") AND (LIMIT-TO (DOCTYPE, "ar")) |
| Web of Science: | (TS = ("fog water" OR "fog interception" OR "fog characteristic" OR "fog intensity" OR "fog event" OR "fog occurrence" OR "fog formation" OR "fog frequency")) AND (TS = ("remote sensing" OR "satellite imagery" OR "remote-sensing" OR "hyperspectral" OR "GIS" OR "image enhancement" OR "remote sensing images")) AND (DT = ("ARTICLE"))                       |

In the network visualization map, the higher the occurrence of a term, the larger its label and circle, and the distance between words indicates their relationship [34]. The distribution was analyzed over the years of publication and the country of the lead author of each paper, along with the average year of publication of the documents in which a particular word appears, to understand the trend of the studied terms.

As this study is a scoping review, a formal risk of bias assessment was not conducted. The focus was on mapping research trends rather than critically appraising study quality. In addition, no effect measures were calculated or synthesized.

### 3. Results and Discussion

#### 3.1. Analysis of Publications

A total of 154 references were retrieved in the information search, with 101 from the Scopus database and 53 from Web of Science. Forty-eight papers were eliminated due to duplication, 68 due to exclusion criteria, and four because the full text was not available (Figure 1). Two of these were published in Chinese without available English translations, and two others were inaccessible despite attempts to retrieve them via institutional subscriptions and author contact. We acknowledge that the exclusion of non-English publications—particularly in Chinese, a dominant language in fog-related research—may introduce a potential language and regional bias into our findings. Nonetheless, we prioritized methodological consistency and transparency in the selection process, ensuring that all included studies met the same accessibility and evaluability criteria.

As a result of identifying studies through the databases, 34 studies were included in the review. Additionally, four records were identified based on citation searches, as they were references continuously found while reading the database texts. Papers that appeared in more than four different studies were considered as part of the citation search.

The countries with the most publications were China, Germany, India, and Canada, with 12, eight, five, and four studies, respectively (Figure 2). It was noted that both Chinese and Canadian researchers focused their studies on fog events in the oceans, the Yellow Sea and Bohai Sea in China [14–16,35–41] and the Atlantic Ocean off the coast of Nova Scotia and Newfoundland and Labrador in Canada [12,42–44].

In India, studies focused on the detection of continental fog within the Indo-Gangetic Plain, which includes parts of Pakistan, Bangladesh, and northern India [45–49]. In countries with fewer publications, the research objective was the impact of fog on vegetation in the Atacama Desert in Chile [19,50], port activities in Casablanca, Morocco [17,51], and



agricultural production in California, United States [13,52]. It was found that the areas studied by German authors were located outside Germany, primarily researching the spatiotemporal distribution patterns of ecosystems with high ecological importance. Some of the regions studied include arid zones on the west coast of southern Africa [20,53] and fog forests and lowland tropical rainforests in South America and Taiwan [18,54–57].

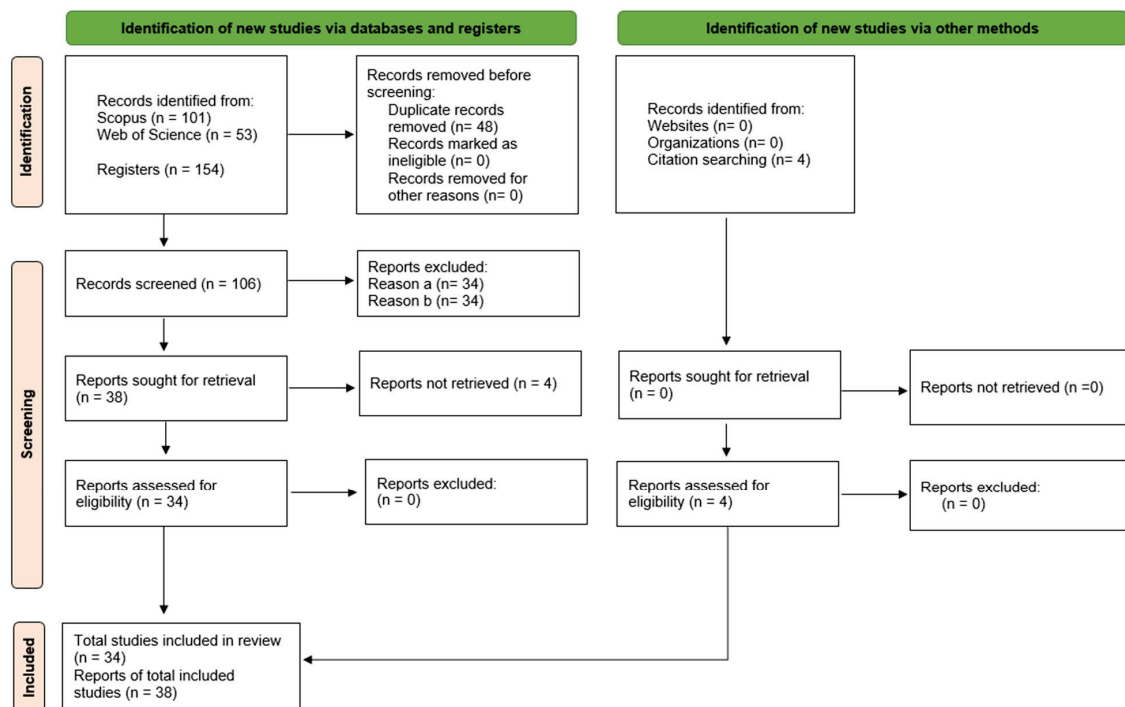


Figure 1. Flowchart of paper selection for the review. Source: Adapted from [32].

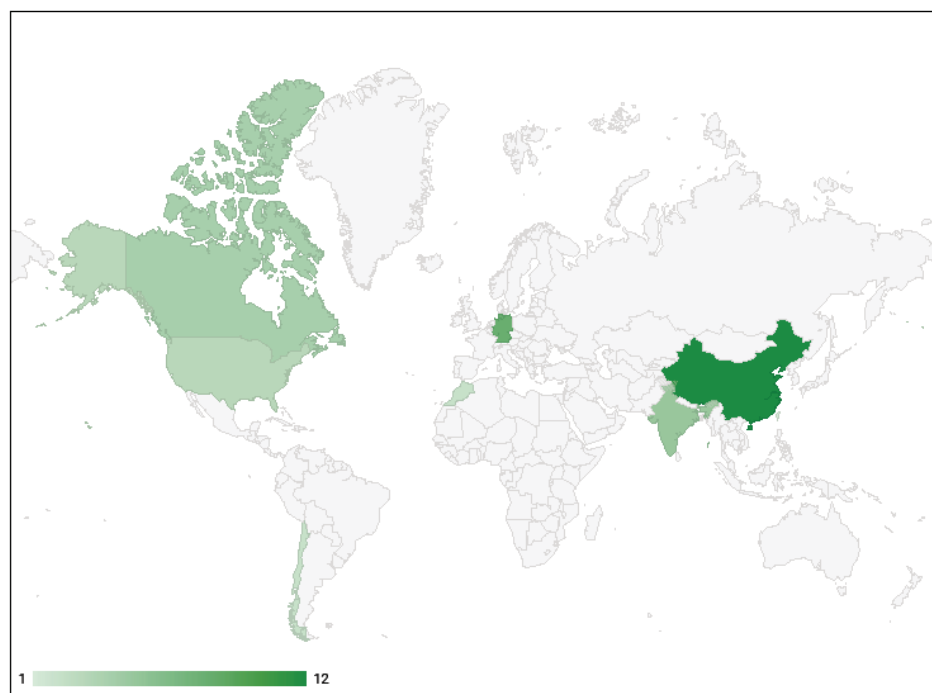


Figure 2. Distribution of the number of publications by the country of the lead author of each paper.

While some regions identified in this review—such as China, Germany, India, and Canada—benefit from advanced research infrastructure and extensive satellite monitoring programs, this concentration of studies in a limited number of geographic areas introduces a significant geographic bias. The majority of reviewed studies focused on specific zones, including the Yellow Sea and Bohai Sea (China), the Atlantic coast of eastern Canada, the Indo-Gangetic Plain (South Asia), and arid or ecologically significant regions studied by German institutions across Africa, South America, and Taiwan. This uneven distribution limits the generalizability of the findings, as methodologies calibrated in well-studied regions may not translate effectively to areas with different climatic, topographic, or ecological conditions.

Our analysis acknowledges a critical limitation in generalizing the findings to underrepresented regions, notably Africa and South America. As highlighted by other studies, atmospheric science research in these areas faces significant barriers—including limited infrastructure, funding constraints, and reduced participation in international collaborations—resulting in a shortage of region-specific studies that hampers the development of comprehensive models and limits our understanding of fog dynamics across diverse ecological and climatic contexts [58]. Consequently, the insights presented in this review—primarily drawn from regions with higher research output—may not fully capture the variability and characteristics of fog in data-scarce areas. Addressing this gap requires coordinated efforts to strengthen local research capacity, promote equitable collaboration, and prioritize studies in these critically important but understudied regions. As a result, current methodologies, which are often calibrated and validated using data from well-studied regions, may perform suboptimally when applied elsewhere. This underscores the need for future research to expand coverage to data-scarce regions, thereby enhancing the global applicability of satellite-based fog detection methods.

Regarding the distribution of publications by year, there was a trend of increasing numbers of publications, but it was not statistically significant over time (Figure 3). The increase seems to be driven by the effects of global warming on fog formation processes, such as the increase in Arctic fog resulting from sea ice melting [59]. Other observed effects include the spatial evolution of the fog life cycle over cities [17,49,60] and the importance of understanding variations in surface heat fluxes in plant ecosystems [19,52,54].

### 3.2. Term Co-Occurrence Analysis

In the network visualization map, built based on text data extracted from the titles and abstracts, seven thematic clusters were identified, grouping the most related terms within the same color (Figure 4). Terms with the highest co-occurrence, such as “fog”, “satellite”, “data”, “satellite imagery”, “fog event”, and “remote sensing” are located at the center of the map, as they represent the main topics of this review.

Although some terms or words appeared less frequently compared to the main ones, their occurrence was enough to highlight them as secondary themes in their respective groups. In the yellow cluster, for example, terms such as “horizontal visibility”, “air temperature”, and “modeling study” gather studies that use visible satellite imagery and climate variables to model advection cooling fog formation processes. The advection fog occurs when warm, humid air is cooled to its dew point by a cold underlying sea surface until it condenses [35,36,61]. However, some studies used the difference between the temperature at the top of the fog and low stratus (FLS) and the surface temperature, considering that positive values imply the existence of a temperature inversion in the marine boundary layer, a condition favorable to fog occurrence [16,59].

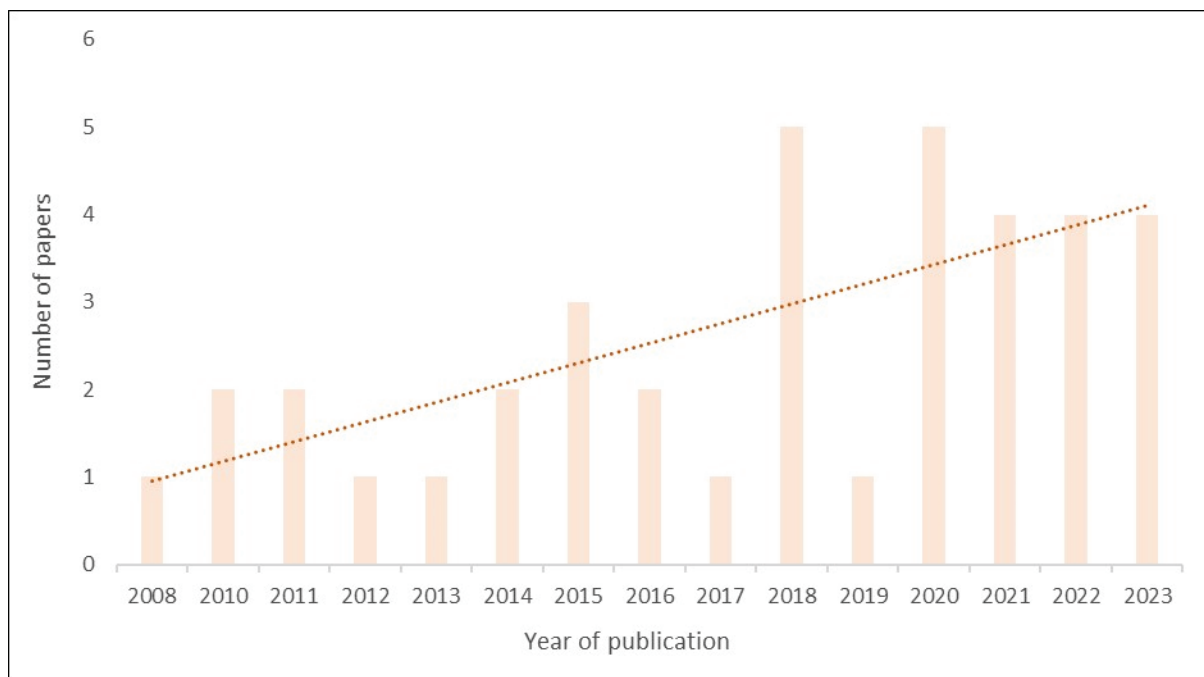


Figure 3. Distribution of studies by year of publication.

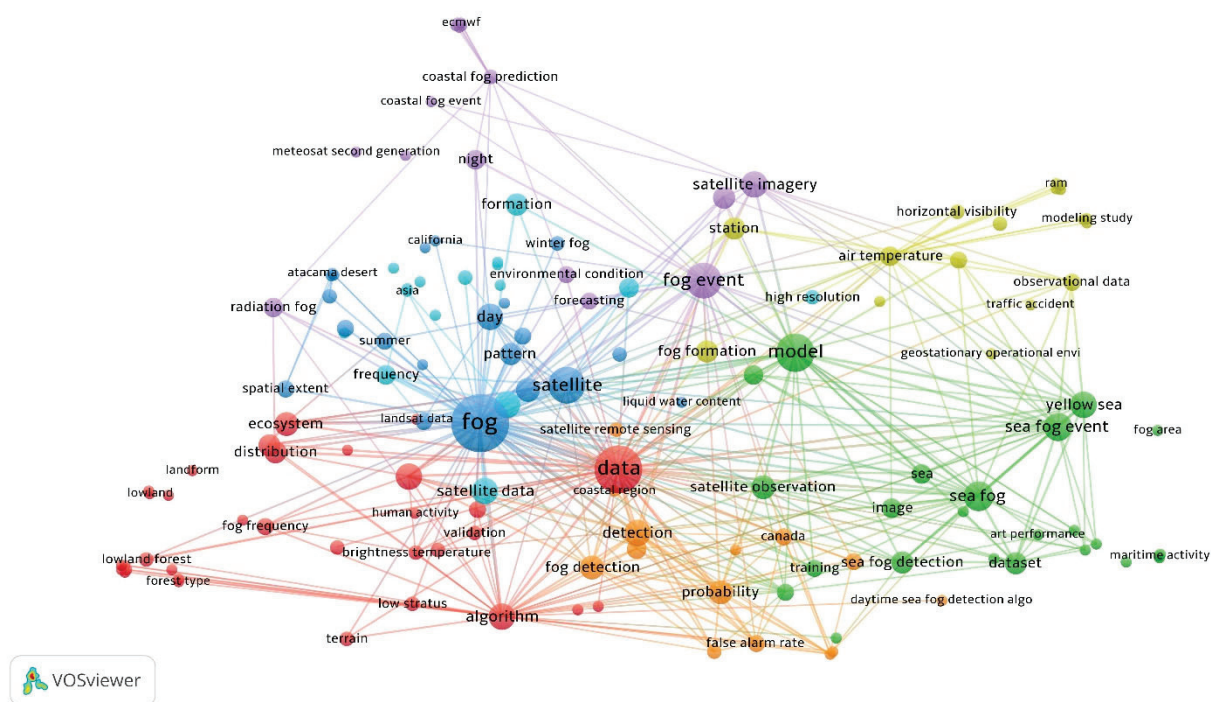


Figure 4. Network visualization map based on text data.

Unlike the first three papers in the yellow cluster [35,36,61], which only use visible spectrum bands, the latter suggest a comprehensive dynamic threshold algorithm with infrared bands, combining cloud brightness temperature (BT) with climatological monthly mean data and sea surface BT values [16]. The green cluster shows secondary themes such as “satellite observation”, “sea fog”, and “Yellow Sea”, referring to studies on sea fog. Although connected with the yellow cluster terms, the methodologies in the green cluster for detecting maritime fog are based on the threshold method (TM) between thermal

infrared and mid-infrared bands for fog detection [12,44,62]. Smaller droplets in fog are associated with lower emissivity in the 3.7  $\mu\text{m}$  band than in the 10.8  $\mu\text{m}$  band, while the emissivity for larger droplets in high- and mid-level clouds is approximately the same. This characteristic results in a significant contrast in BT due to the emissivity difference [48].

The term “model” from the green cluster is close to secondary terms such as “satellite imagery”, “forecasting”, and “environmental condition” from the purple cluster. This can be explained by the use of climate models, such as the Weather Research and Forecasting Model (WRF) [42] or the Mesoscale Nonhydrostatic Model (Meso NH) [17,51], along with reanalysis data, meteorological station information, and visible satellite imagery observations to study fog events and predict their formation in coastal cities.

Although the term “radiation fog” belongs to the purple cluster, related to modeling for a better understanding of the different physical processes of advection and radiation in coastal fog [51], it is more closely related to the blue and red clusters, which have other research objectives in continental areas. Radiation fog explains how certain factors, including local radiative cooling, high relative humidity, and calm winds, allow air near the ground to cool to its dew point, favoring fog formation associated with vegetation cover [49]. In this context, the red cluster moves away from clusters that predominantly focus on maritime fog formation studies (mainly formed by advection processes) and their forecasting, focusing instead on methodologies to study the spatial extent of fog over continental areas.

In the red cluster, terms such as “ecosystem”, “distribution”, “fog frequency”, and “algorithm” were identified as secondary terms that group papers focused on when developing methodologies to study the distribution and frequency of fog over time. This approach generates frequency maps to identify potential locations of cloud forests [18,56,57], as understanding the spatial occurrence of FLS is essential for detecting potential clusters of this type of forest [54]. In this group of studies, the incorporation of relief and topography data into detection algorithms was observed. These factors interact with climate variables in fog formation processes [56,63], as well as with the time of year and seasonal variations in which the studies are conducted [52].

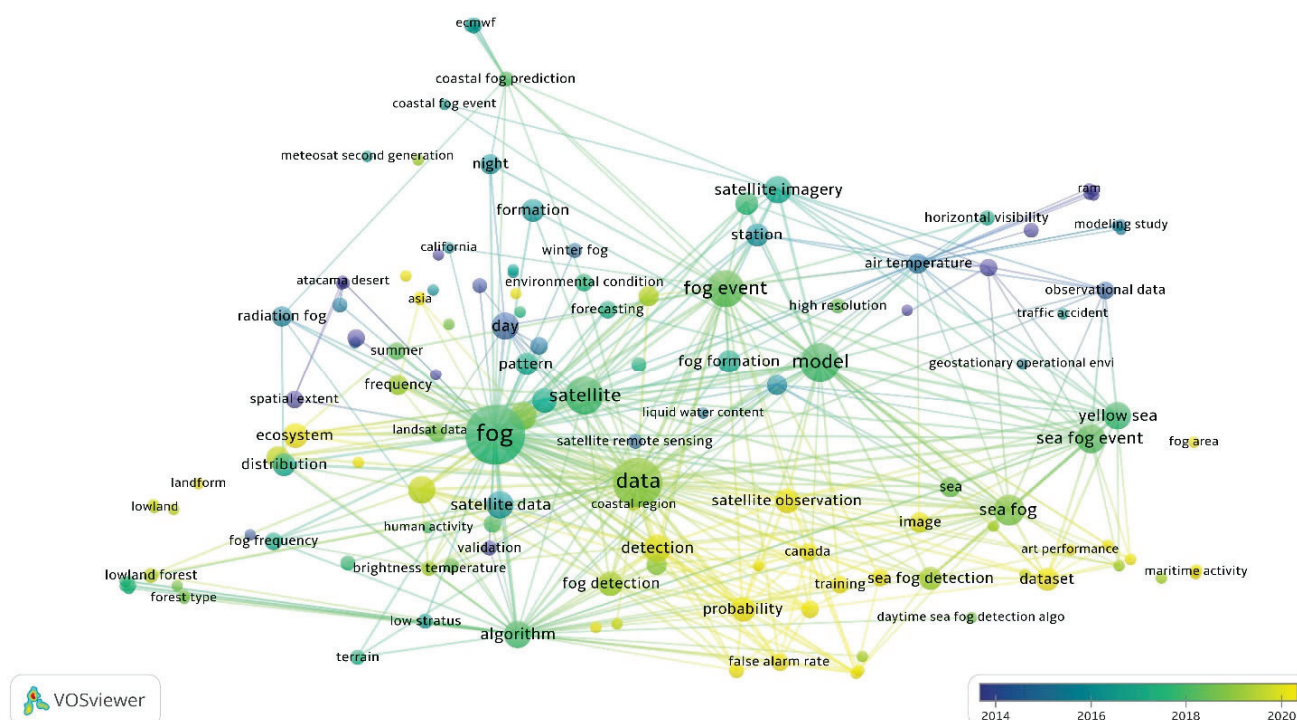
The orange cluster presents the terms “fog detection”, “probability”, and “false alarm rate”, which refer to a group of studies with more complex methodologies, using a deep learning approach based on artificial neural network algorithms. These methodologies employ algorithms such as TMs between thermal bands, climate reanalysis data, radar, meteorological stations, satellite and ground observations, and fog event databases [38,40,41,62]. This dataset is used to train convolutional neural networks to improve fog event detection and forecasting.

Although most of the literature in this cluster focuses on applying these technologies to maritime fog detection, showing an overlap with terms from the green cluster such as “accuracy” and “training”, other investigations used deep learning algorithms to enhance fog forecasting and support traffic planning in northern China [39]. This connection helps explain the proximity of terms from the red cluster, such as “algorithm” and “human activity”.

It was observed that studies employing machine learning techniques for fog detection have shown a notable increase in recent years. Of the total ML-focused studies identified in this review, 75% were published between 2020 and 2023. This pattern aligns with the overlay visualization (Figure 5), where terms related to deep learning and training appear predominantly in more recent publications. However, recent trends in publications were also observed focusing on ecosystem research, such as studying the impact of climate change on the spatiotemporal dynamics of fog forests using learning methods [64]. Other



studies found that fog can contribute to forest resilience to climate change in the Amazon, creating refuges for vulnerable organisms in areas with high fog frequency during the dry season [55].



**Figure 5.** Network visualization map with the overlay of paper publication years.

### 3.3. Satellite Data in Fog Detection

The type of satellite imagery used in detection methodologies depended on the characteristics of the study areas in each paper. It is known that the temporal resolution of polar-orbiting satellites is lower than that of geostationary satellites, and polar-orbiting satellites have better spatial resolution [14]. Thus, the choice of satellite responded to the time and space information needs of the fog events being studied.

Of the total studies selected for this review, 47.34% (18 out of 38) worked with observations from sensors aboard polar-orbiting satellites (95% CI: 31.5–63.2%), while 52.63% (20 out of 38) were associated with observations using geostationary satellites (95% CI: 36.7–68.5%) (Table S1 in Supplementary Materials). Due to the limited temporal resolution of sensors mounted on polar-orbiting systems compared to geostationary systems, the latter became the more suitable option for providing data at short measurement intervals [65].

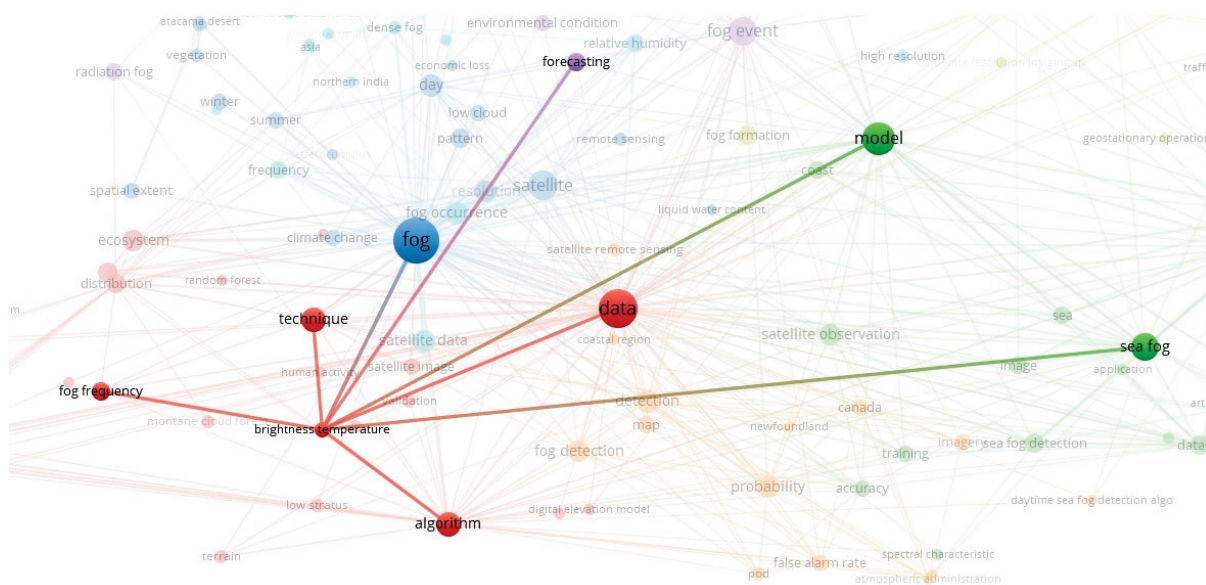
The use of geostationary system data was found to be more common in studies with shorter temporal series, where fog events were researched over intervals of days, months, or, at most, four years [17,35,36,39,43,44,46,50,51,60]. Images from the Geostationary Operational Environmental Satellite (GOES) 15 were used to study the effect of coastal fog on agricultural systems in California. The 30 min temporal resolution of GOES-15 was crucial for studying changes in plant's energy and water balance, as well as changes in the spatial extent of fog [13]. For maritime fog research, observations from the Meteorological Imager (MI) aboard the Communication Ocean and Meteorological Satellite (COMS) were chosen. The advantages of short-term temporal resolution, a wide observation area, and many spectral channels from visible to long-wave infrared were the reasons for selecting this sensor's data [62]. In contrast, the temporal series in studies using data from polar-orbiting satellites were longer, ranging from five to more than 15 years [38,47–49,52,54,55,57,59,63].



Polar-orbiting satellites cover many areas around the world with better spatial resolution due to their lower operating altitude compared to geostationary satellites. It was observed that the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor was the most recurrent in the papers reviewed that used data from this type of orbiting satellite. Some research indicated that MODIS data provided new opportunities for monitoring fog dynamics [30], due to their finer spatial resolution and consistent overpass timing, which minimizes the influence of solar zenith angles—an issue observed with geostationary sensors such as SEVIRI during twilight periods [15]. Similarly, MODIS images were selected in other studies due to their spatial resolution of 1 km to 250 m, depending on the distribution of fog forests in Taiwan, which is highly altitude dependent. In mountainous areas, coarse resolution implies a large altitudinal range covered by each pixel [56,57].

### 3.4. Methodologies for Fog Detection

This section analyzes the different methodologies described in the reviewed papers for identifying fog using satellite images. TMs, neural networks, texture analysis, and cluster analysis are commonly used in cloud feature studies. Among them, the TM is the most widely used for detecting maritime fog via satellite [15]. However, despite being widely used for maritime detections, the TM was found in many studies in this review, regardless of the type of coverage. The term “brightness temperature”, a variable used in tests within the TM, is linked to the different thematic clusters in this review (Figure 6), highlighting the versatility and applicability of this method.



**Figure 6.** Thematic connections of the term brightness temperature.

It was found that the TM, based on the brightness temperature difference (BTD) test between two infrared window channels, was applied to both geostationary and polar-orbiting satellite images. Although it is the same technique, it is important to understand that the threshold values can vary from place to place [45] and, therefore, between the bands of each sensor. Data from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) sensor analyzed the difference between the 3.9  $\mu\text{m}$  (IR3.9) and 10.8  $\mu\text{m}$  (IR10.8) channels, based on the fact that low-level water clouds observed in IR3.9 have lower emissivity than the same clouds observed in IR10.8 [17,51]. Other studies that opted for images from the same sensor conducted the analysis using the 12.0  $\mu\text{m}$  and 8.7  $\mu\text{m}$  channels [20]. Data from GOES-16, captured in bands 14 (11.2  $\mu\text{m}$ ) and 7 (3.9  $\mu\text{m}$ ), corresponding to the thermal

infrared and mid-infrared spectra, were used in a similar remote sensing technique to detect nocturnal sea fog [12]. Meanwhile, [44] defined fog boundaries based on the BTD between the 3.9  $\mu\text{m}$  and 10.6  $\mu\text{m}$  bands [44]. To study spatiotemporal dynamics of fog using BT data from the MODIS sensor, the most frequently used channels were 3.9  $\mu\text{m}$  and 10.8  $\mu\text{m}$  [54] and 11  $\mu\text{m}$  and 3.6  $\mu\text{m}$  [19,56], as well as bands 22 (3.929–3.989  $\mu\text{m}$ ) and 31 (10.780–11.280  $\mu\text{m}$ ) [48,49].

Some studies indicated that solar radiation interferes with the BT of the bands, obscuring the signal associated with fog reflective characteristics [41,44]. During daytime, the 3.9  $\mu\text{m}$  band can be contaminated by reflected solar radiation, which reduces the contrast needed for fog identification [30]. Moreover, atmospheric water vapor strongly influences the  $\Delta\text{BT}$  values: as vapor content increases, the  $\Delta\text{BT}_{10.8-3.9}$  becomes more negative, potentially leading to false positives or negatives [18]. Thus, the separation of FLS from other cloud types becomes difficult during daytime, which is why this technique is primarily used at night. Another reason for choosing nighttime data is the fog formation processes, such as in [54] the Amazon plain, where fog occurs at night and early in the morning [54]. In the Indo-Gangetic Plain, INSAT-3D data were used to distinguish fog from clouds and snow via two threshold methods: BTD between TIR1 (10.8  $\mu\text{m}$ ) and MIR (3.9  $\mu\text{m}$ ) for nighttime, and a combination of visible and TIR1 bands for daytime detection [54]. Although the study did not report formal accuracy metrics, it qualitatively compared satellite-derived fog with visibility data from METAR stations, showing consistent temporal patterns, especially at night. This highlights the need for future studies to incorporate quantitative validation using reanalysis or LIDAR data.

For daytime fog detection in the Arctic, measurements from MODIS band 31 (10.7  $\mu\text{m}$ ) and surface temperature provided by standard MODIS data products were used [59]. Based on this information, it is possible to detect an FLS event if the temperature difference between the cloud top and the surface exceeds the established threshold for a given region. The threshold test using visible and near-infrared data was another technique used in TM. Studies in the Yellow Sea in China used a combination of MODIS band 3 (479 nm), band 6 (1.652  $\mu\text{m}$ ), and band 7 (2.155  $\mu\text{m}$ ) to distinguish between cloud types, as liquid water absorbs more red and shortwave infrared radiation [36]. Other studies opted for data from MODIS channels 1 (620–670 nm), 2 (841–876 nm), 18 (931–941 nm), and 19 (915–965 nm) to differentiate between fog, high clouds, and the sea surface reflected in water vapor [15]. Although TM is widely used for being stable and simple, some authors believe that simple threshold combinations are insufficient to monitor maritime fog in complex scenarios [38]. As a result, other studies have used multiple threshold combinations with more statistical analysis and a combination of other data types and techniques. In northern China, three different methodologies were applied using data from the Advanced Himawari Imager (AHI) aboard the Himawari-8 satellite, considering that fog has consistent spatial and temporal characteristics within adjacent 30 min intervals [39]. The texture and spectral difference, BTD, and visual interpretation methodologies were employed to recover daytime, nighttime, and twilight fog, respectively.

One promising approach is a hybrid pipeline that integrates TM with convolutional neural networks (CNNs). For example, BTD values derived from thermal bands can serve as pre-classified inputs or feature maps to train CNNs capable of distinguishing fog from other low-level clouds [62]. This combined architecture leverages the physical interpretability of TM and the pattern recognition power of ML. A dual-branch neural network applied to geostationary satellite data has been shown to improve fog detection accuracy when compared to traditional TM, as demonstrated in a recent study using GOCI imagery [41].

It was observed that in all the studies included in this review, the methodologies validated fog detection using reanalysis data such as ERA-5, data from active satellites like LIDAR, meteorological stations, and ground-based observations. It was clear that the purpose of the analyzed research was to define combinations of fog detection techniques according to the processes and areas studied. Despite recent progress, fog detection using satellite imagery still faces key limitations, particularly in distinguishing fog from low stratus and in detecting nocturnal fog due to sensor spectral constraints. For example, while high-resolution sensors such as Sentinel-2 offer detailed spatial information and advanced cloud masking algorithms, their lack of thermal infrared bands limits their applicability for detecting FLS during nighttime periods [54].

#### 4. Conclusions

A total of 38 studies were considered in this literature review. While our search strategy was methodologically rigorous and broad in scope, the exclusion of non-English studies and papers with inaccessible full texts may have constrained the geographical and linguistic diversity of the included literature. This is especially relevant given the high volume of fog detection research conducted in China. Future reviews would benefit from the integration of multilingual search strategies and translation tools to ensure more comprehensive global representation. The country with the most publications was China, followed by Germany and India. The methodologies found varied according to the type of coverage, research objective, type of satellite, sensor characteristics, and detection techniques. Seven thematic clusters were identified, grouping publications based on the fog formation process studied and the research objective.

The primary objective of the studies was fog forecasting in the sea and large urban centers, motivated by the impact on maritime, air, and land transportation activities. Fewer papers focused on ecosystems where the spatiotemporal dynamics of fog have high ecological importance, such as deserts, fog forests, or agricultural systems. Around 53% of the papers used data from sensors aboard geostationary satellites. The high temporal resolution of these satellites allowed the research to have shorter time series (from days to up to four years). A total of 47.34% of the studies worked with observations from sensors aboard polar-orbiting satellites, with the MODIS sensor being the most frequently used in this category. These studies, unlike the ones mentioned earlier, investigated fog distribution overtime series ranging from five to more than 15 years. The fog detection techniques found did not differentiate between image types, satellite, coverage studied, or time series. The BTM technique was the most frequently used among TM and in datasets from more complex methodologies, such as those based on deep learning. It was found that more recent papers employed machine learning and deep learning methodologies, as well as CNNs. This pattern aligns with the overlay visualization, where terms related to deep learning and training appear predominantly in more recent publications. Further studies could benefit from combining data from complementary sensors—including geostationary platforms like Himawari-8, hyperspectral imagers, and active sensors such as LIDAR—to improve temporal coverage and classification accuracy in complex terrestrial environments. Although most existing works have focused on maritime fog forecasting, there are also studies addressing the effects of climate change on vulnerable ecosystems. Advancing this field requires the development of integrated methodologies that combine traditional threshold-based techniques—such as brightness temperature difference—with machine learning approaches, including convolutional neural networks. This combination can leverage the strengths of both strategies, improving classification accuracy by coupling physically grounded indicators with data-driven pattern recognition. Identifying these

methodologies can help generate new research that provides information on fog patterns in vulnerable ecosystems, one of the most recent and least explored themes found in this review.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/cli13060110/s1>: Table S1. Sensors used in the studies included in the literature review.

**Author Contributions:** Methodology, A.G.A.C. and K.C.T.; formal analysis, A.G.A.C., K.C.T. and R.H.T.; investigation, A.G.A.C.; writing—original draft, A.G.A.C.; writing—review and editing, A.G.A.C., K.C.T. and R.H.T.; supervision, K.C.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was financial supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)—CODE 001, Brazilian National Council for Scientific and Technological Development (CNPq)—grant 312562/2021-7 and the São Paulo Research Foundation (FAPESP)—grant 2021/11697-9.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- World Meteorological Organization. *International Meteorological Vocabulary*; World Meteorological Organization: Geneva, Switzerland, 1992; Volume 182.
- Bruijnzeel, L.A.; Mulligan, M.; Scatena, F.N. Hydrometeorology of Tropical Montane Cloud Forests: Emerging Patterns. *Hydrol. Process.* **2011**, *25*, 465–498. [CrossRef]
- Gultepe, I.; Tardif, R.; Michaelides, S.C.; Cermak, J.; Bott, A.; Bendix, J.; Müller, M.D.; Pagowski, M.; Hansen, B.; Ellrod, G.; et al. Fog Research: A Review of Past Achievements and Future Perspectives. *Pure Appl. Geophys.* **2007**, *164*, 1121–1159. [CrossRef]
- Gultepe, I.; Sharman, R.; Williams, P.D.; Zhou, B.; Ellrod, G.; Minnis, P.; Trier, S.; Griffin, S.; Yum, S.S.; Gharabaghi, B.; et al. A Review of High Impact Weather for Aviation Meteorology. *Pure Appl. Geophys.* **2019**, *176*, 1869–1921. [CrossRef]
- Hameed, S.; Mirza, M.I.; Ghauri, B.M.; Siddiqui, Z.R.; Javed, R.; Khan, A.R.; Rattigan, O.V.; Qureshi, S.; Husain, L. On the Widespread Winter Fog in Northeastern Pakistan and India. *Geophys. Res. Lett.* **2000**, *27*, 1891–1894. [CrossRef]
- Gupta, S.K.; Elumalai, S.P. Adverse Impacts of Fog Events during Winter on Fine Particulate Matter, CO and VOCs: A Case Study of a Highway near Dhanbad, India. *Weather* **2018**, *73*, 396–402. [CrossRef]
- Martin, A.E. Mortality and Morbidity Statistics and Air Pollution. *R. Soc. Med.* **1964**, *57*, 969–975. [CrossRef] [PubMed]
- Holder, C.D. Fog Precipitation in the Sierra de Las Minas Biosphere Reserve, Guatemala. *Hydrol. Process.* **2003**, *17*, 2001–2010. [CrossRef]
- Bruijnzeel, L.S.; Eugster, W.; Burkard, R. Fog as a Hydrologic Input. In *Encyclopedia of Hydrological Sciences*; Anderson, M.G., Ed.; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2005.
- Bruijnzeel, L.A.; Scatena, F.N.; Hamilton, L.S. *Tropical Montane Cloud Forests: Science for Conservation and Management*; Cambridge University Press: Cambridge, UK, 2011; ISBN 9780511778384.
- García-Santos, G.; Bruijnzeel, L.A. Rainfall, Fog and Throughfall Dynamics in a Subtropical Ridge Top Cloud Forest, National Park of Garajonay (La Gomera, Canary Islands, Spain). *Hydrol. Process.* **2011**, *25*, 411–417. [CrossRef]
- Amani, M.; Mahdavi, S.; Bullock, T.; Beale, S. Automatic Nighttime Sea Fog Detection Using GOES-16 Imagery. *Atmos. Res.* **2020**, *238*, 104712. [CrossRef]
- Baguskas, S.A.; Clemesha, R.E.S.; Loik, M.E. Coastal Low Cloudiness and Fog Enhance Crop Water Use Efficiency in a California Agricultural System. *Agric. For. Meteorol.* **2018**, *252*, 109–120. [CrossRef]
- Yuan, Y.; Qiu, Z.; Sun, D.; Wang, S.; Yue, X. Daytime Sea Fog Retrieval Based on GOCI Data: A Case Study over the Yellow Sea. *Opt. Express* **2016**, *24*, 787. [CrossRef] [PubMed]
- Wu, D.; Lu, B.; Zhang, T.; Yan, F. A Method of Detecting Sea Fogs Using CALIOP Data and Its Application to Improve MODIS-Based Sea Fog Detection. *J. Quant. Spectrosc. Radiat. Transf.* **2015**, *153*, 88–94. [CrossRef]
- Zhang, S.; Yi, L. A Comprehensive Dynamic Threshold Algorithm for Daytime Sea Fog Retrieval over the Chinese Adjacent Seas. *Pure Appl. Geophys.* **2013**, *170*, 1931–1944. [CrossRef]



17. Bari, D.; Bergot, T.; El Khelifi, M. Numerical Study of a Coastal Fog Event over Casablanca, Morocco. *Q. J. R. Meteorol. Soc.* **2015**, *141*, 1894–1905. [CrossRef]
18. Obregon, A.; Gehrig-Downie, C.; Gradstein, S.R.; Bendix, J. The Potential Distribution of Tropical Lowland Cloud Forest as Revealed by a Novel MODIS-Based Fog/Low Stratus Night-Time Detection Scheme. *Remote Sens. Environ.* **2014**, *155*, 312–324. [CrossRef]
19. Lehnert, L.W.; Thies, B.; Trachte, K.; Achilles, S.; Osses, P.; Baumann, K.; Schmidt, J.; Samolov, E.; Jung, P.; Leinweber, P.; et al. A Case Study on Fog/Low Stratus Occurrence at Las Lomitas, Atacama Desert (Chile) as a Water Source for Biological Soil Crusts. *Aerosol Air Qual. Res.* **2018**, *18*, 254–269. [CrossRef]
20. Andersen, H.; Cermak, J.; Fuchs, J.; Knippertz, P.; Gaetani, M.; Quinting, J.; Sippel, S.; Vogt, R. Synoptic-Scale Controls of Fog and Low-Cloud Variability in the Namib Desert. *Atmos. Chem. Phys.* **2020**, *20*, 3415–3438. [CrossRef]
21. Parde, A.N.; Ghude, S.D.; Dhangar, N.G.; Bhautmage, U.P.; Wagh, S.; Lonkar, P.; Govardhan, G.; Kumar, R.; Biswas, M.; Chen, F. Challenges in Simulating Prevailing Fog Types Over Urban Region of Delhi. *J. Geophys. Res. Atmos.* **2024**, *129*, e2023JD039772. [CrossRef]
22. Hunt, G.E. Radiative Properties of Terrestrial Clouds at Visible and Infra-red Thermal Window Wavelengths. *Q. J. R. Meteorol. Soc.* **1973**, *99*, 346–369. [CrossRef]
23. Eyre, J.R.; Brownscombe, J.L.; Allam, R.J. Detection of Fog at Night Using Advanced Very High Resolution Radiometer (AVHRR) Imagery. *Meteorol. Mag.* **1984**, *113*, 266–271.
24. Turner, J.; Allam, R.J.; Maine, D.R. AA Case Study of the Detection of Fog at Night Using Channel 3 and 4 on the Advanced Very High Resolution Radiometer (AVHRR). *Meteorol. Mag.* **1986**, *115*, 285–290.
25. Bendix, J. A Case Study on the Determination of Fog Optical Depth and Liquid Water Path Using AVHRR Data and Relations to Fog Liquid Water Content and Horizontal Visibility. *Int. J. Remote Sens.* **1995**, *16*, 515–530. [CrossRef]
26. Bendix, J.; Bachmann, M. A Method for Detection of Fog Using AVHRR-Imagery of NOAA Satellites Suitable for Operational Purposes. *Meteorol. Rundsch.* **1991**, *43*, 169–178.
27. Ellrod, G. Advances in the Detection and Analysis of Fog at Night Using GOES Multispectral Infrared Imagery. *Weather. Forecast.* **1995**, *4*, 606–619. [CrossRef]
28. Wetzel, M.A.; Borys, R.D.; Xu, L.E. Satellite Microphysical Retrievals for Land-Based Fog with Validation by Balloon Profiling. *J. Appl. Meteorol.* **1996**, *5*, 810–829. [CrossRef]
29. Greenwald, T.J.; Christopher, S.A. The GOES I-M Imagers: New Tools for Studying Microphysical Properties of Boundary Layer Stratiform Clouds. *Bull. Am. Meteorol. Soc.* **2000**, *81*, 2607–2620. [CrossRef]
30. Bendix, J.; Thies, B.; Cermak, J.; Nauß, T. Ground Fog Detection from Space Based on MODIS Daytime Data-A Feasibility Study. *Am. Meteorol. Soc.* **2005**, *20*, 989–1005. [CrossRef]
31. Bendix, J. A Satellite-Based Climatology of Fog and Low-Level Stratus in Germany and Adjacent Areas. *Atmos. Res.* **2002**, *64*, 3–18. [CrossRef]
32. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, *372*, n71.
33. Prancutė, R. Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World. *Publications* **2021**, *9*, 12. [CrossRef]
34. van Eck, J.N.; Waltman, L. *Manual for VOSviewer*, Version 1.6.20; 2023. Available online: [https://www.vosviewer.com/documentation/Manual\\_VOSviewer\\_1.6.20.pdf](https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.20.pdf) (accessed on 13 January 2024).
35. Fu, D.; Wang, S.; Chen, D.; Pang, H.; Li, P. Comparison Study between Observation and Simulation for Sea Fog over the Yellow Sea in May 2009. *J. Ocean Univ. China* **2012**, *11*, 290–300. [CrossRef]
36. Fu, G.; Li, P.; Crompton, J.G.; Guo, J.; Gao, S.; Zhang, S. An Observational and Modeling Study of a Sea Fog Event over the Yellow Sea on 1 August 2003. *Meteorol. Atmos. Phys.* **2010**, *107*, 149–159. [CrossRef]
37. Guo, Y.; Xu, T.; Cheng, J.; Wei, G.; Lin, Y. Above- and Belowground Biodiversity Drives Soil Multifunctionality along a Long-Term Grassland Restoration Chronosequence. *Sci. Total Environ.* **2021**, *772*, 145010. [CrossRef]
38. Lu, H.; Ma, Y.; Zhang, S.; Yu, X.; Zhang, J. Daytime Sea Fog Identification Based on Multi-Satellite Information and the ECA-TransUnet Model. *Remote Sens.* **2023**, *15*, 3949. [CrossRef]
39. Ran, Y.; Ma, H.; Liu, Z.; Wu, X.; Li, Y.; Feng, H. Satellite Fog Detection at Dawn and Dusk Based on the Deep Learning Algorithm under Terrain-Restriction. *Remote Sens.* **2022**, *14*, 4328. [CrossRef]
40. Yan, H.; Su, S.; Wu, M.; Xu, M.; Zuo, Y.; Zhang, C.; Huang, B. SeaMAE: Masked Pre-Training with Meteorological Satellite Imagery for Sea Fog Detection. *Remote Sens.* **2023**, *15*, 4102. [CrossRef]
41. Zhou, Y.; Chen, K.; Li, X. Dual-Branch Neural Network for Sea Fog Detection in Geostationary Ocean Color Imager. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4208617. [CrossRef]
42. Chen, C.; Zhang, M.; Perrie, W.; Chang, R.; Chen, X.; Duplessis, P.; Wheeler, M. Boundary Layer Parameterizations to Simulate Fog Over Atlantic Canada Waters. *Earth Space Sci.* **2020**, *7*, 2019EA000703. [CrossRef]



43. Mahdavi, S.; Amani, M.; Bullock, T.; Beale, S. A Probability-Based Daytime Algorithm for Sea Fog Detection Using GOES-16 Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1363–1373. [CrossRef]
44. Gultepe, I.; Pardyjak, E.; Hoch, S.W.; Fernando, H.J.S.; Dorman, C.; Flagg, D.D.; Krishnamurthy, R.; Wang, Q.; Gaberšek, S.; Creegan, E.; et al. Coastal-Fog Microphysics Using In-Situ Observations and GOES-R Retrievals. *Boundary Layer. Meteorol.* **2021**, *181*, 203–226. [CrossRef]
45. Ahmed, R.; Dey, S.; Mohan, M. A Study to Improve Night Time Fog Detection in the Indo-Gangetic Basin Using Satellite Data and to Investigate the Connection to Aerosols. *Meteorol. Appl.* **2015**, *22*, 689–693. [CrossRef]
46. Banerjee, S.; Padmakumari, B. Spatiotemporal Variability and Evolution of Day and Night Winter Fog over the Indo Gangetic Basin Using INSAT-3D and Comparison with Surface Visibility and Aerosol Optical Depth. *Sci. Total Environ.* **2020**, *745*, 140962. [CrossRef] [PubMed]
47. Saraf, A.K.; Bora, A.K.; Das, J.; Rawat, V.; Sharma, K.; Jain, S.K. Winter Fog over the Indo-Gangetic Plains: Mapping and Modelling Using Remote Sensing and GIS. *Nat. Hazards* **2011**, *58*, 199–220. [CrossRef]
48. Singh, M.K.; Gautam, R. Developing a Long-Term High-Resolution Winter Fog Climatology over South Asia Using Satellite Observations from 2002 to 2020. *Remote Sens. Environ.* **2022**, *279*, 113128. [CrossRef]
49. Gautam, R.; Singh, M.K. Urban Heat Island Over Delhi Punches Holes in Widespread Fog in the Indo-Gangetic Plains. *Geophys. Res. Lett.* **2018**, *45*, 1114–1121. [CrossRef]
50. Cereceda, P.; Larrain, H.; Osses, P.; Farías, M.; Egaña, I. The Spatial and Temporal Variability of Fog and Its Relation to Fog Oases in the Atacama Desert, Chile. *Atmos. Res.* **2008**, *87*, 312–323. [CrossRef]
51. Bari, D.; Bergot, T. Influence of Environmental Conditions on Forecasting of an Advection-Radiation Fog: A Case Study from the Casablanca Region, Morocco. *Aerosol Air Qual. Res.* **2018**, *18*, 62–78. [CrossRef]
52. Baldocchi, D.; Waller, E. Winter Fog Is Decreasing in the Fruit Growing Region of the Central Valley of California. *Geophys. Res. Lett.* **2014**, *41*, 3251–3256. [CrossRef]
53. Haensler, A.; Cermak, J.; Hagemann, S.; Jacob, D. Will the Southern African West Coast Fog Be Affected by Future Climate Change?: Results of an Initial Fog Projection Using a Regional Climate Model. *Erdkunde* **2011**, *65*, 261–275. [CrossRef]
54. Pohl, M.J.; Lehnert, L.; Bader, M.Y.; Gradstein, S.R.; Viehweger, J.; Bendix, J. A New Fog and Low Stratus Retrieval for Tropical South America Reveals Widespread Fog in Lowland Forests. *Remote Sens. Environ.* **2021**, *264*, 112620. [CrossRef]
55. Pohl, M.J.; Lehnert, L.W.; Thies, B.; Seeger, K.; Berdugo, M.B.; Gradstein, S.R.; Bader, M.Y.; Bendix, J. Valleys Are a Potential Refuge for the Amazon Lowland Forest in the Face of Increased Risk of Drought. *Commun. Earth Environ.* **2023**, *4*, 198. [CrossRef]
56. Schulz, H.M.; Thies, B.; Chang, S.C.; Bendix, J. Detection of Ground Fog in Mountainous Areas from MODIS (Collection 051) Daytime Data Using a Statistical Approach. *Atmos. Meas. Tech.* **2016**, *9*, 1135–1152. [CrossRef]
57. Schulz, H.M.; Li, C.F.; Thies, B.; Chang, S.C.; Bendix, J. Mapping the Montane Cloud Forest of Taiwan Using 12 Year MODIS-Derived Ground Fog Frequency Data. *PLoS ONE* **2017**, *12*, E0172663. [CrossRef]
58. Garland, R.M.; Altieri, K.E.; Dawidowski, L.; Gallardo, L.; Mbandi, A.; Rojas, N.Y.; Touré, N.E. Opinion: Strengthening Research in the Global South—Atmospheric Science Opportunities in South America and Africa. *Atmos. Chem. Phys.* **2024**, *24*, 5757–5764. [CrossRef]
59. Yi, L.; Li, K.F.; Chen, X.; Tung, K.K. Arctic Fog Detection Using Infrared Spectral Measurements. *J. Atmos. Ocean. Technol.* **2019**, *36*, 1643–1656. [CrossRef]
60. Wang, J.L.; Li, S.M.; Liu, X.L.; Wu, X.J. An Analysis of the Fog Distribution in Beijing for the 2001–2005 Period Using NOAA and FY Data. *Atmos. Res.* **2010**, *96*, 575–589. [CrossRef]
61. Guo, J.; Xu, J.; Xu, X. An Observational and Modeling Study of a Sea Fog Event over the Yellow and East China Seas on 17 March 2014. *Trop. Cyclone Res. Rev.* **2021**, *10*, 182–190. [CrossRef]
62. Kim, Y.; Ryu, H.S.; Hong, S. Data-to-Data Translation-Based Nowcasting of Specific Sea Fog Using Geostationary Weather Satellite Observation. *Atmos. Res.* **2023**, *290*, 106792. [CrossRef]
63. Ball, L.; Tzanopoulos, J. Interplay between Topography, Fog and Vegetation in the Central South Arabian Mountains Revealed Using a Novel Landsat Fog Detection Technique. *Remote Sens. Ecol. Conserv.* **2020**, *6*, 498–513. [CrossRef]
64. Li, H.J.; Lo, M.H.; Juang, J.Y.; Wang, J.; Huang, C. Assessment of Spatiotemporal Dynamics of Diurnal Fog Occurrence in Subtropical Montane Cloud Forests. *Agric. For. Meteorol.* **2022**, *317*, 108899. [CrossRef]
65. Egli, S.; Thies, B.; Bendix, J. A Hybrid Approach for Fog Retrieval Based on a Combination of Satellite and Ground Truth Data. *Remote Sens.* **2018**, *10*, 628. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## Article

# Vegetation Response to Flash Drought Events Considering Resilience in Southwestern China

Liangliang Jiang <sup>1,2,3</sup>, Guangming Wu <sup>2,3</sup>, Qijin Li <sup>2,3</sup> and Xiaoran Liu <sup>1,\*</sup><sup>1</sup> Chongqing Institute of Meteorological Sciences, Chongqing 401147, China; jiang@cqnu.edu.cn<sup>2</sup> School of Geography and Tourism, Chongqing Normal University, Chongqing 401331, China; 2023110514089@stu.cqnu.edu.cn (G.W.); 2023110514079@stu.cqnu.edu.cn (Q.L.)<sup>3</sup> Chongqing Key Laboratory of GIS Application, Chongqing 401331, China

\* Correspondence: liuxiaoran8283@126.com

**Abstract:** Flash drought events occur frequently in Southwestern China, and a notable upward trend is predicted for the future. Attention should be given to how the severity of flash droughts and vegetation vulnerability hinder vegetation from recovering to their original state, leading to losses. Vegetation resilience and vulnerability to flash droughts was assessed in dry years by adopting a ‘resistance–resilience’ framework from a new perspective, and we measured the significance of various drought characteristics in affecting vegetation reduction by using the boosted regression tree (BRT) model. The results showed that croplands in the Sichuan Basin displayed low resistance to flash droughts, whereas grasslands and forests in mountainous areas had high resistance. Croplands in the Sichuan Basin demonstrated high vegetation resilience, while Guizhou province showed low vegetation resilience. Most regions experienced high vegetation vulnerability to flash droughts, especially in the Sichuan Basin and Yunnan province. We found that croplands and forests in 2006 exhibited a significant decrease in LAI during flash drought events. Croplands experienced a significant decrease in LAI in regions where the drought duration (DD) exceeded 60 days, and the drought interval (DIV) ranged from 30 to 40 days. Forest regions with a DD exceeding 60 days and a DIV below 20 days experienced a high reduction in LAI. Furthermore, croplands and shrubs could recover once their vulnerability fell below thresholds of 0.34 and 0.30, respectively. The impact of species richness on vegetation resilience can be explored in future research. This study reveals the spatial patterns of vegetation vulnerability and provides information on preventing and managing vegetation deterioration in Southwestern China.

**Keywords:** flash drought events; vegetation vulnerability; vegetation loss; vegetation recovery; Southwestern China

## 1. Introduction

Flash droughts have become the new normal. They occur abruptly and have a short period and high intensity, adversely affecting vegetation productivity and attracting widespread global attention [1,2]. Flash droughts are rapid-onset drought events that develop and intensify quickly, typically within a few days or weeks, as opposed to the gradual onset of more traditional droughts [1]. Drought intensity (DI) and the frequency of flash droughts have increased. The Intergovernmental Panel on Climate Change (IPCC) Special Report on Extreme Events reported a shift toward a higher occurrence of flash droughts in 74% of the world’s regions over the past half-century. Flash droughts are expected to occur

in most global land areas, with more significant increases anticipated under higher future emission scenarios [1]. Moreover, flash drought events have become increasingly common in Southwestern China, indicating a significant upward trend in flash drought risk [3]. Flash droughts adversely affect food, water security, and environmental sustainability. In the summer of 2006 and 2022, a rare flash drought occurred in the southwestern region, encompassing the entire Yangtze River basin. The affected agricultural land area in 2022 was 6090.2 thousand hectares, resulting in direct economic losses of CNY 51.28 billion. Thus, adapting to the increased frequency of flash droughts in a warmer future is essential. There has been a burgeoning academic interest in understanding vegetation responses to rapid drought events [4].

Unlike normal droughts that evolve slowly, flash droughts can be triggered by unusually high evapotranspiration (ET) rates, leading to the rapid depletion of soil moisture, significantly increasing vegetation stress [5]. Furthermore, due to the abrupt arrival of flash droughts, there is minimal opportunity to prevent the detrimental consequences of vegetation loss. Therefore, flash droughts have more substantial adverse effects on vegetation ecosystems than normal drought events [6]. Monitoring flash droughts and evaluating vegetation ecosystem responses pose significant challenges [7]. Furthermore, vegetation undergoes notable fluctuations during flash droughts. Evaluating vegetation conditions solely on short-term fluctuations during flash droughts can result in an inaccurate assessment. Thus, understanding vegetation responses to flash drought events is crucial, and vegetation resistance and recovery must be considered [8,9]. This information is required to enhance our comprehension of processes leading to vegetation degradation and improve the efficacy of prevention and management measures [8]. Vegetation has a self-regulating capacity known as resilience. When flash droughts exceed this capacity, the vegetation may not recover to its original condition and may degrade [10]. Vegetation resilience refers to the capacity to self-regulate during and after flash droughts. This resilience process is crucial in monitoring the responses of vegetation to flash droughts.

Traditional methods for assessing the vegetation response to droughts include partial correlation, residual, and seasonal abrupt trend analyses. These methods have produced valuable results and improved the understanding of vegetation change patterns and drought responses [11–13]. They were used to examine vegetation responses to climate change during the drought or smaller time intervals in the mutation year [14]. However, vegetation is more sensitive during the drought period [15]. Thus, it is necessary to consider vegetation resilience during and after flash drought effects and analyze the vegetation resistance and recovery periods. Despite progress in flash drought studies, no uniform definition of flash droughts exists [1,16]. Soil moisture anomalies are valuable in delineating drought initiation, especially for flash droughts. A rapid drop in soil moisture levels can indicate a flash drought [17]. Current definitions of flash drought state that the drought ends when soil moisture returns to the 20th percentile during recovery. However, soil moisture variability affects vegetation health [18]. A flash drought may not last long enough to have a detrimental effect on vegetation ecosystems. Researchers have assessed the impact of flash drought events on vegetation considering vegetation resilience. Principal component analysis, autoregressive models, and remote sensing technology were used to monitor vegetation resilience to climate change [19]. Technological advancements have significantly improved the evaluation of vegetation resilience. Vegetation resilience was assessed by establishing a linear model of the historical state and the drought index during a given period based on autoregressive models [20]. A flash drought event was detected using a censored time series to determine the beginning and end of flash droughts and their effects. Autoregressive models do not provide insights into the vegetation resistance and recovery

duration. Moreover, the IPCC suggested that assessing vegetation vulnerability should consider vegetation resistance and its capacity to recover from the drought effects [21]. Thus, it is necessary to determine vegetation vulnerability, especially resistance and resilience, to drought events [22]. Furthermore, the run theory, which is extensively employed in drought research, can be used to identify drought events and their fundamental characteristics, such as intensity, duration, and frequency [23]. The vegetation loss is closely related to drought characteristics and vegetation resilience. Attention should be given to how the severity of flash droughts and vegetation vulnerability hinder vegetation from recovering to their original state, leading to losses. Therefore, it is crucial to examine the resilience, vulnerability, and loss of vegetation to flash droughts in various regions and consider different vegetation types to obtain an accurate estimate of vegetation response process to flash droughts. Understanding the thresholds of vegetation vulnerability for vegetation loss after flash droughts can be used to prevent and manage vegetation degradation in Southwestern China.

A notable rise in the likelihood of flash droughts has been observed throughout China, particularly in the southwestern regions [3]. The intensity of flash droughts will likely increase substantially by the middle and end of the current century. It is expected to rise by more than 20% across most of Southwestern China [3]. Southwestern China experienced extreme flash drought events in 2006 and 2022 marked by a rainfall deficit and an exceptionally high temperature anomaly, resulting in significant economic and agricultural losses. The flash droughts of 2006 and 2022 were the most severe in recent decades [24,25]. However, few studies have assessed the resistance and resilience of vegetation to flash drought events and vegetation loss in Southwestern China, considering the characteristics of these events across different regions. Understanding the thresholds of vegetation vulnerability for vegetation recovery after flash droughts is essential for guiding decision-making to prevent vegetation deterioration. Hence, the typical flash drought years of 2006 and 2022 were analyzed and compared in this study. The objectives were as follows: (1) to analyze the features of flash drought events in 2006 and 2022, (2) to evaluate the resilience and vulnerability of vegetation to flash drought events in both years, and (3) to measure vegetation loss during flash drought events and explore the thresholds of vegetation vulnerability for vegetation recovery after flash droughts.

## 2. Materials and Methods

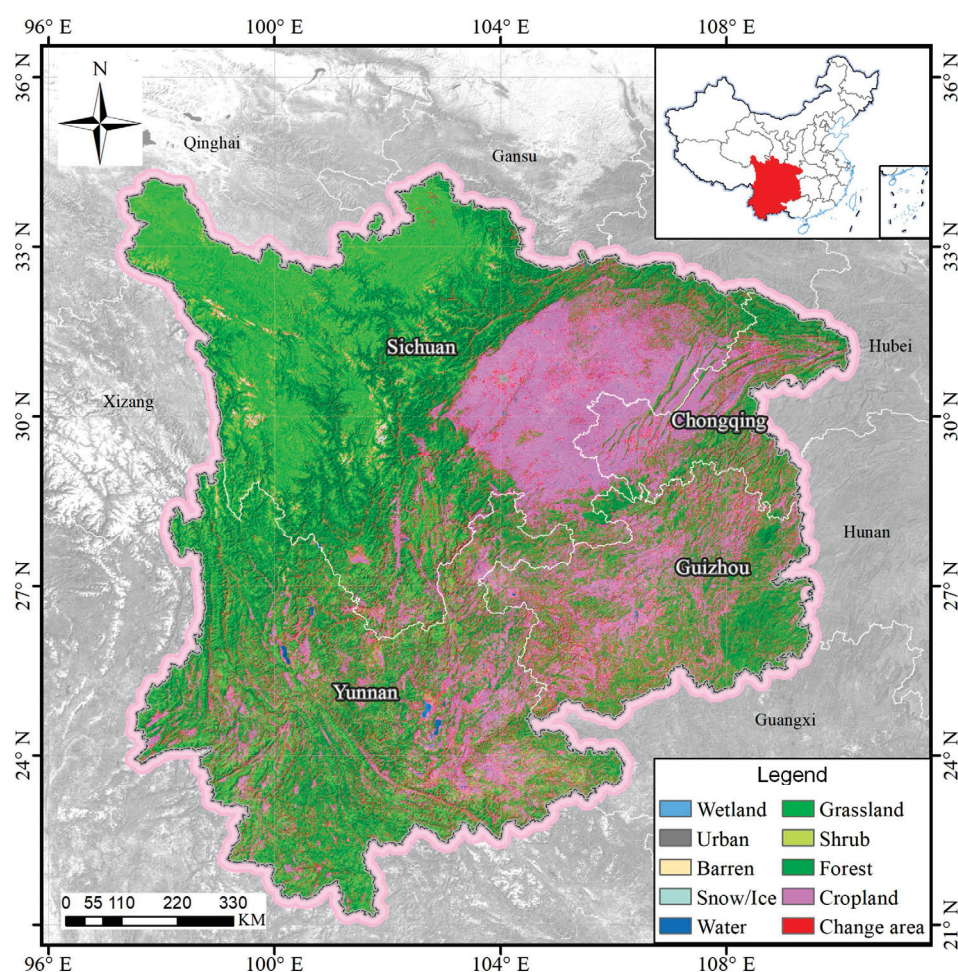
### 2.1. Study Area

The narrowest concept of Southwest China includes Sichuan, Chongqing, Yunnan, and Guizhou (Figure 1), representing the study area. Southwest China encompasses towering mountain ranges, fertile river valleys, lush tropical forests, and arid plateaus. Most of the area is drained by the Yangtze River, which shapes the Three Gorges in the upper reaches. The area is known for its subtropical monsoon climate, featuring a unique karst topography. The elevation increases from the southeast to the northwest [26]. The unique geography of Southwestern China contributes to its rich biodiversity and is a subject of interest for scholars [27].

Southwest China has one of the world's most extensive continuous karst landscapes and is home to over 30 million people, a significant portion living in poverty. Rocky desertification is the most severe land degradation issue in Southwestern China. The vegetation vulnerability in this region is primarily due to the effects of geology, topography, climate, and human activities [28]. Southwest China has diverse land cover categories, such as croplands (24.68%), grasslands (18.14%), forests (56.33%), and shrublands (0.85%) (Figure 1). Croplands are mainly concentrated in the Sichuan Basin. Most grasslands



occur in the northwest, whereas forests are predominantly found in the southwestern part. Shrublands have the smallest proportion and are scattered throughout the study area.



**Figure 1.** Vegetation types in Southwestern China.

Flash droughts are occurring increasingly in Southwest China [3]. The region's vulnerability to rapid and intense flash drought events has significant implications for agriculture, water resources, and ecosystems. Vegetation vulnerability to flash droughts in Southwestern China is a significant concern with potential long-term ecological consequences. Thus, this study conducts an in-depth examination of vegetation loss, vulnerability, and resilience to flash droughts to provide information on implementing proactive measures to mitigate and adapt to these events.

## 2.2. Data Sources

Leaf area index (LAI), soil moisture, and land use data were utilized in this study (Table S1, Supplementary Materials).

The MCD15A3H LAI MODIS product is a global, high-resolution dataset depicting the distribution and density of green vegetation (<https://ladsweb.modaps.eosdis.nasa.gov/search/> (accessed on 15 October 2024)). MCD15A3H is a 4-day composite dataset with 500 m spatial resolution [29]. It is derived from time-series MODIS observations, making it a valuable resource for monitoring vegetation health over time. The MCD15A3H dataset provides valuable information for various research applications related to environmental monitoring, land cover analysis, climate studies, and more. The MCD15A3H LAI data



were used in this study to monitor changes in vegetation health and dynamics over time. This information is crucial for assessing the impact of flash droughts.

Soil moisture data were obtained from the National Tibetan Plateau/Third Pole Environment Data Center (<https://data.tpdc.ac.cn/en/data/e1f24e35-6235-40b2-b3d7-677dfb249e39/>) (accessed on 13 October 2024)). This product was of 1 km resolution and was produced by downscaling AMSR-E and AMSR-2 data, encompassing all periods of operation for the two radiometers from 2003 to 2022 [30]. The daily soil moisture data covered the study area almost completely. The product was validated using soil moisture measurements from observation stations. The evaluation results showed high accuracy, with an unbiased root mean square error of  $0.053 \text{ cm}^3/\text{cm}^3$  to  $0.056 \text{ cm}^3/\text{cm}^3$  [30], indicating the reliability of the data for monitoring flash drought events in Southwestern China.

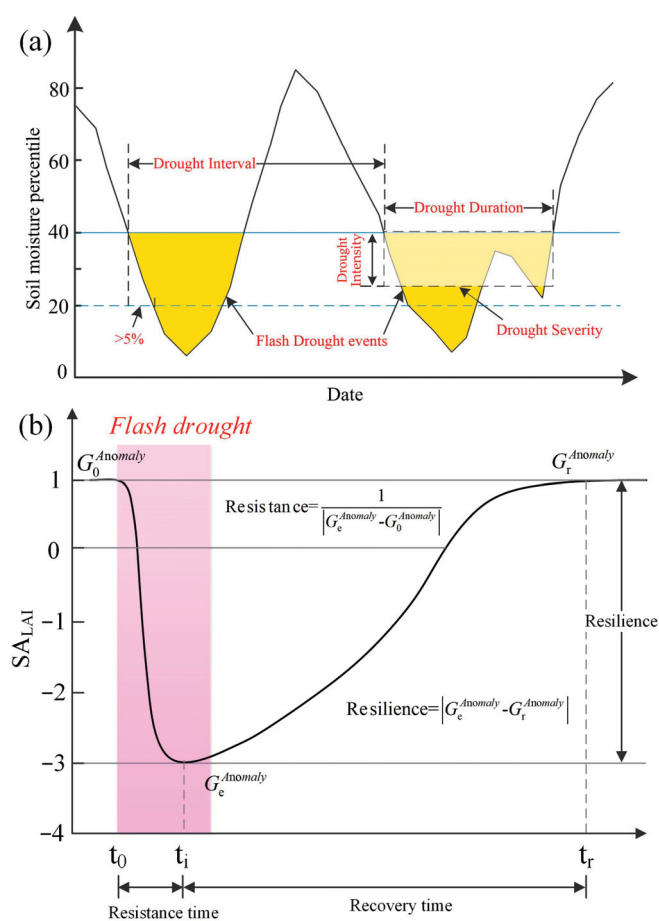
The land use data were the 30 m resolution annual land cover datasets in China from 1985 to 2022 [31]. This dataset was created using 335,709 Landsat images (<https://zenodo.org/record/8176941>) (accessed on 13 October 2024)). Numerous temporal metrics were generated using all accessible Landsat data and input into a random forest classifier to derive the land use/land cover dataset (Table S2, Supplementary Materials). Land use changes during the study period were not considered in this study, and the vegetation types without land use changes between 2006 and 2022 were extracted to ensure statistical comparability. According to the previous studies [32,33], high spatial resolution remote sensing data were resampled to a lower spatial resolution, and the results from the low resolution data remain accurate within an acceptable range. Thus, the LAI and land use data were resampled to 1 km resolution using bilinear interpolation to align the data with the spatial resolution of the soil moisture data, obtain a uniform resolution, and enable precise comparisons [34,35].

### 2.3. Methods

#### 2.3.1. Flash Drought Events Definition

A flash drought is defined by a sudden commencement, gradual intensification, and a short period, i.e., several weeks to a few months. Currently, most studies have relied on soil moisture to track flash drought events [1,3,17,36]. Soil moisture anomalies are valuable for identifying the onset of droughts, especially for flash droughts [1]. Flash drought events in this study were distinguished by a sudden and significant decline in soil moisture. The definition of a flash drought event and its characteristics based on the run theory are displayed in Figure 2. The theory describes a “run” as a continuous sequence of values in a time series that either rise above or fall below a particular threshold. A positive run happens when the values exceed the threshold, whereas a negative run occurs when the values stay below it. A flash drought is a negative run characterized by a rapid decline [37]. The average daily soil moisture for four days was used to match the temporal resolution of the LAI data.

Soil moisture exhibits substantial regional differences. It cannot be used directly for flash drought monitoring [2,18]. The soil moisture percentile was used as a straightforward and resilient approach to relate soil moisture to flash droughts and enable temporal and spatial comparisons. The following criteria were used to detect flash drought events [1]: (1) a continuous duration of a flash drought event below the 40th percentile for more than 20 days; (2) a minimum soil moisture value of less than the 20th percentile; (3) the transition of the soil moisture from greater than the 40th percentile to the 20th percentile should have a minimum average decrease rate of 5%. Its characteristics were quantified using the run theory [37], as shown in Figure 2a.



**Figure 2.** The definition of a flash drought event and its characteristics (a) and vegetation resistance and resilience during a flash drought event (b).

The run theory was used to analyze the basic properties of flash drought events. It involves identifying consecutive periods of drought events or other drought-related variables to quantify different aspects of flash drought events. Drought intensity (DI) is typically quantified by calculating the mean soil moisture during flash drought events. Drought severity (DS) is closely related to intensity but considers the cumulative impact of the drought events. Drought interval (DIV) refers to the time between consecutive drought events. It measures how often droughts occur and offers insights into the recurrence of drought in a region. The number of droughts (DT) represents the number of flash drought events within a specified period. Drought duration (DD) is the length of time that a drought event persists. DS and DI are quantified using the following equations:

$$DS = \sum_{i=1}^{DD} |SM_i| \quad (1)$$

$$DI = \frac{\sum_{i=1}^{DD} |SM_i|}{DD} \quad (2)$$

where DD represents drought duration; DI is drought intensity; and DS denotes drought severity.

### 2.3.2. Vegetation Resilience and Vulnerability to Flash Drought Events

Based on the resistance–resilience framework, vegetation resistance and resilience to flash drought events were calculated in this study (Figure 2b). This framework provides

valuable insights for tracking vegetation changes in response to flash droughts disturbances [38,39]. A standardized anomaly (SA) is a statistical measure used to quantify and compare anomalies or deviations from a reference or baseline value. The SAs for LAI and soil moisture were calculated as follows [40]:

$$SA_i = \frac{x_i - \mu(x)}{\delta(x)} \quad (3)$$

where  $SA_i$  is the SA on the  $i$ th day;  $x_i$  denotes the data point on the  $i$ th day;  $\mu(x)$  is the average value; and  $\delta(x)$  represents the standard deviation of the dataset.

Vegetation resistance was quantified by considering the pre-drought vegetation conditions and calculating the disparity between the SA before and during the flash drought events [38,39]. Vegetation resistance time is the duration from the onset of a flash drought event to the day when vegetation experiences the most severe impact. Vegetation resistance is determined as follows:

$$Resistance = \frac{1}{|G_e^{Anomaly} - G_0^{Anomaly}|} \quad (4)$$

$$G_0^{Anomaly} = \frac{G_{onset}^{Anomaly} + G_{onset-1}^{Anomaly}}{2} \quad (5)$$

where  $G_e^{Anomaly}$  refers to the SA in the LAI on the most severe flash drought day, and  $G_0^{Anomaly}$  represents the SA in the LAI before the initiation of the flash drought.  $G_{onset}^{Anomaly}$  and  $G_{onset-1}^{Anomaly}$  denotes the SA in the LAI at the start and before the drought events, respectively.

The resilience of vegetation is identified as the relative anomalies during and after flash drought events, indicating the ability of vegetation restoration [41]. Vegetation recovery time is the period from the day of the most severe impact to the day when the vegetation reaches the most favorable condition. The most favorable day is the day when the LAI has a positive anomaly after a flash drought event ends, with a subsequent decrease the following day. Vegetation resilience during drought events can be quantified as follows [41]:

$$Resilience = |G_e^{Anomaly} - G_r^{Anomaly}| \quad (6)$$

where  $G_r^{Anomaly}$  denotes the SA of the LAI on the most favorable recovery day.

If several cycles of resistance and resilience occur in a dry year, the average levels of resistance and resilience are determined as follows:

$$Resistance = \frac{1}{|\sum G_e^{Anomaly} - \sum G_0^{Anomaly}|} \quad (7)$$

$$Resilience = |\sum G_e^{Anomaly} - \sum G_r^{Anomaly}| \quad (8)$$

The resistance value reflects the ability of vegetation to withstand flash drought events. A higher value indicates stronger resistance. Conversely, higher resilience values show a stronger capacity for vegetation recovery. Vegetation exhibiting high resistance and resilience values is more robust in enduring and recovering from drought damage.

The robust resistance and capacity of vegetation to recover from flash droughts can significantly mitigate the adverse impacts of arid conditions, reducing vulnerability to

drought [42,43]. The assessment of vegetation vulnerability considering resilience was conducted as follows [43]:

$$Vulnerability = \frac{1}{Resistance} \times \frac{1}{Resilience} = \frac{|G_e^{Anomaly} - G_0^{Anomaly}|}{|G_e^{Anomaly} - G_r^{Anomaly}|} \quad (9)$$

where  $G_e^{Anomaly}$  refers to the SA in the LAI on the most severe flash drought day, and  $G_0^{Anomaly}$  represents the SA in the LAI before the initiation of the flash drought.  $G_r^{Anomaly}$  denotes the SA of the LAI on the most favorable recovery day.

### 2.3.3. Geographically Weighted Regression

The convergence analysis of vegetation vulnerability to flash droughts has revealed a negative association between the discrepancies in the vulnerability value during a given period and the initial vulnerability value [44]. Convergence analysis involves comparing the average change in the vulnerability value over time with the initial vulnerability value, assuming that regions with the lowest initial values experience a faster transition toward the mean vulnerability value. Thus, regions with low or high vulnerability values at the start of the analysis period tend to approach the mean vulnerability value for their respective areas. Therefore, a consistent negative relationship exists between the vegetation vulnerability value at the beginning and end of the analysis period across all regions [44]. The strength of this negative relationship indicates the extent to which vulnerability values in all regions converge to the mean value.

The average vegetation vulnerability values for the administrative subdivisions are presented in Figure S1. Subsequently, we computed the average annual change rates for different time intervals from 2006 to 2022. We employed an ordinary least squares (OLS) regression model to detect the spatial convergence of vulnerability in Southwest China. The convergence analysis followed the methodology of previous studies [45,46].

$$\Delta VI_{(2022-2006,i)} = b_0 + b_1 \times VI_{(2006,i)} + \varepsilon \quad (10)$$

where  $\Delta VI_{(2022-2006,i)}$  denotes the change in vegetation vulnerability during a given period from 2006 to 2022;  $VI_{(2006,i)}$  denotes the initial vulnerability value in the  $i$ th spatial unit in 2006; and  $\varepsilon$  represents the error term.  $b_1$  and  $b_0$  are the coefficients of the explanatory variables and intercept parameter, respectively.

Convergence analysis was performed using geographically weighted regression (GWR). GWR is a spatial regression technique that is increasingly used in geography. GWR creates a localized model to understand or predict variables by fitting regression equations to each feature in the dataset. Individual equations are established to describe the relationships between the dependent and explanatory variables of features within the range of target features [47]. It is assumed that the relationship between variables differs at different locations. This method considers the geographical context and spatial heterogeneity, providing a precise and localized understanding of variable relationships [47]. We assessed the statistical significance of the coefficients ( $b_1$ ) at the 95% confidence level using F-tests. Positive coefficients reflect a divergence trend in vulnerability, and negative coefficients indicate a convergence trend.

### 2.3.4. Boosted Regression Trees

A boosted regression tree (BRT) model was used to determine the impact of flash drought characteristics on vegetation reduction over flash droughts [48]. The BRT model,

also known as a gradient boosting machine (GBM), is a powerful machine learning technique that combines the principles of regression trees and boosting. It is an ensemble learning method that utilizes sequential regression trees. Each tree is designed to correct errors made by the previous ones, improving the overall predictive performance [49]. Boosting merges several weak models to create a robust predictive model. A BRT sequentially fits new trees to the residual errors of the previous trees. Regression trees are decision trees used for predicting continuous outcomes. A regression tree splits the data into branches to make predictions based on the input features [49]. The model starts with a simple regression tree that makes initial predictions. Then, the difference between the actual and predicted values (residuals) is calculated, and a new tree is trained on the residuals to correct the errors made by the previous tree [50]. Subsequently, the model is updated by combining the predictions of the new tree. The iterative process is repeated for a specified number of iterations or until the error reaches an acceptable level [51]. BRT achieves high accuracy by iteratively reducing the model error. It can handle various data complexities, including nonlinearity and variable interactions [52].

In this study, the impact of flash drought characteristics on vegetation reduction over flash droughts is a non-linear relationship. Furthermore, different drought characteristics interact and jointly influence vegetation reduction over flash droughts. Thus, this model was selected in this study. The model requires the specification of the following three key parameters: (1) learning rate (0.0001–0.1), (2) tree complexity (1–10), and (3) bag fraction (0–1). To identify the optimal parameters, all combinations were evaluated based on the lowest prediction error. The findings indicated that the optimal fit was achieved with a learning rate of 0.005, a tree complexity of 6, and a bag fraction of 0.5, resulting in high predictive accuracy.

### 2.3.5. Vegetation Recovery Calculation

Vegetation recovery is the recovery condition of vegetation after the end of the last flash drought event. The condition of vegetation recovery following flash droughts was calculated by the following formula:

$$Re = \sum_{i=n}^m (LAI_i - \overline{LAI}_i) \quad (11)$$

where  $Re$  represents the vegetation anomaly after flash droughts;  $LAI_i$  refers the value of LAI in the  $i$ th day;  $\overline{LAI}_i$  is the mean value of LAI in the  $i$ th day;  $n$  is the day after the last flash drought in the dry year; and  $m$  is the end day of the year.

To identify the thresholds of vegetation vulnerability for vegetation recovery following flash droughts, cubic polynomial analysis was applied to describe the relationships between vegetation vulnerability and vegetation recovery conditions [53]. The influence coefficient refers to the change in vegetation recovery brought by every 1 unit increase in the vegetation vulnerability, which represents the intensity and efficiency of the impact of the vegetation vulnerability. The influence coefficient was obtained by differentiating the cubic polynomial of the vegetation recovery and vegetation vulnerability. The calculation is as follows [53]:

$$E = \frac{d(f(\text{vegetation vulnerability}))}{d(\text{vegetation vulnerability})} \quad (12)$$

where  $f(\text{vegetation vulnerability})$  is the expression function between the vegetation recovery and vegetation vulnerability, and  $E$  is the influence coefficient.

To help clarify the workflow, the methodological framework was provided in Figure 3.



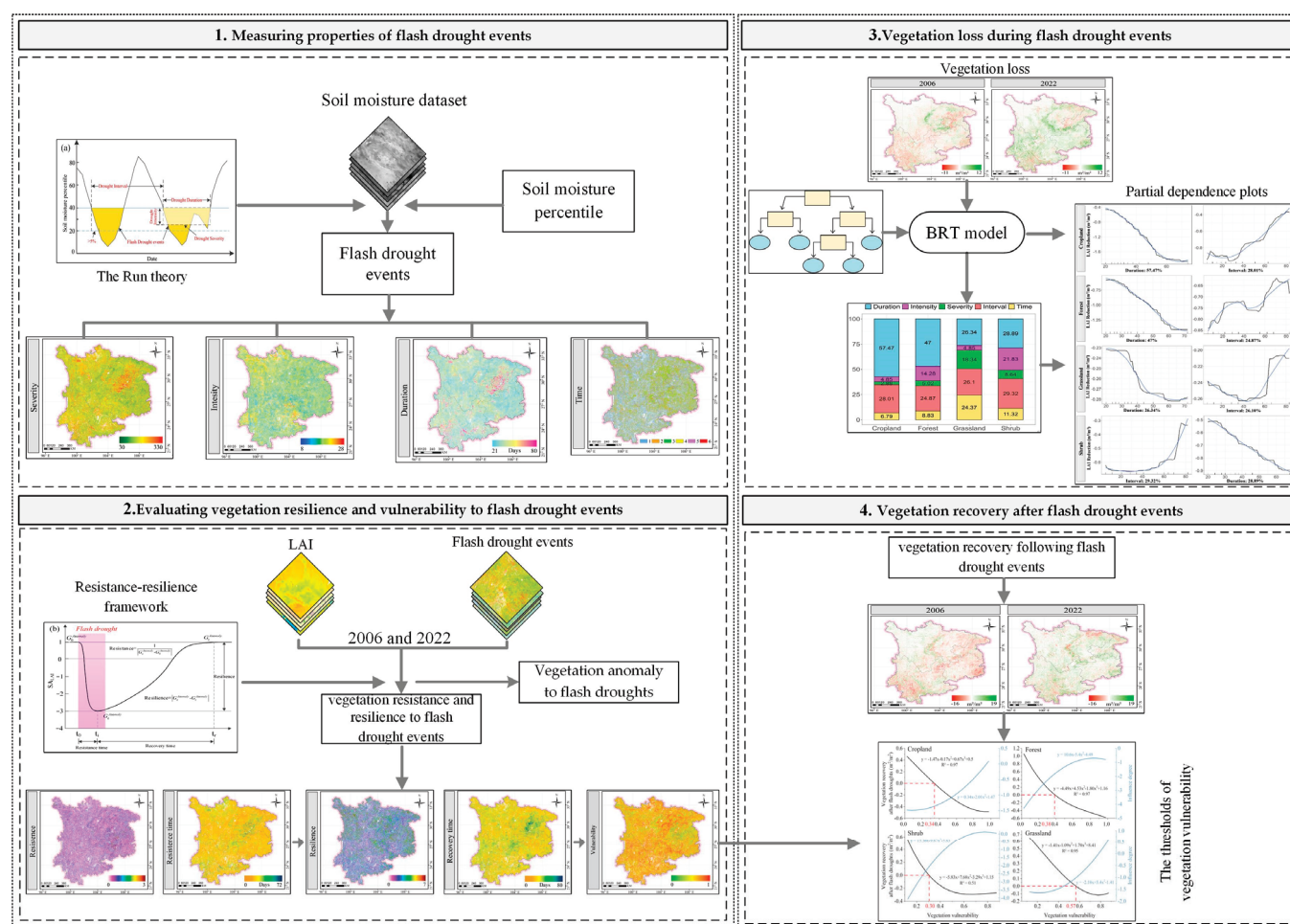
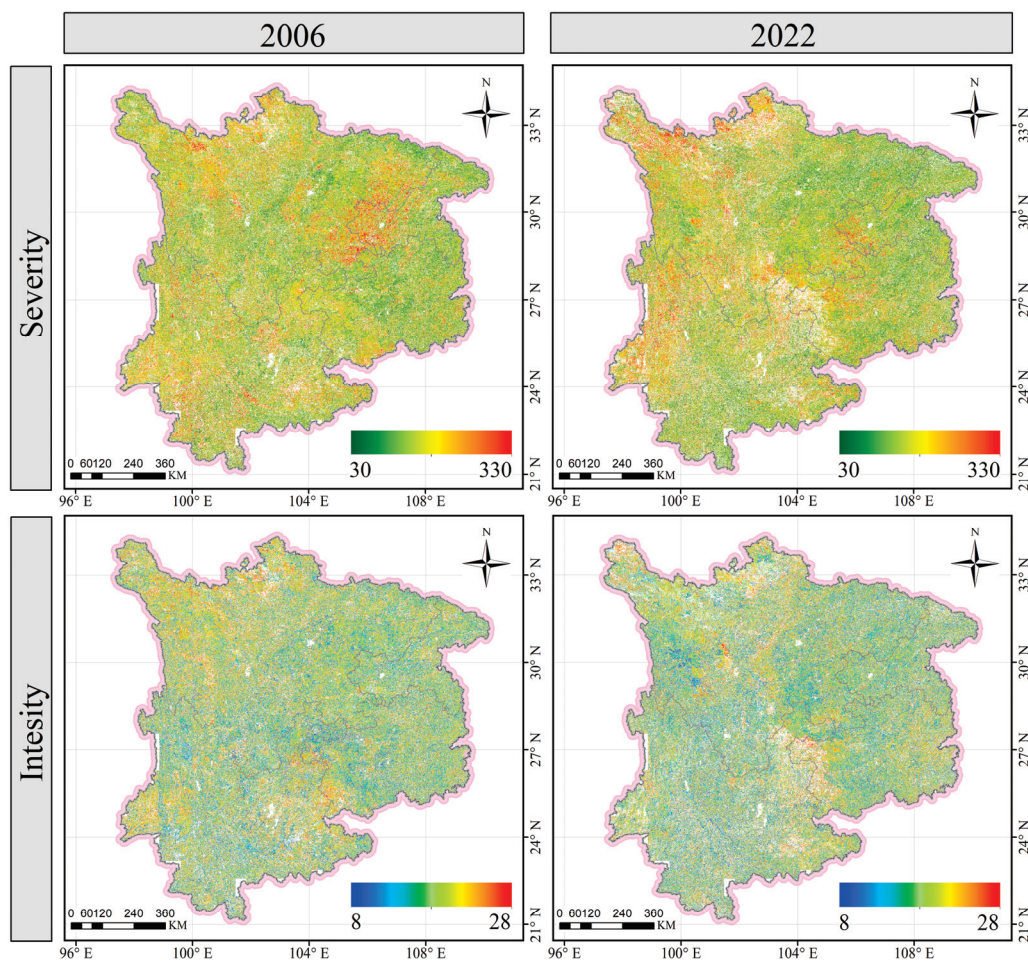


Figure 3. The methodological framework in this study.

### 3. Results

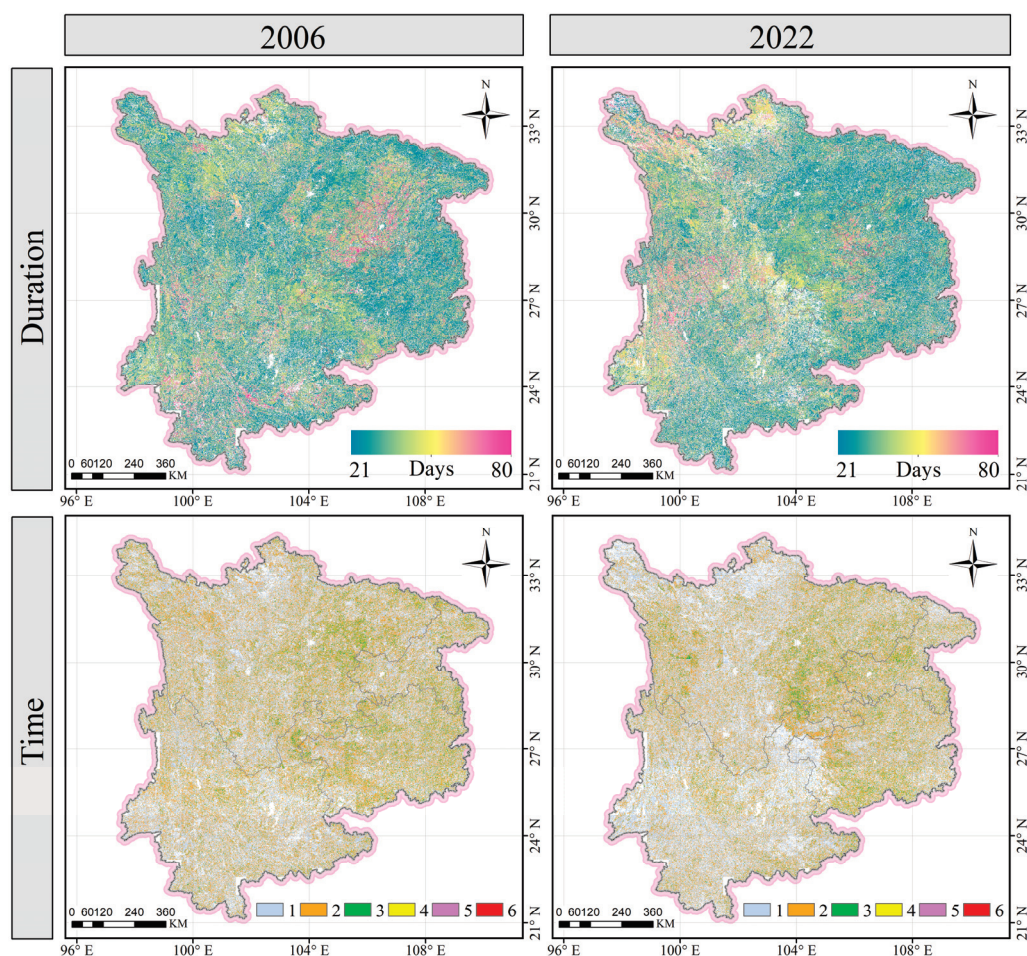
#### 3.1. Measuring Properties of Flash Drought Events

The typical flash drought years 2006 and 2022 in Southwestern China were selected. The mean drought severity (DS) and DI in 2006 and 2022 are shown in Figure 4. The DS in 2006 showed regional variability in Southwest China. Clusters of higher DS occurred in Western Chongqing and Eastern Sichuan, whereas lower DS was observed in the mountainous areas of western Sichuan. However, the dry hot valleys (Figure S2, Supplementary Materials) in Yunnan province exhibited high DS, especially in Yuanjiang. In contrast, DS was lower in 2022 in most regions, particularly in Western Chongqing, whereas the Northwest region experienced higher DS than in 2006. The DI in 2006 was higher in mountainous areas than in plain areas in the Sichuan Basin (Figure S2, Supplementary Materials). High DI was observed in western Sichuan and the border between Guizhou and Yunnan provinces. In 2022, the spatial distribution characteristics of DI were similar to those of 2006. The DI in the western mountainous areas was lower than in 2016. In general, the DS and DI were higher in 2006 than in 2022.



**Figure 4.** The mean severity and intensity of flash drought events in dry years.

The drought duration (DD) and drought time (DT) are depicted in Figure 5. Western Chongqing, Eastern Sichuan, and the dry hot valleys in Yunnan province experienced longer DDs in 2006 (approximately 80 days). The flash drought events in most of Southwest China lasted 21–50 days. The DD was lower in 2022 in most regions, especially in Western Chongqing and the dry hot valleys in Yunnan province. However, the DD in the western region was longer than in 2006. In dry years, the DT was significantly larger in the eastern region than in the western regions. A relatively high frequency of flash drought events was observed in the Sichuan Basin. Yunnan province experienced lower drought frequency than other provinces. Furthermore, the DT in Chongqing was higher in 2022 than in 2006. An opposite spatial distribution was observed for the DT and drought interval (DIV) in Southwest China (Figure S3, Supplementary Materials). A shorter DIV occurred in 2006 than in 2022 in the Sichuan Basin and the dry hot valleys in Yunnan province. The latter experienced a higher frequency of flash drought events. However, many regions lacked data, and only a single flash drought occurred in these areas. Overall, higher frequency and shorter duration drought events occurred in 2022, whereas fewer drought events with longer durations were observed in 2006, particularly in the Sichuan Basin. Furthermore, the plain areas experienced a higher frequency of flash drought than mountainous regions.

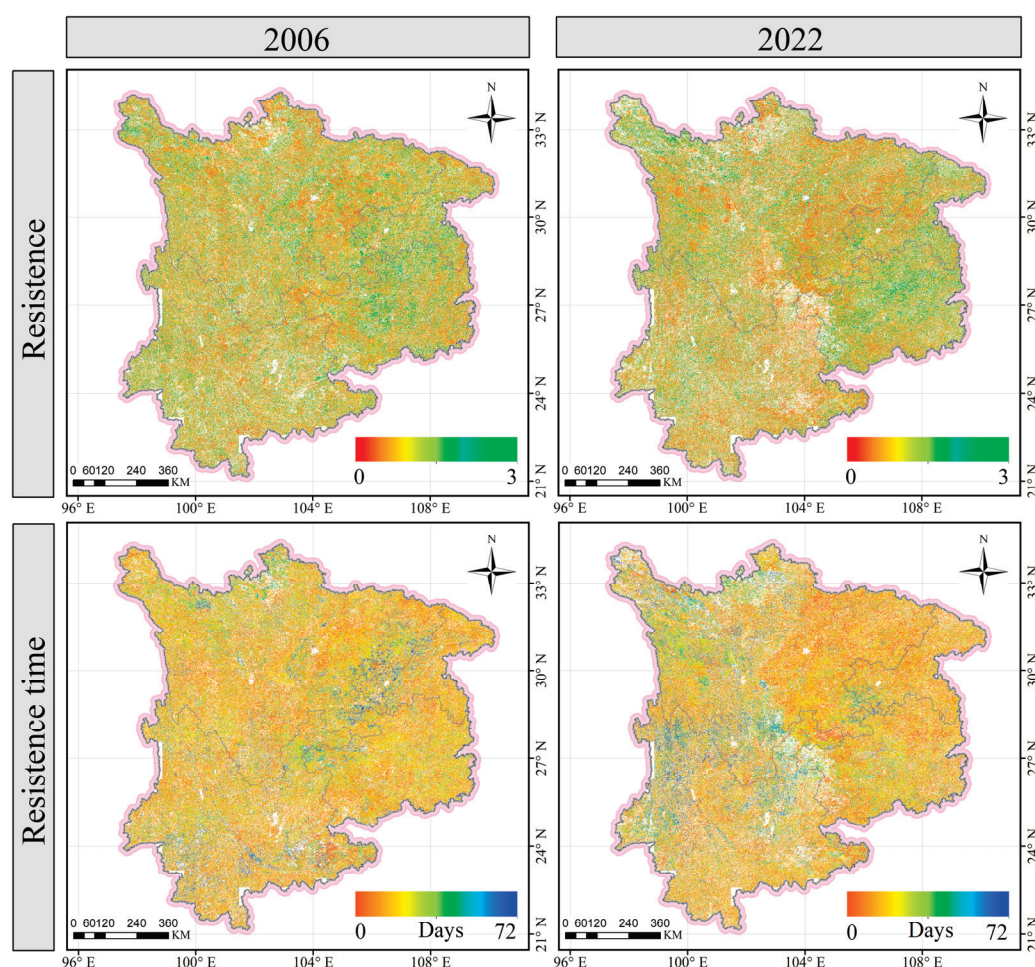


**Figure 5.** The average duration and times of flash drought events affecting vegetation in dry years.

### 3.2. Evaluating Vegetation Resilience and Vulnerability to Flash Drought Events

Most regions showed low vegetation resistance and a short resistance duration (Figure 6), indicating a limited ability of vegetation to withstand flash drought events. The LAI rapidly declined during the resistance period. Vegetation resistance to flash drought events was higher in the eastern region than in the western area in 2006 (Figure 6). Higher vegetation resistance was observed in Guizhou province, and lower vegetation resistance occurred in most of the Sichuan Basin and the dry hot valleys in Yunnan province. Furthermore, we found that mountainous areas around the Sichuan Basin had high vegetation resistance. The spatial patterns of vegetation resistance were similar in 2022 and in 2006. Vegetation resistance in Guizhou province was higher in 2022 than in 2006. Longer vegetation resistance times were observed in 2006 in Western Chongqing and the dry hot valleys in Yunnan province. Notably, the vegetation resistance time in mountainous areas in 2022 exceeded that in 2006, especially in the western region of Sichuan. This result suggests that the vegetation in these regions can endure prolonged cumulative flash drought events. However, the vegetation resistance in the Sichuan Basin had a shorter duration in 2022 than in 2006. This phenomenon cannot be ignored. Because rainfed croplands are concentrated in the Sichuan Basin, they rely primarily on rainfall for the water supply. Severe flash droughts were observed in this region, significantly affecting rainfed crop production [54].



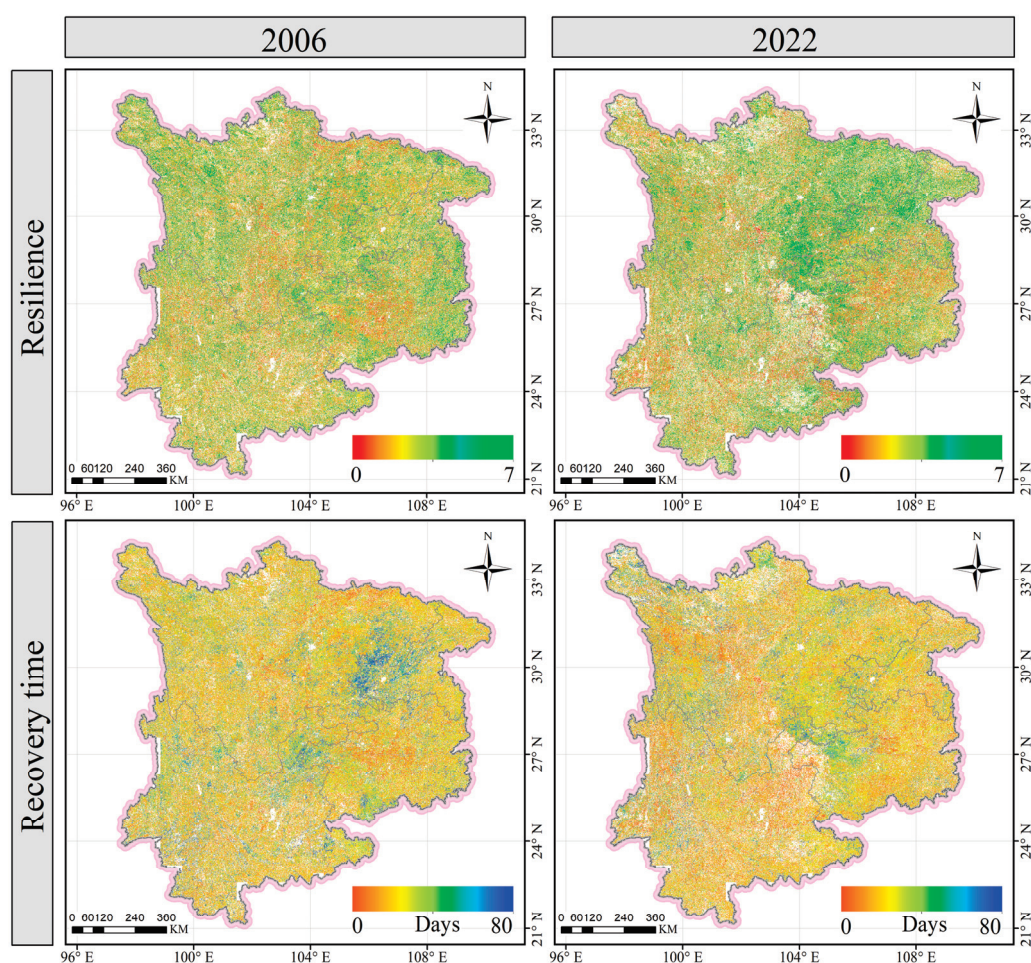


**Figure 6.** Vegetation resistance and resistance time to flash droughts in dry years. The value of vegetation resistance indicates the ability of vegetation to withstand flash drought events. The value of vegetation resistance time is the duration from the onset of a flash drought event to the day when vegetation experiences the most severe impact.

In 2006, a higher 50th percentile of vegetation resistance was observed in shrublands than in croplands (Figure S4, Supplementary Materials). The 75th percentile of vegetation resistance values was markedly higher in shrublands than in other vegetation types. The magnitude of the vegetation resistance was similar in 2022 and 2006. However, a higher 50th percentile of vegetation resistance occurred in grasslands. The variability of the vegetation resistance time for different vegetation types was small in 2006. The 75th percentile vegetation resistance time was higher in grasslands and shrublands than in other vegetation types. In contrast, notable differences in the vegetation resistance time of different vegetation types were observed in 2022. Grasslands showed longer vegetation resistance times, and croplands displayed shorter ones. In general, natural vegetation exhibited high resistance with a long resistance time, whereas croplands displayed low resistance with a short resistance time. This phenomenon was attributed to the prevalence of stress-tolerant natural vegetation, which exhibits various response mechanisms to drought events, including the capacity to access water from deeper soil layers [10].

Higher vegetation resilience to flash droughts was observed in the Sichuan Basin, reflecting a greater capacity for vegetation recovery. In contrast, lower vegetation resilience occurred in Yunnan province in 2006, reflecting a reduced capacity for recovery (Figure 7). Vegetation resilience was higher in 2022 in most regions, especially in the Sichuan Basin.

Notably, the majority of low vegetation resilience areas occurred in Guizhou province. The area has significant rock desertification. Intense surface erosion and subterranean leakage in Guizhou province have decreased the soil depth, exposed bedrock extensively, and caused frequent instances of rocky desertification [55]. Most regions showed a short recovery time (Figure 7), indicating a stronger capacity for vegetation recovery after flash drought events. The LAI could increase rapidly during the recovery period. A longer vegetation recovery time was observed in 2006 in Western Chongqing and the dry, hot valleys in Yunnan province. Most mountainous areas in the western region experienced short recovery times. The vegetation recovery time was much lower in 2022 in most regions, particularly in Western Chongqing and the dry, hot valleys in Yunnan province. Furthermore, mountainous areas exhibited shorter vegetation recovery periods than plain areas. This finding implies that vegetation in the plain regions was more vulnerable to the impact of flash droughts, especially in 2006.



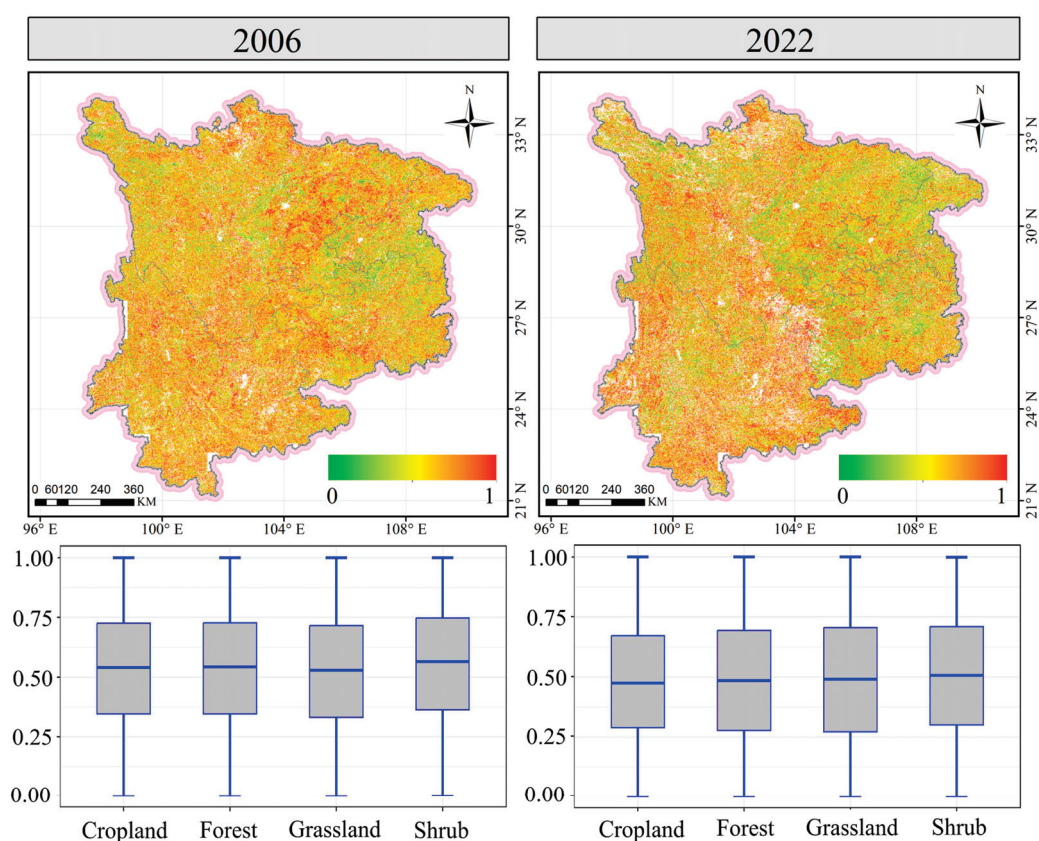
**Figure 7.** Vegetation resilience and recovery time in dry years. The value of vegetation resilience indicates a capacity for vegetation recovery after flash droughts. The value of vegetation time is the period from the day of the most severe impact to the day when the vegetation reaches the most favorable condition.

In 2006, a higher 50th percentile of vegetation resilience occurred in grasslands than in other vegetation types, and a lower one was observed in shrublands (Figure S5, Supplementary Materials). The magnitudes were similar for forests and croplands. However, croplands displayed a higher 50th percentile of vegetation resilience than other vegeta-



tion types in 2022. The lowest 50th percentile of vegetation resilience was observed in grasslands. Slight variations in the vegetation recovery time were observed for different vegetation types in dry years. The 50th and 75th percentiles in croplands in 2006 exceeded those in other vegetation types, but the variability was low for other vegetation types. In 2022, grasslands exhibited a higher 50th percentile of vegetation recovery time than shrubs and croplands. The lowest 50th percentile was observed in forests. Thus, differences occurred in the response patterns of vegetation to different flash droughts. The response patterns were more pronounced in 2006 than in 2022.

Most regions experienced high vegetation vulnerability in 2006 (Figure 8), especially in the Sichuan Basin and Yunnan province, suggesting that the vegetation in this region is highly prone to changes during flash drought events [56], whereas low vegetation vulnerability values were concentrated in the southern part of Chongqing province and the northern part of Guizhou province. In contrast, vegetation vulnerability was low in 2022 in most regions. Flash droughts did not significantly affect vegetation in the Sichuan Basin, Chongqing province, and Yunnan province. However, high vegetation vulnerability was identified in the central region of the Sichuan Basin. No flash drought events occurred in some regions of the western mountainous areas. Overall, the vegetation was more vulnerable in the western mountainous areas than the eastern regions.



**Figure 8.** Vegetation vulnerability to flash droughts. The boxplots display the ranges of vegetation vulnerability. The 50th percentile is the median value, representing the value at which 50% of the observed vegetation resilience values fall below and 50% fall above. The percentiles were calculated by using all pixel values.

The vegetation vulnerability values at the 50th percentile were noticeably higher in shrublands than in the other vegetation types (Figure 8). The 50th percentiles are the median value and were calculated using all pixel values in the study area. Grasslands

exhibited a lower 50th percentile of vegetation vulnerability than other vegetation types. The magnitudes of vegetation vulnerability were similar in croplands and forests in 2006. In contrast, the vegetation vulnerability values were lower in 2022 than in 2006 in most vegetation types, indicating that the impact of flash droughts on vegetation was higher in 2006 than in 2022. The degree of vegetation vulnerability in 2022 was higher in shrublands than in croplands, indicating that the vulnerability of vegetation increased as the vegetation cover decreased. Additionally, cropland vegetation exhibited lower vulnerability to flash drought events than natural vegetation.

Figure S6 depicts the estimated coefficients obtained from spatial convergence analysis with the GWR model in the administrative regions. The majority of Southwestern China exhibited negative coefficients of vegetation vulnerability, indicating convergence from 2006 to 2022. Negative coefficients of vegetation vulnerability were detected in most regions in Yunnan and Guizhou provinces, with a high adjusted  $R^2$  (Figure S6, Supplementary Materials). This result indicated that the vegetation vulnerability exhibited convergence in these areas. The Sichuan Basin showed high negative coefficients with high adjusted  $R^2$  values (exceeding 0.6). This area experiences a convergence of vegetation vulnerability. In contrast, positive coefficients of vegetation vulnerability occurred in the northwestern region of Sichuan, suggesting divergence. Vegetation vulnerability to flash drought events generally showed convergence in most regions, and more so in the Sichuan Basin. However, divergence was more common in northwestern Sichuan province.

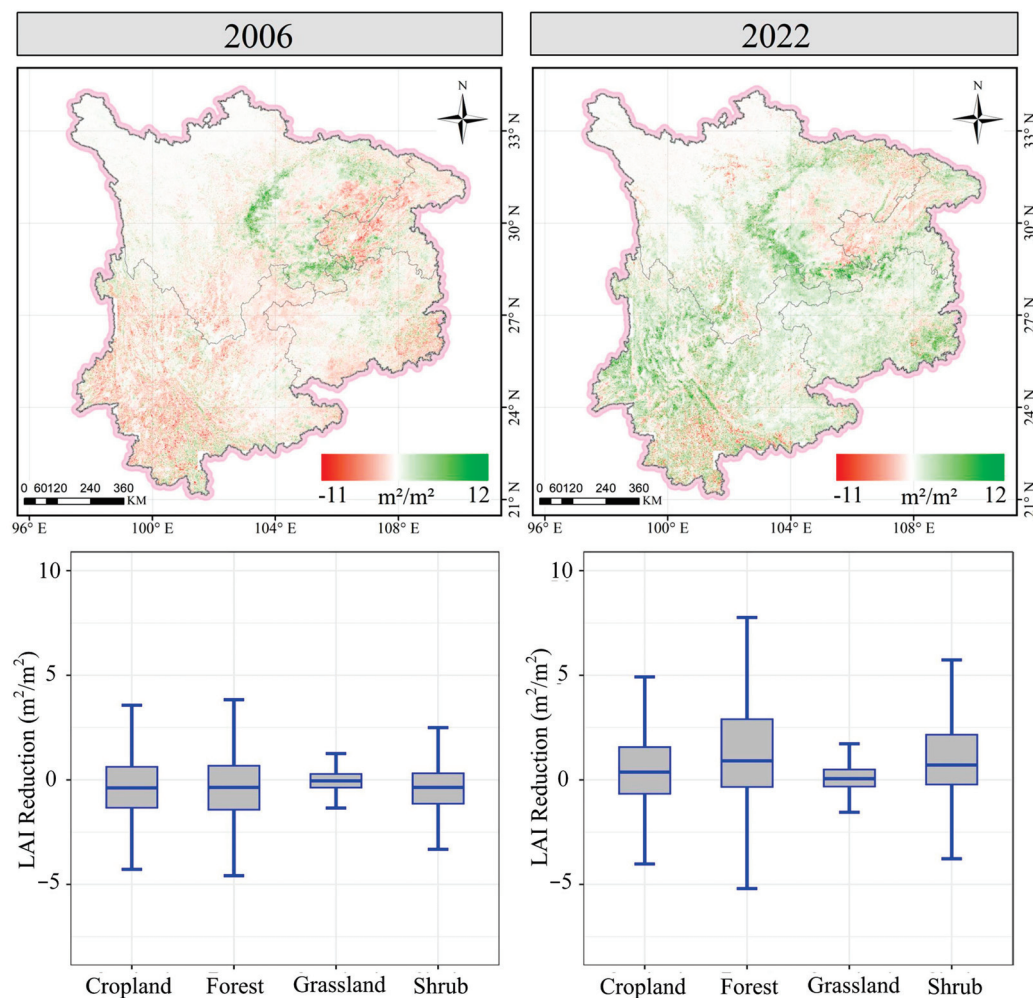
### 3.3. Evaluating Vegetation Anomaly During and After Flash Drought Events

#### 3.3.1. Vegetation Loss During Flash Drought Events

LAI changes during flash drought events were monitored in dry years, as shown in Figure 9. Most regions experienced vegetation loss due to flash drought events in 2006, particularly in Chongqing City and Yunnan province. Small increases in LAI occurred in the mountainous areas surrounding the Sichuan Basin. In 2022, the impact on vegetation had weakened in the Yunnan and Guizhou provinces. Vegetation loss was identified in the central region of the Sichuan Basin, especially in western Chongqing. LAI reduction in these regions was lower than in 2006. Stable areas during flash drought events occurred in the western mountainous areas. The ranges of vegetation change during flash drought are illustrated in Figure 9. A lower 50th percentile of vegetation loss in 2006 was observed in croplands and forests. Furthermore, croplands and forests exhibited higher percentages of vegetation loss, followed by shrublands. Most grasslands were concentrated in the western area, which had low vegetation vulnerability. The magnitudes of LAI changes in this region were small. In contrast, the vegetation loss was lower in 2022 than in 2006 in most vegetation types. The impact of flash drought on croplands was more severe than on other vegetation types. In general, vegetation loss was more significant in 2006 than in 2022. Additionally, croplands and forests exhibited a significant decrease in LAI in 2006 due to flash drought events.

The BRT model was used to determine the importance of flash drought properties on vegetation loss (Figure S7, Supplementary Materials). The model revealed that flash DD and DIV had the most significant impact on vegetation loss in different vegetation types, contributing to 60% of the overall impact. In croplands, DD (57.47%) exerted the strongest influence on vegetation loss, followed by DIV (28.01%). DI, DS, and DT accounted for 14.52% of the explained vegetation change during flash droughts. The effect of DD on vegetation declined over time, whereas the effects of DI and DS exhibited an increase. DI and DS contributed to vegetation loss in forests (19.30%), grasslands (23.19%), and shrubland (30.47%). DT (24.37%) had the most significant impact on vegetation loss in

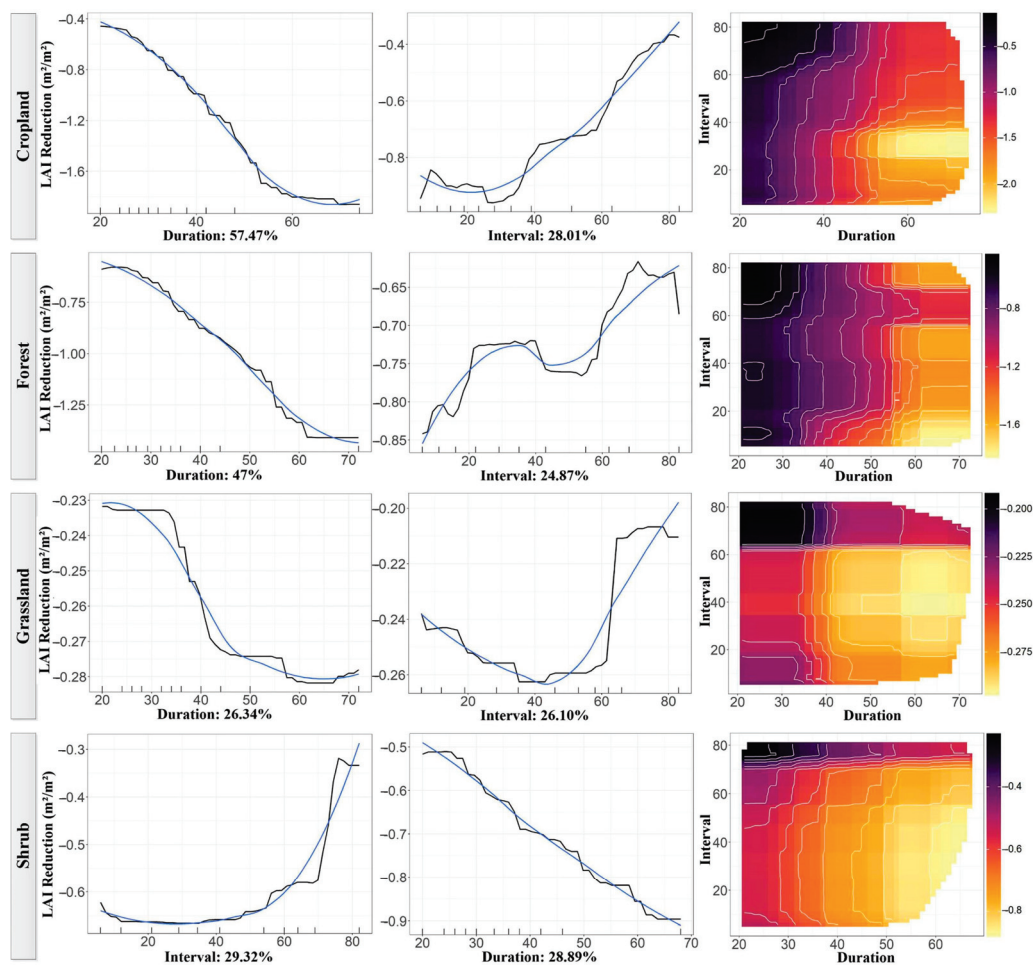
grasslands. In general, DD and DIV were the dominant factors impacting vegetation loss during flash droughts, especially in croplands. Because many rainfed croplands, mainly located in the Sichuan Basin, are highly susceptible to flash droughts. The crop yield in this area is highly dependent on weather conditions, particularly during flash drought events [54].



**Figure 9.** Vegetation loss during flash drought events. The boxplots display the ranges and magnitudes of vegetation loss.

The two most important indicators were selected to investigate the influence of flash drought properties on vegetation loss, as shown in the partial dependence and joint pairwise interaction plots (Figure 10). As the length of the flash drought increased, the decrease in the LAI increased in all vegetation types, particularly for flash droughts lasting less than 50 days. Moreover, the reductions were larger in croplands and forests than in grasslands and shrublands. In contrast, as the interval of the flash drought increased, the reduction in LAI decreased in most vegetation types. The LAI reduction rapidly decreased when the interval of flash droughts exceeded 60 days. The joint pairwise interaction plots for croplands showed that vegetation experienced a significant decrease in LAI in regions where the DD exceeded 60 days, and the DIV ranged from 30 to 40 days. Forest areas with DDs exceeding 60 days and a DIV below 20 days experienced high reductions in LAI. Grasslands exhibited higher decreases in LAI in regions where the DD exceeded 50 days and the DIV was around 40 days. Shrubland saw higher decreases in LAI in regions

where the DD exceeded 50 days, and the DIV was below 40 days. Overall, significant LAI reductions occurred in Southwestern China in areas where the DD exceeded 60 days, and the DIV was below 40.



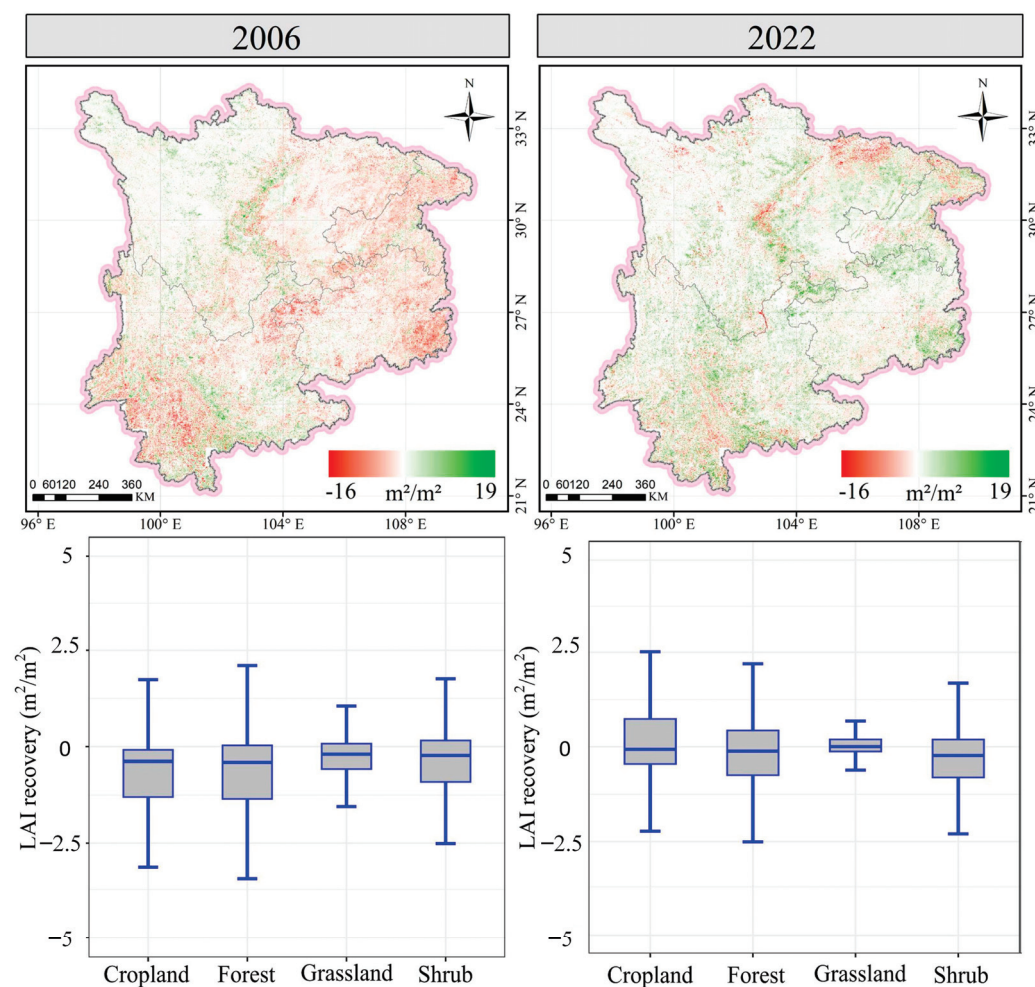
**Figure 10.** Partial dependence plots illustrating the impact of the two most important explanatory factors on vegetation loss in the BRT model. The blue lines indicate LOESS smooth fitting. The vertical axis shows the marginal effect in the model. The rug plots on the horizontal axis illustrate the distribution of the driving factors.

### 3.3.2. Vegetation Recovery After Flash Drought Events

Figure 11 displays the state of vegetation recovery following flash drought events. Vegetation in most regions did not recover after flash drought events in 2006, particularly in Chongqing City and Southwest Yunnan province. Some vegetation was able to recover following flash drought events in Western Sichuan province. In 2022, most regions can recover after flash drought events. However, some vegetation in the Sichuan Basin and the dry hot valleys in Yunnan province did not recover and continued to be impacted by flash droughts. The ranges in vegetation changes following flash drought events are illustrated in Figure 11. A lower 50th percentile of vegetation loss after flash droughts in 2006 was observed in croplands and forests. Moreover, these land types showed higher overall percentages of vegetation loss compared to grasslands, which were less affected by the drought conditions. In contrast, the decrease in vegetation following flash droughts was less pronounced in 2022 than in 2006 across most vegetation types. However, the impact of flash droughts was more severe on forests and shrubs compared to other vegetation types.

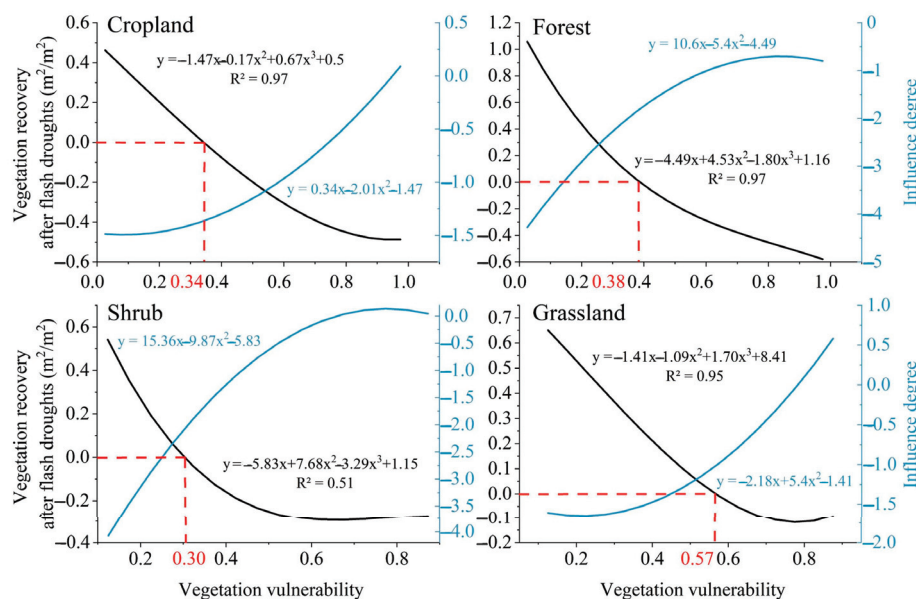


Overall, vegetation in most regions did not fully recover after the flash drought events of 2006. Additionally, croplands and forests exhibited low restoration in 2006.



**Figure 11.** Vegetation recovery after flash drought events. The boxplots display the ranges and magnitudes of vegetation recovery.

The most significant impacts of flash drought on vegetation were analyzed at the pixel level for the years 2006 and 2022. Thresholds for vegetation vulnerability, crucial for recovery after these droughts, were identified for different vegetation types, as shown in Figure 12. As vegetation vulnerability increased, vegetation recovery weakened across all vegetation types. Specifically, croplands and shrubs can recover when their resilience fell below the thresholds of 0.34 and 0.30, respectively. In contrast, when the vegetation vulnerability was less than the threshold of 0.38 and 0.57 for grasslands and forests, respectively, these vegetation types could recover after flash drought events. When vegetation vulnerability was greater than the threshold, vegetation cannot recover after flash drought events. The relationship between vegetation vulnerability and restoration follows a distinct pattern. The impact increases for forests and shrubs rapidly at first and then slows down. In contrast, croplands and grasslands exhibit the opposite trend, with the impact increasing slowly at the beginning and then accelerating.



**Figure 12.** The thresholds of vegetation vulnerability for vegetation recovery after flash droughts in different vegetation types. The Y-axis right coordinate represents the influence degree of vegetation resilience on vegetation restoration. The thresholds of vegetation vulnerability (red value) were determined based on the value of 0 in vegetation recovery following flash droughts. If vegetation vulnerability exceeds the threshold, vegetation cannot recover after flash droughts.

## 4. Discussion

### 4.1. Vegetation Resistance to Flash Drought Events

The number and frequency of flash droughts have significantly increased since 2000, largely due to the widespread and simultaneous occurrence of extreme heat and dryness due to global warming [57,58]. Higher temperatures amplify these compound extremes by increasing the atmospheric demand for water significantly, potentially causing dry conditions even if precipitation remains the same or increases slightly [59,60]. Anthropogenic warming, caused by higher concentrations of greenhouse gases, has heightened the variability in precipitation and temperature by increasing the upward trend in flash drought frequency [3]. Furthermore, substantial increases in flash drought risks have occurred in Southwestern China due to concurrent hot and dry conditions [3], while flash drought risks was decreased in northwest China [3]. Southwestern China experienced extreme flash drought events in 2006 and 2022, causing extensive damage to vegetation during critical vegetation growth stages [60,61]. We found that the flash DS, DI, and DD were higher in 2006 than in 2022 (Figures 4 and 5). During the summer, the circulation pattern in subtropical and midlatitude regions does not favor the convergence of cold–dry air from the north and warm, moist air from the south, leading to severe flash droughts [62]. Negative precipitation anomalies can reduce soil moisture, decreasing evapotranspiration, which in turn causes higher temperatures and vapor pressure deficits (VPDs) due to increased sensible heat flux and lower near-surface humidity. However, an elevated VPD increases the atmospheric evaporative demand, further accelerating the depletion of soil moisture. These land–atmosphere feedbacks strengthen due to global warming and can lead to concurrent anomalies in precipitation, temperature, and VPDs [63,64]. The combination of a severe drought and elevated summer temperatures in 2006 led to water scarcity, affecting at least 18 million residents. Crop failures occurred on 311,300 hectares, resulting in economic losses of CNY 11.74 billion [62]. Furthermore, the Sichuan Basin and dry hot valleys in Yunnan province experienced more pronounced flash drought events

than other regions (Figures 4 and 5). Flash droughts in the Sichuan Basin result from complex interactions between local precipitation patterns and water from nearby rivers. Additionally, the impact of the Qinghai-Tibet Plateau and the Qinling Mountains create obstacles to air circulation internally and externally. This phenomenon causes the southeast monsoon to descend over central China [65,66]. A higher drought frequency in the dry hot valleys in Yunnan province is attributed to the topography that significantly impacts the spatial distribution of drought events [67]. Inversions occur in the dry, hot valleys in this subtropical climate. Drought events are likely in the dry, hot valleys, including the basins of the Jinsha River, Lancang River, Nujiang River, Yuanjiang River, and Nanpan River [68]. Thus, flash drought events in these regions should be a primary focus.

Vegetation resistance to flash drought events varied in different areas. Vegetation in the Sichuan Basin exhibited low resistance and shorter durations of resistance (Figure 6). Rainfed croplands are concentrated in the Sichuan Basin. They rely primarily on rainfall for the water supply. Severe flash droughts were observed in this region, significantly affecting rainfed crop production [54]. Moreover, the occurrences of severe and extreme drought in this area are projected to rise significantly in the future, which will cause an imbalance in the water demand and supply and less water availability [69,70]. These conditions will exacerbate the existing water shortages in rainfed croplands in the Sichuan Basin. In contrast, most irrigated croplands are located in the western region of the Sichuan Basin. Irrigation agriculture is less prone to drought impacts [71]. Consequently, the vegetation resistance was higher in these areas than in rainfed croplands, especially in 2022. The additional water provided by irrigation increases vegetation resistance in irrigated agricultural lands. The result is consistent with a previous study, which found that rainfed croplands in Northern Kazakhstan had higher resistance than irrigated agriculture in the Amu Darya delta [43].

Grasslands and forests in mountainous areas exhibited high resistance to flash drought events and had longer durations of resistance (Figure 6 and Figure S4). This phenomenon was attributed to the prevalence of stress-tolerant natural vegetation, which exhibits various response mechanisms to drought events, including the capacity to access water from deeper soil layers [10]. Furthermore, higher forest canopy heights are related to higher drought resistance. Forests with taller trees exhibited higher vulnerability and susceptibility to growth reduction than those with shorter trees during severe drought conditions. In contrast, the latter displayed higher drought resistance [72]. Larger and taller trees have lower drought resistance than smaller and shorter trees due to their higher ET rates and water demand. A water deficit reduces the transpiration rate, and the force to transport water is insufficient, resulting in lower drought resistance [73]. Thus, forests in Guizhou province with shorter trees displayed higher vegetation resistance to flash drought events (Figure 6). However, forests and grasslands exhibited longer resistance durations than other types of vegetation in Southwestern China (Figure S4, Supplementary Materials). This result is attributed to the ability of forest soils to retain large quantities of water that is gradually released over a prolonged period, alleviating the immediate impacts of climate change.

#### *4.2. Vegetation Resilience to Flash Drought Events*

Rainfed croplands in the Sichuan Basin exhibited high vegetation resilience due to the extensive cultivation of rainfed rice, particularly in 2022 (Figure 7). The negative effects on vegetation in rainfed croplands were more pronounced during the drought period [74]. Flash drought events significantly affected rainfed rice, resulting in a considerable reduction in crop yield (rice, wheat, corn, etc.) [75,76]. Moreover, rainfed croplands have a higher

potential for rapid recovery due to re-wetting following flash drought events, either through seed germination or rhizome regeneration. Nevertheless, the degree of recovery depends on the availability of post-drought water [77]. Rainfed croplands demonstrated a higher recovery ability after drought events. However, rainfed croplands can self-regulate in response to drought disturbances. When drought occurred in 2006, the self-regulation ability of rainfed croplands was insufficient, and they could not recover [78]. Thus, rainfed croplands in the Sichuan Basin exhibited lower vegetation resilience and longer recovery times in 2006 than in 2022 (Figure 7 and Figure S5). Conversely, irrigated croplands in the western Sichuan Basin exhibited low vegetation resilience to flash drought events (Figure 7). Irrigated croplands typically require a substantial and consistent water supply, and a deficiency in irrigation during severe droughts causes water stress because some species lack adaptive strategies. Irrigated croplands are more susceptible to flash droughts than rainfed croplands due to a lower ability to buffer against high DI [77]. The result is consistent with a previous study. We found that irrigated agriculture in arid areas had higher resilience than rainfed croplands [43].

Low vegetation resilience occurred in Guizhou province in drought years (Figure 7). The area has significant rock desertification, resulting in ecological damage and poverty. Intense surface erosion and subterranean leakage in Guizhou province have decreased the soil depth, exposed bedrock extensively, and caused frequent instances of rocky desertification [55]. The soil depth of soil significantly affects vegetation resilience in karst regions, particularly when the soil depth is less than 25 cm [55]. Moreover, ecological restoration projects, the Karst Rocky Desertification Comprehensive Control, and the Grain to Green Project have been implemented in Guizhou province since the early 2000s [79]. The vegetation resilience in southwestern Guizhou was higher in 2022 than in 2006 (Figure 7). However, the northern region did not experience a significant improvement in vegetation resilience. Afforestation is the primary approach to ecological restoration in this region. The structure and function of the trees are unstable until they mature [80]. Meanwhile, most afforestation projects in this region have utilized a single tree species, resulting in a uniform forest structure [55]. As a result, the afforested areas were more prone to drought than natural forests [81,82]. Thus, we suggest the conventional afforestation approach in karst regions, which aims to increase forest cover, should be refined to improve vegetation resilience. In addition, species richness could be enriched to enhance vegetation resilience in this region [83]. Shrublands in Southwestern China in 2006 exhibited high resistance but low resilience to flash drought events (Figures 6 and 9). Shrublands have marginal stability and are susceptible to transitioning into desert conditions in response to drought events [84]. As a result, shrublands took longer to recover to the average state following a drought and demonstrated lower resilience than other types of vegetation (Figure S4, Supplementary Materials). However, the lower intensity of the flash drought in 2022 than in 2006 resulted in a reversal of the phenomenon observed in the vegetation resilience of shrublands.

#### *4.3. Vegetation Vulnerability and Loss During Flash Drought Events*

Spatial variations in vegetation vulnerability were detected in different vegetation types (Figure 8). The Sichuan Basin had high vegetation vulnerability in 2006, suggesting that the vegetation in this region is highly prone to changes during flash drought events [56]. More prolonged and severe flash drought events were observed in this basin than in other areas (Figures 4 and 5). Rainfed croplands, mainly located in the Sichuan Basin, are highly susceptible to flash droughts. The crop yield in this area is highly dependent on weather conditions, particularly during flash drought events, such as in 2006. Crop



failures occurred on approximately 0.3 million hectares, and 18 million people experienced water scarcity [85,86]. Thus, a large decrease in LAI during flash droughts was observed in this region (Figure 9). Furthermore, vegetation vulnerability to flash drought events showed convergence in the region (Figure S6, Supplementary Materials), suggesting an increase in vegetation vulnerability in most of the Sichuan Basin in the future. The spatial convergence pattern observed in this area can be regarded as an early indicator of vegetation vulnerability [44,46]. Most croplands in Southwestern China are located in the Sichuan Basin, where flash droughts affect agricultural land, leading to direct economic losses. We suggest reducing rainfed farmland and increasing irrigated farmland in areas of high vulnerability the central region of the Sichuan Basin, especially in Western Chongqing. In addition, this area can cultivate crops with low agricultural water demand and high drought tolerance.

Forests in the southern part exhibited high vegetation vulnerability to flash droughts and vegetation loss, especially in Yunnan province (Figures 8 and 9). The reason is that Yunnan province frequently experiences droughts from winter to spring, and their influences on the annual gross primary productivity fluctuations are more pronounced than during summer droughts [87]. Furthermore, the vegetation phenology in Yunnan province is prone to changes during flash droughts, especially at the start and end of the growing season, significantly decreasing gross primary productivity [87]. Low vegetation resistance and resilience were observed in Yunnan province, especially in the dry hot valleys. Areas exhibiting low vegetation resistance and resilience can act as an early warning, indicating the likelihood of degradation and underscoring the need for proactive conservation measures.

The DD and DIV were the dominant drivers affecting vegetation loss during flash droughts in Southwest China (Figure S7, Supplementary Materials). As the DD increased and the DIV decreased in dry years, the negative impact on vegetation intensified, significantly reducing vegetation productivity [88]. Increasing evidence suggests frequent droughts disrupt vegetation stability, hasten environmental shifts, and diminish vegetation resilience [89]. In drought years with high DDs and low DIVs, the vegetation could not recover from damage, potentially causing vegetation mortality [90]. Furthermore, flash droughts have become significantly more frequent and tend to occur more rapidly. They will likely last longer in the future. The primary reason is the increased frequency of concurrent short-period hot and extreme dry events in a warming climate [57]. Thus, a high flash drought frequency increases the likelihood of vegetation damage and decreases vegetation productivity. Moreover, the degree of damage differed for different vegetation types due to unique adaptation strategies. We found that croplands experienced a significant decrease in LAI in regions where the DD exceeded 60 days, and the DIV ranged from 30 to 40 days. In contrast, forest areas with DDs exceeding 60 days and DIVs of less than 20 days experienced high reductions in LAI (Figure 10). Importantly, as the vegetation cover decreases, the impact of the DS and DI on vegetation susceptibility becomes more pronounced. Shrublands and grasslands exhibited higher susceptibility than forests and croplands to the severity and intensity of flash droughts (Figure S7, Supplementary Materials). In contrast, sparse vegetation in arid areas showed lower susceptibility to drought [43], because it has evolved physiological adaptations to survive in drought-prone environments. These adaptations involve enhanced water uptake and mechanisms to minimize water loss [91].

In this study, remote sensing products, including MODIS data, soil moisture, and land use information, were utilized. The limitations of these products may influence this study's results. Because these remote sensing products have been verified as having high accuracy in the previous studies [29–31], the research results remain within an acceptable error range and exhibit a high level of reliability.

## 5. Conclusions

The objectives of this study focused on how the severity of flash droughts and vegetation vulnerability hinder recovery to the original state, resulting in losses. The significant impact of flash drought properties on vegetation loss were revealed in this study, and the thresholds of vegetation vulnerability were determined for vegetation recovery. The main findings are as follows:

- (1) The Sichuan Basin and the dry, hot valleys in Yunnan province experienced more pronounced flash drought events than other regions. Croplands in the Sichuan Basin exhibited low resistance to flash drought events and shorter resistance periods, whereas grasslands and forests in mountainous areas had high resistance to flash drought events and longer resistance periods.
- (2) A divergent pattern was observed for vegetation resilience and resistance to flash droughts. Croplands in the Sichuan Basin demonstrated high vegetation resilience to flash drought events, whereas those in Guizhou province exhibited low vegetation resilience. Afforested areas in Guizhou province were more susceptible to flash drought events than natural forests.
- (3) Most regions experienced high vegetation vulnerability to flash drought events, especially in the Sichuan Basin and Yunnan province. Croplands and forests in 2006 exhibited a significant decrease in LAI to flash drought events. We suggest reducing rainfed farmland and increasing irrigated farmland in areas of high vulnerability the central region of the Sichuan Basin, especially in western Chongqing.
- (4) Croplands experienced a significant decrease in LAI in regions where the DD exceeded 60 days, and the DIV ranged from 30 to 40 days. Forest regions with DDs exceeding 60 days and DIVs of less than 20 days experienced high reductions in LAI. Grasslands and shrublands saw higher decreases in LAI in regions where the DD exceeded 50 days, and the DIV was less than 40 days. Furthermore, croplands and shrubs could recover once their vulnerability fell below thresholds of 0.34 and 0.30, respectively. In contrast, when the vegetation vulnerability was less than the threshold of 0.38 and 0.57 for grasslands and forests, respectively, these vegetation types could recover after flash drought events.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/w17050653/s1>, Figure S1. Distribution map of administrative subdivisions in Southwestern China; Figure S2. Topographic map with specific regions in Southwestern China; Figure S3. Spatial patterns of the average interval of flash drought events affecting vegetation in dry years; Figure S4. Boxplots of vegetation resistance and resistance time in dry years. The 50th percentile is the median value, representing the value at which 50% of the observed vegetation resilience values fall below and 50% fall above. The percentiles were calculated by using all pixel values; Figure S5. Boxplots of vegetation resilience and recovery time in dry years. The 50th percentile is the median value, representing the value at which 50% of the observed vegetation resilience values fall below and 50% fall above. The percentiles were calculated by using all pixel values; Figure S6. Spatial convergence of vegetation vulnerability in Southwestern China; Figure S7. The relative importance of drought properties affecting vegetation loss during flash droughts; Table S1. Attribute information for different datasets; Table S2. Classification of land cover types provided by the CLCD dataset.

**Author Contributions:** Conceptualization, L.J. and X.L.; methodology, L.J. and G.W.; software, L.J. and Q.L.; validation, X.L. and G.W.; formal analysis, L.J. and Q.L.; investigation, Q.L. and G.W.; resources, L.J. and X.L.; data curation, L.J., Q.L. and G.W.; writing—original draft preparation, L.J. and G.W.; writing—review and editing, L.J. and X.L.; visualization, L.J., Q.L. and G.W.; supervision,

L.J. and X.L.; project administration, L.J. and X.L.; funding acquisition, L.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the China Postdoctoral Science Foundation [Grant No. 2023M740433], the National Natural Science Foundation of China [Grant No. 42201124], the Innovation and Development Joint Fund of Chongqing Natural Science Foundation [Grant No. CSTB2024NSCQ-LZX0064], and the Chongqing Entrepreneurship and Innovation Support Program for Overseas Chinese [Grant No. 2204012976986965].

**Data Availability Statement:** The data supporting the plots within this paper and other study findings are available from the corresponding author upon reasonable request.

**Acknowledgments:** We would like to thank NASA for providing the Modis dataset and express our gratitude to the National Tibetan Plateau/Third Pole Environment Data Center for sharing the soil moisture dataset. Their willingness to share these valuable resources has played a crucial role in driving forward our study.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Yuan, X.; Wang, Y.; Ji, P.; Wu, P.; Sheffield, J.; Otkin, J.A. A global transition to flash droughts under climate change. *Science* **2023**, *380*, 187–191. [CrossRef] [PubMed]
2. Sungmin, O.; Park, S.K. Flash drought drives rapid vegetation stress in arid regions in Europe. *Environ. Res. Lett.* **2023**, *18*, 014028.
3. Yuan, X.; Wang, L.; Wu, P.; Ji, P.; Sheffield, J.; Zhang, M. Anthropogenic shift towards higher risk of flash drought over China. *Nat. Commun.* **2019**, *10*, 4661. [CrossRef] [PubMed]
4. Seddon, A.W.R.; Macias-Fauria, M.; Long, P.R.; Benz, D.; Willis, K.J. Sensitivity of global terrestrial ecosystems to climate variability. *Nature* **2016**, *531*, 229–232. [CrossRef] [PubMed]
5. Nguyen, H.; Wheeler, M.C.; Otkin, J.A.; Cowan, T.; Frost, A.; Stone, R. Using the evaporative stress index to monitor flash drought in Australia. *Environ. Res. Lett.* **2019**, *14*, 064016. [CrossRef]
6. Yao, T.; Liu, S.; Hu, S.; Mo, X. Response of vegetation ecosystems to flash drought with solar-induced chlorophyll fluorescence over the Hai River Basin, China during 2001–2019. *J. Environ. Manag.* **2022**, *313*, 114947. [CrossRef] [PubMed]
7. Pendergrass, A.G.; Meehl, G.A.; Pulwarty, R.; Hobbins, M.; Hoell, A.; AghaKouchak, A.; Bonfils, C.J.; Gallant, A.J.; Hoerling, M.; Hoffmann, D. Flash droughts present a new challenge for subseasonal-to-seasonal prediction. *Nat. Clim. Change* **2020**, *10*, 191–199. [CrossRef]
8. Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems (SR2). 2017. Available online: [https://archive.ipcc.ch/report/srcl/pdf/sr2\\_background\\_report\\_final.pdf](https://archive.ipcc.ch/report/srcl/pdf/sr2_background_report_final.pdf) (accessed on 19 January 2025).
9. Wang, M.; Wang, S.; Zhao, J.; Ju, W.; Hao, Z. Global positive gross primary productivity extremes and climate contributions during 1982–2016. *Sci. Total Environ.* **2021**, *774*, 145703. [CrossRef] [PubMed]
10. Li, X.; Piao, S.; Wang, K.; Wang, X.; Wang, T.; Ciais, P.; Chen, A.; Lian, X.; Peng, S.; Penuelas, J. Temporal trade-off between gymnosperm resistance and resilience increases forest sensitivity to extreme drought. *Nat. Ecol. Evol.* **2020**, *4*, 1075–1083. [CrossRef] [PubMed]
11. Lin, Q.; Wu, Z.; Singh, V.P.; Sadeghi, S.; He, H.; Lu, G. Correlation between hydrological drought, climatic factors, reservoir operation, and vegetation cover in the Xijiang Basin, South China. *J. Hydrol.* **2017**, *549*, 512–524. [CrossRef]
12. Cao, S.; Zhang, L.; He, Y.; Zhang, Y.; Chen, Y.; Yao, S.; Yang, W.; Sun, Q. Effects and contributions of meteorological drought on agricultural drought under different climatic zones and vegetation types in Northwest China. *Sci. Total Environ.* **2022**, *821*, 153270. [CrossRef]
13. von Keyserlingk, J.; De Hoop, M.; Mayor, A.; Dekker, S.; Rietkerk, M.; Foerster, S. Resilience of vegetation to drought: Studying the effect of grazing in a Mediterranean rangeland using satellite time series. *Remote Sens. Environ.* **2021**, *255*, 112270. [CrossRef]
14. Horion, S.; Ivits, E.; De Keersmaecker, W.; Tagesson, T.; Vogt, J.; Fensholt, R. Mapping European ecosystem change types in response to land-use change, extreme climate events and land degradation. *Land Degrad. Dev.* **2019**, *30*, 951–963. [CrossRef]
15. Jiang, L.; Liu, W.; Liu, B.; Yuan, Y.; Bao, A. Monitoring vegetation sensitivity to drought events in China. *Sci. Total Environ.* **2023**, *893*, 164917. [CrossRef] [PubMed]
16. Otkin, J.A.; Svoboda, M.; Hunt, E.D.; Ford, T.W.; Anderson, M.C.; Hain, C.; Basara, J.B. Flash droughts: A review and assessment of the challenges imposed by rapid-onset droughts in the United States. *Bull. Am. Meteorol. Soc.* **2018**, *99*, 911–919. [CrossRef]

17. Qing, Y.; Wang, S.; Ancell, B.C.; Yang, Z.-L. Accelerating flash droughts induced by the joint influence of soil moisture depletion and atmospheric aridity. *Nat. Commun.* **2022**, *13*, 1139. [CrossRef] [PubMed]
18. Yang, L.; Wang, W.; Wei, J. Assessing the response of vegetation photosynthesis to flash drought events based on a new identification framework. *Agric. For. Meteorol.* **2023**, *339*, 109545. [CrossRef]
19. Yuan, Y.; Bao, A.; Liu, T.; Zheng, G.; Jiang, L.; Guo, H.; Jiang, P.; Yu, T.; De Maeyer, P. Assessing vegetation stability to climate variability in Central Asia. *J. Environ. Manag.* **2021**, *298*, 113330. [CrossRef] [PubMed]
20. Kusch, E.; Davy, R.; Seddon, A.W. Vegetation-memory effects and their association with vegetation resilience in global drylands. *J. Ecol.* **2022**, *110*, 1561–1574. [CrossRef]
21. Mitchell, P.J.; O'Grady, A.P.; Pinkard, E.A.; Brodribb, T.J.; Arndt, S.K.; Blackman, C.J.; Duursma, R.A.; Fensham, R.J.; Hilbert, D.W.; Nitschke, C.R. An ecoclimatic framework for evaluating the resilience of vegetation to water deficit. *Glob. Change Biol.* **2016**, *22*, 1677–1689. [CrossRef]
22. Liang, M.; Cao, R.; Di, K.; Han, D.; Hu, Z. Vegetation resistance and resilience to a decade-long dry period in the temperate grasslands in China. *Ecol. Evol.* **2021**, *11*, 10582–10589. [CrossRef] [PubMed]
23. Yevjevich, V.M. *Objective Approach to Definitions and Investigations of Continental Hydrologic Droughts*; Colorado State University: Fort Collins, CO, USA, 1967.
24. Wang, Y.; Yuan, X. Anthropogenic speeding up of South China flash droughts as exemplified by the 2019 Summer-Autumn transition season. *Geophys. Res. Lett.* **2021**, *48*, e2020GL091901. [CrossRef]
25. Zhang, M.; Yuan, X.; Otkin, J.A.; Ji, P. Climate warming outweighs vegetation greening in intensifying flash droughts over China. *Environ. Res. Lett.* **2022**, *17*, 054041. [CrossRef]
26. Dong, G.; Fan, L.; Fensholt, R.; Frappart, F.; Ciais, P.; Xiao, X.; Sitch, S.; Xing, Z.; Yu, L.; Zhou, Z. Asymmetric response of primary productivity to precipitation anomalies in Southwest China. *Agric. For. Meteorol.* **2023**, *331*, 109350. [CrossRef]
27. Dong, B.; Yu, Y.; Pereira, P. Non-growing season drought legacy effects on vegetation growth in southwestern China. *Sci. Total Environ.* **2022**, *846*, 157334. [CrossRef] [PubMed]
28. Zhang, X.; Yue, Y.; Tong, X.; Wang, K.; Qi, X.; Deng, C.; Brandt, M. Eco-engineering controls vegetation trends in southwest China karst. *Sci. Total Environ.* **2021**, *770*, 145160. [CrossRef] [PubMed]
29. Myneni, R.; Park, T. *MODIS/Terra+ Aqua Leaf Area Index/FPAR 4-Day L4 Global 500 m SIN Grid V061*; The Land Processes Distributed Active Archive Center (LP DAAC): Sioux Falls, SD, USA, 2021.
30. Song, P.; Zhang, Y.; Guo, J.; Shi, J.; Zhao, T.; Tong, B. A 1 km daily surface soil moisture dataset of enhanced coverage under all-weather conditions over China in 2003–2019. *Earth Syst. Sci. Data* **2022**, *14*, 2613–2637. [CrossRef]
31. Yang, J.; Huang, X. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. *Earth Syst. Sci. Data* **2021**, *13*, 3907–3925. [CrossRef]
32. Lyons, M.B.; Keith, D.A.; Phinn, S.R.; Mason, T.J.; Elith, J. A comparison of resampling methods for remote sensing classification and accuracy assessment. *Remote Sens. Environ.* **2018**, *208*, 145–153. [CrossRef]
33. Jiang, L.; Liu, B.; Yuan, Y. Quantifying Vegetation Vulnerability to Climate Variability in China. *Remote Sens.* **2022**, *14*, 3491. [CrossRef]
34. Li, Y.; Zhang, W.; Schwalm, C.R.; Gentile, P.; Smith, W.K.; Ciais, P.; Kimball, J.S.; Gazol, A.; Kannenberg, S.A.; Chen, A. Widespread spring phenology effects on drought recovery of Northern Hemisphere ecosystems. *Nat. Clim. Change* **2023**, *13*, 182–188. [CrossRef]
35. Schwalm, C.R.; Anderegg, W.R.L.; Michalak, A.M.; Fisher, J.B.; Biondi, F.; Koch, G.; Litvak, M.; Ogle, K.; Shaw, J.D.; Wolf, A.; et al. Global patterns of drought recovery. *Nature* **2017**, *548*, 202–205. [CrossRef] [PubMed]
36. Li, J.; Wang, Z.; Wu, X.; Chen, J.; Guo, S.; Zhang, Z. A new framework for tracking flash drought events in space and time. *Catena* **2020**, *194*, 104763. [CrossRef]
37. Ma, Q.; Li, Y.; Liu, F.; Feng, H.; Biswas, A.; Zhang, Q. SPEI and multi-threshold run theory based drought analysis using multi-source products in China. *J. Hydrol.* **2023**, *616*, 128737. [CrossRef]
38. Isbell, F.; Craven, D.; Connolly, J.; Loreau, M.; Schmid, B.; Beierkuhnlein, C.; Bezemer, T.M.; Bonin, C.; Bruehlheide, H.; de Luca, E.; et al. Biodiversity increases the resistance of ecosystem productivity to climate extremes. *Nature* **2015**, *526*, 574–577. [CrossRef] [PubMed]
39. Vogel, A.; Scherer-Lorenzen, M.; Weigelt, A. Grassland resistance and resilience after drought depends on management intensity and species richness. *PLoS ONE* **2012**, *7*, e36992. [CrossRef] [PubMed]
40. Deng, Y.; Wang, X.; Wang, K.; Ciais, P.; Tang, S.; Jin, L.; Li, L.; Piao, S. Responses of vegetation greenness and carbon cycle to extreme droughts in China. *Agric. For. Meteorol.* **2021**, *298*, 108307. [CrossRef]



41. Liu, Y.; You, C.; Zhang, Y.; Chen, S.; Zhang, Z.; Li, J.; Wu, Y. Resistance and resilience of grasslands to drought detected by SIF in inner Mongolia, China. *Agric. For. Meteorol.* **2021**, *308*, 108567. [CrossRef]
42. Zhang, L.; Ameca, E.I.; Cowlshaw, G.; Pettorelli, N.; Foden, W.; Mace, G.M. Global assessment of primate vulnerability to extreme climatic events. *Nat. Clim. Change* **2019**, *9*, 554–561. [CrossRef]
43. Jiang, L.; Liu, B.; Guo, H.; Yuan, Y.; Liu, W.; Jiapaer, G. Assessing vegetation resilience and vulnerability to drought events in Central Asia. *J. Hydrol.* **2024**, *634*, 131012. [CrossRef]
44. Jiang, L.; Bao, A.; Jiapaer, G.; Guo, H.; Zheng, G.; Gafforov, K.; Kurban, A.; De Maeyer, P. Monitoring land sensitivity to desertification in Central Asia: Convergence or divergence? *Sci. Total Environ.* **2019**, *658*, 669–683. [CrossRef] [PubMed]
45. Canetti, R.M.; España, M.D. Convergence analysis of the least-squares identification algorithm with a variable forgetting factor for time-varying linear systems. *Automatica* **1989**, *25*, 609–612. [CrossRef]
46. Tombolini, I.; Colantoni, A.; Renzi, G.; Sateriano, A.; Sabbi, A.; Morrow, N.; Salvati, L. Lost in convergence, found in vulnerability: A spatially-dynamic model for desertification risk assessment in Mediterranean agro-forest districts. *Sci. Total Environ.* **2016**, *569–570*, 973–981. [CrossRef] [PubMed]
47. Zhang, D.; Jia, Q.; Wang, P.; Zhang, J.; Hou, X.; Li, X.; Li, W. Analysis of spatial variability in factors contributing to vegetation restoration in Yan'an, China. *Ecol. Indic.* **2020**, *113*, 106278. [CrossRef]
48. Rahmati, O.; Falah, F.; Dayal, K.S.; Deo, R.C.; Mohammadi, F.; Biggs, T.; Moghaddam, D.D.; Naghibi, S.A.; Bui, D.T. Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. *Sci. Total Environ.* **2020**, *699*, 134230. [CrossRef] [PubMed]
49. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [CrossRef] [PubMed]
50. Hastie, T.; Tibshirani, R.; Friedman, J.H.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: Berlin/Heidelberg, Germany, 2009; Volume 2.
51. Naghibi, S.A.; Pourghasemi, H.R.; Dixon, B. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environ. Monit. Assess.* **2016**, *188*, 44. [CrossRef] [PubMed]
52. Yang, R.-M.; Zhang, G.-L.; Liu, F.; Lu, Y.-Y.; Yang, F.; Yang, F.; Yang, M.; Zhao, Y.-G.; Li, D.-C. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecol. Indic.* **2016**, *60*, 870–878. [CrossRef]
53. Ma, S.; Wang, L.-J.; Jiang, J.; Chu, L.; Zhang, J.-C. Threshold effect of ecosystem services in response to climate change and vegetation coverage change in the Qinghai-Tibet Plateau ecological shelter. *J. Clean. Prod.* **2021**, *318*, 128592. [CrossRef]
54. Zhang, Y.; Qu, H.; Yang, X.; Wang, M.; Qin, N.; Zou, Y. Cropping system optimization for drought prevention and disaster reduction with a risk assessment model in Sichuan Province. *Glob. Ecol. Conserv.* **2020**, *23*, e01095. [CrossRef]
55. Chen, T.; Wang, Q.; Wang, Y.; Peng, L. Differentiation characteristics of karst vegetation resilience and its response to climate and ecological restoration projects. *Land Degrad. Dev.* **2023**, *34*, 5055–5070.
56. Ge, W.; Li, X.; Xie, M.; Yu, B.; Jiao, J.; Han, J.; Wang, F. Quantitative evaluation of drought risk related to vegetation productivity in China. *J. Hydrol.* **2023**, *623*, 129877. [CrossRef]
57. Zeng, Z.; Wu, W.; Peñuelas, J.; Li, Y.; Jiao, W.; Li, Z.; Ren, X.; Wang, K.; Ge, Q. Increased risk of flash droughts with raised concurrent hot and dry extremes under global warming. *npj Clim. Atmos. Sci.* **2023**, *6*, 134. [CrossRef]
58. Christian, J.I.; Martin, E.R.; Basara, J.B.; Furtado, J.C.; Otkin, J.A.; Lowman, L.E.; Hunt, E.D.; Mishra, V.; Xiao, X. Global projections of flash drought show increased risk in a warming climate. *Commun. Earth Environ.* **2023**, *4*, 165. [CrossRef]
59. Alizadeh, M.R.; Adamowski, J.; Nikoo, M.R.; AghaKouchak, A.; Dennison, P.; Sadegh, M. A century of observations reveals increasing likelihood of continental-scale compound dry-hot extremes. *Sci. Adv.* **2020**, *6*, eaaz4571. [CrossRef] [PubMed]
60. Christian, J.I.; Hobbins, M.; Hoell, A.; Otkin, J.A.; Ford, T.W.; Cravens, A.E.; Powlen, K.A.; Wang, H.; Mishra, V. Flash drought: A state of the science review. *Wiley Interdiscip. Rev. Water* **2024**, *11*, e1714. [CrossRef]
61. Yuan, X.; Ma, Z.; Pan, M.; Shi, C. Microwave remote sensing of short-term droughts during crop growing seasons. *Geophys. Res. Lett.* **2015**, *42*, 4394–4401. [CrossRef]
62. Lin, W.; Wen, C.; Wen, Z.; Gang, H. Drought in Southwest China: A review. *Atmos. Ocean. Sci. Lett.* **2015**, *8*, 339–344.
63. Zhou, S.; Williams, A.P.; Berg, A.M.; Cook, B.I.; Zhang, Y.; Hagemann, S.; Lorenz, R.; Seneviratne, S.I.; Gentile, P. Land-atmosphere feedbacks exacerbate concurrent soil drought and atmospheric aridity. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 18848–18853. [CrossRef]
64. Mukherjee, S.; Mishra, A.K. Global flash drought analysis: Uncertainties from indicators and datasets. *Earth's Future* **2022**, *10*, e2022EF002660. [CrossRef]

65. Yao, C.; Luo, Z.; Wang, H.; Li, Q.; Zhou, H. GRACE-derived terrestrial water storage changes in the inter-basin region and its possible influencing factors: A case study of the Sichuan Basin, China. *Remote Sens.* **2016**, *8*, 444. [CrossRef]
66. Zhang, L.; Chang, J.; Guo, A.; Zhou, K.; Yang, G.; Zou, D. Ecological drought evolution characteristics under different climatic regions in the Yangtze River basin. *J. Hydrol.* **2024**, *629*, 130573. [CrossRef]
67. Ding, T.; Gao, H. The record-breaking extreme drought in Yunnan Province, Southwest China during spring-early summer of 2019 and possible causes. *J. Meteorol. Res.* **2020**, *34*, 997–1012. [CrossRef]
68. Li, C.; Adu, B.; Wu, J.; Qin, G.; Li, H.; Han, Y. Spatial and temporal variations of drought in Sichuan Province from 2001 to 2020 based on modified temperature vegetation dryness index (TVDI). *Ecol. Indic.* **2022**, *139*, 108883. [CrossRef]
69. Mahto, S.S.; Mishra, V. Increasing risk of simultaneous occurrence of flash drought in major global croplands. *Environ. Res. Lett.* **2023**, *18*, 044044. [CrossRef]
70. Wang, L.; Chen, W.; Zhou, W. Assessment of future drought in Southwest China based on CMIP5 multimodel projections. *Adv. Atmos. Sci.* **2014**, *31*, 1035–1050. [CrossRef]
71. Zhu, X.; Liu, Y.; Xu, K.; Pan, Y. Effects of drought on vegetation productivity of farmland ecosystems in the drylands of northern China. *Remote Sens.* **2021**, *13*, 1179. [CrossRef]
72. Xu, P.; Fang, W.; Zhou, T.; Li, H.; Zhao, X.; Berman, S.; Zhang, T.; Yi, C. Satellite evidence of canopy-height dependence of forest drought resistance in southwestern China. *Environ. Res. Lett.* **2022**, *17*, 025005. [CrossRef]
73. Cochard, H.; Pimont, F.; Ruffault, J.; Martin-StPaul, N. SurEau: A mechanistic model of plant water relations under extreme drought. *Ann. For. Sci.* **2021**, *78*, 55. [CrossRef]
74. Chen, Y.; Taylor, P.; Cuddy, S.; Wahid, S.; Penton, D.; Karim, F. Inferring vegetation response to drought at multiscale from long-term satellite imagery and meteorological data in Afghanistan. *Ecol. Indic.* **2024**, *158*, 111567. [CrossRef]
75. Zhang, G.; Yang, Y.; Wei, Z.; Zhu, X.; Shen, W.; Ma, J.; Lv, S.; Xu, H. The low greenhouse gas emission intensity in water-saving and drought-resistance rice in a rainfed paddy field in Southwest China. *Field Crops Res.* **2023**, *302*, 109045. [CrossRef]
76. Li, X.; Li, Y.; Chen, A.; Gao, M.; Slette, I.J.; Piao, S. The impact of the 2009/2010 drought on vegetation growth and terrestrial carbon balance in Southwest China. *Agric. For. Meteorol.* **2019**, *269*, 239–248. [CrossRef]
77. Chen, Q.; Timmermans, J.; Wen, W.; van Bodegom, P.M. Ecosystems threatened by intensified drought with divergent vulnerability. *Remote Sens. Environ.* **2023**, *289*, 113512. [CrossRef]
78. Fang, O.; Zhang, Q. Tree resilience to drought increases in the Tibetan Plateau. *Glob. Change Biol.* **2019**, *25*, 245–253. [CrossRef]
79. Yue, Y.; Qi, X.; Wang, K.; Liao, C.; Tong, X.; Brandt, M.; Liu, B. Large scale rocky desertification reversal in South China karst. *Prog. Phys. Geogr. Earth Environ.* **2022**, *46*, 661–675. [CrossRef]
80. Paluch, J. The stochastic backward shifts model better corresponds to the fine-scale structural heterogeneity of old-growth Abies-Fagus-Picea forests than the ontogenic life cycle model. *For. Ecol. Manag.* **2021**, *486*, 118978. [CrossRef]
81. Tong, X.; Brandt, M.; Yue, Y.; Horion, S.; Wang, K.; Keersmaecker, W.D.; Tian, F.; Schurgers, G.; Xiao, X.; Luo, Y. Increased vegetation growth and carbon stock in China karst via ecological engineering. *Nat. Sustain.* **2018**, *1*, 44–50. [CrossRef]
82. Tong, X.; Wang, K.; Yue, Y.; Brandt, M.; Liu, B.; Zhang, C.; Liao, C.; Fensholt, R. Quantifying the effectiveness of ecological restoration projects on long-term vegetation dynamics in the karst regions of Southwest China. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *54*, 105–113. [CrossRef]
83. Bai, Y.-H.; Tang, Z. Enhanced effects of species richness on resistance and resilience of global tree growth to prolonged drought. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2410467121. [CrossRef] [PubMed]
84. D'Odorico, P.; Laio, F.; Ridolfi, L. Patterns as indicators of productivity enhancement by facilitation and competition in dryland vegetation. *J. Geophys. Res. Biogeosciences* **2006**, *111*, G03010. [CrossRef]
85. Lu, H.; Wu, Y.; Li, Y.; Liu, Y. Effects of meteorological droughts on agricultural water resources in southern China. *J. Hydrol.* **2017**, *548*, 419–435. [CrossRef]
86. Mokhtar, A.; He, H.; Alsafadi, K.; Mohammed, S.; Ayantobo, O.O.; Elbeltagi, A.; Abdelwahab, O.M.; Zhao, H.; Quan, Y.; Abdo, H.G. Assessment of the effects of spatiotemporal characteristics of drought on crop yields in southwest China. *Int. J. Climatol.* **2022**, *42*, 3056–3075. [CrossRef]
87. Ge, W.; Han, J.; Zhang, D.; Wang, F. Divergent impacts of droughts on vegetation phenology and productivity in the Yungui Plateau, southwest China. *Ecol. Indic.* **2021**, *127*, 107743. [CrossRef]
88. Xu, M.; Zhang, T.; Zhang, Y.; Chen, N.; Zhu, J.; He, Y.; Zhao, T.; Yu, G. Drought limits alpine meadow productivity in northern Tibet. *Agric. For. Meteorol.* **2021**, *303*, 108371. [CrossRef]
89. Mueller, R.C.; Scudder, C.M.; Porter, M.E.; Talbot Trotter, R., III; Gehring, C.A.; Whitham, T.G. Differential tree mortality in response to severe drought: Evidence for long-term vegetation shifts. *J. Ecol.* **2005**, *93*, 1085–1093. [CrossRef]

90. Han, Z.; Huang, Q.; Huang, S.; Leng, G.; Bai, Q.; Liang, H.; Wang, L.; Zhao, J.; Fang, W. Spatial-temporal dynamics of agricultural drought in the Loess Plateau under a changing environment: Characteristics and potential influencing factors. *Agric. Water Manag.* **2021**, *244*, 106540. [CrossRef]
91. Zhang, X.; Zhang, B. The responses of natural vegetation dynamics to drought during the growing season across China. *J. Hydrol.* **2019**, *574*, 706–714. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## Article

# Eutrophication Conditions in Two High Mountain Lakes: The Influence of Climate Conditions and Environmental Pollution

Fátima Goretti García-Miranda <sup>1</sup>, Claudia Muro <sup>1,\*</sup>, Yolanda Alvarado <sup>1</sup>, José Luis Expósito-Castillo <sup>2</sup> and Héctor Víctor Cabadas-Báez <sup>3,4</sup>

<sup>1</sup> Tecnológico Nacional de México, Instituto Tecnológico de Toluca, Av. Tecnológico s/n, Colonia Agrícola Bellavista, Metepec 52149, Mexico; dd20280298@toluca.tecnm.mx (F.G.G.-M.); yoa@toluca.tecnm.mx (Y.A.)

<sup>2</sup> Instituto Interamericano de Tecnología y Ciencias del Agua (IITCA), Universidad Autónoma del Estado de México, Carretera Toluca-Atlaconulco km. 14, Toluca, Mexico; jlexpositoc@uaemex.mx

<sup>3</sup> Escuela Nacional de Conservación, Restauración y Museografía “Manuel del Castillo Negrete” (INAH), Calz. Gral. Anaya 187, San Diego Churubusco, Coyoacán, Ciudad de Mexico 04120, Mexico; victor\_cabadas\_b@encrym.edu.mx

<sup>4</sup> Universidad Autónoma del Estado de México, Facultad de Geografía, Cerro de Coatepec S/N, Ciudad Universitaria, Toluca 50110, Mexico

\* Correspondence: cmurou@toluca.tecnm.mx; Tel.: +52-7227096994

**Abstract:** The lakes known as El Sol and La Luna are high mountain water deposits located in Mexico within an inactive volcanic system. These lakes are of ecological importance because they are unique in Mexico. However, currently, the lakes have experienced changes in their shape and an increase in algae blooms, coupled with the degradation of the basin, which has alerted government entities to the need to address the lakes’ problems. To address the environmental status of El Sol and La Luna, a trophic study was conducted during the period of 2021–2023, including an analysis of the influence of climatic variables, lake water quality, and eutrophication conditions. The trophic state was established based on the eutrophication index. The Pearson correlations defined the eutrophication interrelation between the distinct factors influencing the lakes’ status. El Sol registered higher eutrophication conditions than La Luna. El Sol was identified as seasonal eutrophic and La Luna as transitioning from oligotrophic to mesotrophic, showing high levels of chlorophyll, total phosphorus, and total nitrogen and low water transparency. The principal factors altering the eutrophic conditions were water pollution and climatic variables (precipitation and ambient temperature). Eutrophication was the prime factor impacting perimeter loss at El Sol, whereas at La Luna, it was due to a decline in precipitation.

**Keywords:** eutrophication; high mountain lakes; climate variables; eutrophic parameters; trophic index

## 1. Introduction

The increase in pollution and climate change have had a joint impact on aquatic environments, evidenced by alterations in eutrophication conditions and deterioration processes [1,2]. The principal pollutants producing eutrophication are associated with agrochemicals and nutrients.

Agrochemicals are available in the atmosphere and travel long distances and heights. Due to easy displacement, they are deposited on water surfaces and accumulate in the deepest regions of water ecosystems, causing fertilization, nutrient accumulation, and high levels of pollution, with devastating consequences [3,4]. Agrochemicals comprise



pesticides, fertilizers, and plant growth hormones and contain nitrogen, phosphorus, and heavy metals. In addition, other nutrients from atmospheric deposition, local wildlife, and direct human activities contribute to water pollution [5], which is manifested by the increase in phytoplankton biomass, algae blooms, water degradation, and complete eutrophication due to water and biota loss.

Currently, eutrophication has reached lakes; 50% of them have suffered environmental alterations due to eutrophication [6]. This is disturbing because lakes are a vital part of the terrestrial hydrosphere, providing fresh water, a habitat for wild flora and fauna, and maintaining ecological balance; however, the high risk of water quality modifications and biodiversity affections is latent [7].

High mountain lakes have also registered analogous environmental and eutrophication risks to the lakes systems, breaking the non-pollution status in those places.

The present data reveal that high mountain lakes are threatened by the relentless march of climate change and intensified human activities. Nutrient accumulation and changes to water quality, excess algae, and the predominance of biota eutrophication alert us to the risk of water loss and biological communities in high mountain lakes [8].

The most studied alpine lakes are mainly located in Europe and Asia; the reports describe a similar eutrophication panorama. Kuczynsky et al. [9] analyzed the lakes of Aotearoa, New Zealand. However, seasonal changes did not permit the environmental category of lake eutrophication to be defined.

Magyari et al. [10] analyzed trophic modifications in Lake Balaton, East-Central Europe's largest shallow lake, using palaeoecological (chironomids, pollen) and geochemical (sediment chlorophyll, TOC, TS, TN, and C/H) ratios. In this case, a rapid increase in benthivore fish biomass and eutrophication-related alterations were observed in the lakes, indicating significant eutrophic degradation. Zuo et al. [11] investigated environmental changes in 59 lakes and reservoirs in eastern China, measuring their impact on cyanobacterial communities. The authors found that cyanobacteria diversity declined due to lake deterioration caused by anthropogenic activities. Han et al. [12] studied an alpine lake in the Tianshan Mountains, Central Asia (2015–2021), and detected several changes attributed to climate warming and atmospheric nutrient deposition. Similarly, Han et al. [13] explored Lake Yamzho Yumco (Southern QTP) over the past three decades, focusing on eutrophic taxa. The lake exhibited alterations in 2004, which accelerated in 2014, showing an increase in eutrophic taxa due to nutrient enrichment. These eutrophic conditions were attributed to anthropogenic emissions from the catchment.

Poikane et al. [14] analyzed nutrient targets for five major ecoregions of Europe—Northern, Central Baltic, Alpine, Mediterranean, and Eastern Continental—by developing regressions between nutrient concentrations and ecological quality ratios (EQRs) based on phytoplankton and macrophyte communities. The authors found that these relationships were stronger for phosphorus than for nitrogen and more pronounced for phytoplankton than for macrophytes. They also established the nutrient concentrations required for lakes to achieve good ecological status.

Jacquemin et al. [15] investigated the impact of an atmospheric nutrient deposition regime on the taxonomic and functional composition of phytoplankton communities in six French alpine lakes with varying catchment sizes, vegetation cover, and contrasting nitrogen (N) and phosphorus (P) deposition regimes. Photoautotrophic taxa dominated the phytoplankton assemblages, while mixotrophic taxa were even more favored in lakes with large, vegetated catchments.

In the Americas, the eutrophication of high mountain lakes has received little attention. Studies on the eutrophication of high mountain lakes in Mexico are absent.

In Mexico, two high mountain lakes, El Sol and La Luna (as they are commonly known), are located at more than 3000 m above sea level (MASL) [16]. These lakes are currently experiencing visible ecological alterations, including algae proliferation and water loss, increasing their risk of eutrophication. However, environmental data on these lakes are scarce and insufficient to determine their trophic status. As a result, there are no established strategies for lake restoration or eutrophication management.

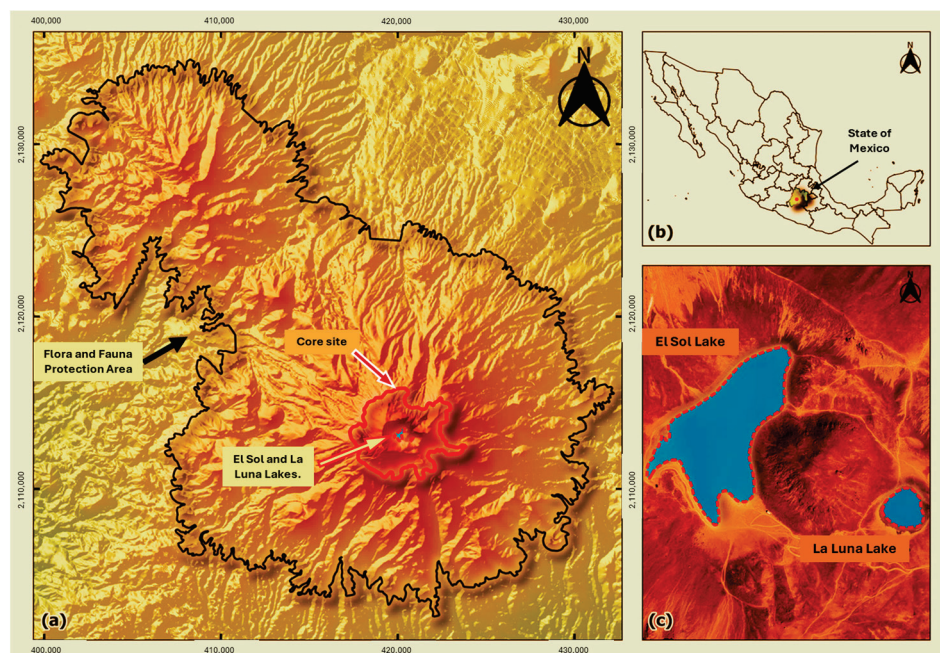
This research involved the environmental monitoring of Lakes El Sol and La Luna to assess their eutrophic status and the influence of water pollution and climatic factors. This study is part of the first conservation plan organized by government institutions in Mexico, including the National Commission for Natural Protected Areas (CONANP) and the National Water Commission (CONAGUA). Monitoring and eutrophication studies were conducted between 2021 and 2023.

## 2. Materials and Methods

### 2.1. Study Area

Lakes El Sol and La Luna are crater alpine lakes situated within an extinct volcano, Nevado de Toluca, located in the core zone of the Flora and Fauna Protection Area in Mexico.

Figure 1 illustrates the Protection Area, the volcano within the core site, the volcano's location in Mexico, and the position of Lakes El Sol and La Luna within the volcanic crater.



**Figure 1.** (a) Flora and Fauna Protection Area of the Nevado de Toluca (black line), volcano, and the core site of Lakes El Sol and La Luna (red line). (b) The volcano's location within Mexico. (c) Lakes El Sol and La Luna in the volcano (blue area).

The lakes are in an endorheic basin and are classified as continuous warm polymictic. Both lakes are located at an approximate altitude of 4200 MASL. The surface area of Lake El Sol is 237,321 m<sup>2</sup>; Lake la Luna measures 31,083 m<sup>2</sup>. Lake El Sol has a depth of 15 m, while Lake La Luna is 10 m deep [16–19].

The lake basin is characterized by andesitic–dacitic boulders and a high-altitude, subhumid semi-cold climate, with an average annual temperature ranging from  $-2^{\circ}\text{C}$  to  $7^{\circ}\text{C}$ .

The basin's vegetation consists primarily of alpine grassland, typical of mountainous regions, where grasses, lichens, and mosses dominate. The area also contains patches of pine–oak, pine, fir, and tropical deciduous forest [20].

The lakes receive water primarily from rainfall in summer, snow in winter, and meltwater in spring. They are open to tourism, attracting visitors for hiking and mountain climbing. The main economic activities in the region are agriculture and livestock.

Lakes El Sol and La Luna serve as reservoirs for phytoplankton and zooplankton, typical of alpine lakes. Despite their geographic proximity, each lake has a distinct diatom community, with La Luna exhibiting lower abundance and species richness [21–25]. Trout were introduced into both lakes in the 1960s, but they only survived in Lake El Sol.

In recent decades, isolated studies have classified Lake El Sol as oligotrophic and La Luna as ultraoligotrophic [17]. However, both lakes now show visible contamination, including a reduced water perimeter and algal blooms [26].

These environmental changes in the lakes have raised concerns among government agencies, requiring monitoring to assess the factors contributing to the lakes' eutrophication.

To monitor the lakes, the studies included measuring the lakes' perimeters, recording climatic variables, analyzing water pollution, and evaluating water quality variables and eutrophic conditions from 2021 to 2023 to determine the trophic status of Lakes El Sol and La Luna and the causes.

## 2.2. Climatic Variables Information

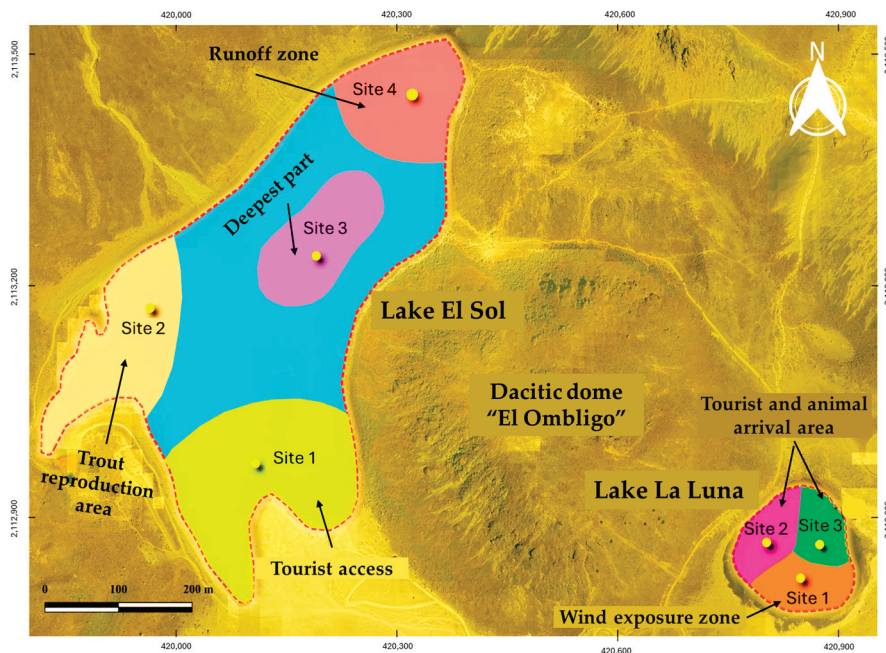
The analysis of climatic variables included records of water evaporation, precipitation, and ambient temperature in the lakes during the study period. Data on these climatic variables were collected from the monthly summaries of the National Meteorological System (SMN), specifically from station 15072-Nevado de Toluca ( $19^{\circ}07'07''\text{ N}$ ,  $99^{\circ}46'53''\text{ W}$ ), 1.4 km NE of the crater [27,28]. The station is located at a higher altitude in the protected natural area.

## 2.3. Lake Perimeter Measurement

The lake perimeter measurement established polygons of Lakes El Sol and La Luna using GPS tracking. Then, Geographic Information Systems were used to determine the perimeter of each lake.

## 2.4. Analysis of Water Quality and Eutrophication Parameters

Figure 2 shows the sampling sites for water quality and eutrophication parameters in the lakes. Sampling was conducted at four locations in Lake El Sol and three in Lake La Luna. The sites were selected based on previous studies and represent areas of influence where pollution tends to concentrate [25,26]. The sampling sites in Lake El Sol are identified as follows: Site 1 is the main access area where tourists and animals arrive; Site 2 is the trout reproduction area; Site 3 is the deepest part of the lake; and Site 4 is where the runoff from the high mountains converges, coinciding with a tourist access point after crossing the surrounding peaks. For Lake La Luna, Site 1 is fully exposed to the wind due to the absence of high mountains around the lake. Sites 2 and 3 are located at points where tourists and animals reach the lake by crossing a dome that separates Lakes El Sol and La Luna. In addition, these sites receive runoff from the high mountains.



**Figure 2.** Sampling sites for water from Lakes El Sol and La Luna (delimited by red dash lines, corresponding to the current lakes' perimeters).

Water samples from the lakes were collected during three seasons each year, i.e., dry (March–June), wet (July–October), and cold (October–February), over the 2021–2023 period.

Each sample consisted of 2 L of water per sampling site in each lake, collected from the epilimnion at a depth of 1 m, using a Van Dorn bottle. A total of 12 samples were taken from Lake El Sol and 9 from Lake La Luna per sampling day.

The samples were stored in glass flasks and preserved at 4 °C in the laboratory for analysis.

Water quality parameters were determined by measuring pH, electrical conductivity (EC), and total solids (TS). Eutrophication factors were assessed based on chlorophyll-a (Chl-a), total nitrogen (TN), total phosphorus (TP), and water column transparency, which was measured using a Secchi disk (SD).

The water parameters, including pH, EC, TS, TN, TP, and Chl-a, were measured following Standard Methods [29]. Additionally, the analysis of TN and TP was conducted in accordance with Mexican regulations NMX-AA-026-SCFI-2010 and NMX-AA-029-SCFI-2001 [30,31].

### 2.5. Trophic State Determination of the Lakes

The trophic state of Lakes El Sol and La Luna was determined using the Trophic State Index (TSI) of Carlson and the eutrophic parameter values established by Carlson and the Organization for Economic Co-operation and Development (OECD). The TSI of Carlson was calculated using Equations (1)–(3), incorporating eutrophication parameters such as SD, TP, and Chl-a.

$$\text{TSICT (SD)} = 60 - 14.4 \ln \text{SD}, \quad (1)$$

$$\text{TSICT (TP)} = 14.42 \ln \text{TP} + 4.15, \quad (2)$$

$$\text{TSIChl-a} = 9.81 \ln \text{Chl-a} + 30.6. \quad (3)$$

Based on the TSI of Carlson intervals and the corresponding average values of TP, SD, and Chl-a, the eutrophication status of the water was classified into five categories:



ultraoligotrophic (UOlig), oligotrophic (Olig), mesotrophic (Mtrop), eutrophic (Eutrop), and hypertrophic (Hiptrop). Table 1 presents the trophic classification according to the TSI of Carlson, along with the eutrophic parameter intervals defined by Carlson and the adjusted eutrophic parameters of the OECD [32–34].

**Table 1.** Trophic classification by index and average values of the trophic parameters of the water for TSICT and TSIOECD.

| TSI of Carlson                |        |                      |                         | TSI of the OECD               |           |                      |                         |
|-------------------------------|--------|----------------------|-------------------------|-------------------------------|-----------|----------------------|-------------------------|
| Eutrophic Parameters/Category | SD (M) | TP mg/m <sup>3</sup> | Chl-a mg/m <sup>3</sup> | Eutrophic Parameters/Category | SD Mean m | TP mg/m <sup>3</sup> | Chl-a mg/m <sup>3</sup> |
| Oligotrophic                  | 64     | 0.75                 | 0.04                    | Ultraoligotrophic             | 12–6      | <4                   | <2.5                    |
|                               | 32     | 1.5                  | 0.12                    | Oligotrophic                  | 6–3       | <10                  | 2.5–8                   |
|                               | 16     | 3                    | 0.34                    |                               |           |                      |                         |
|                               | 8      | 6                    | 0.94                    |                               |           |                      |                         |
| Mesotrophic                   | 4      | 12                   | 2.6                     |                               |           |                      |                         |
|                               | 2      | 24                   | 6.4                     | Mesotrophic                   | 3–1.5     | 10–35                | 8–25                    |
|                               | 1      | 48                   | 20                      |                               |           |                      |                         |
|                               | 0.5    | 96                   | 56                      |                               |           |                      |                         |
| Eutrophic                     | 0.25   | 192                  | 154                     | Eutrophic                     | 1.5–0.7   | 35–100               | 25–75                   |
|                               | 0.12   | 384                  | 427                     |                               |           |                      |                         |
| Hypereutrophic                | 0.06   | 768                  | 1183                    | Hypereutrophic                | >0.7      | >100                 | >75                     |

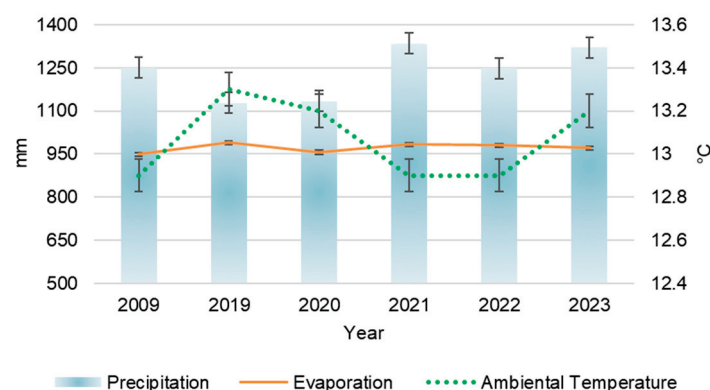
## 2.6. Statistical Analysis

The collected data represent the average values of three samples and three measurements. The standard deviation was used to assess the variability in the data. An ANOVA analysis was performed to compare variable values, while the statistical relationship between water quality, meteorological factors, and eutrophication during each study period was evaluated using the Pearson correlation coefficient ( $p < 0.05$ ). A positive correlation (values  $> 0$ ) indicated an association between variables. All statistical analyses were conducted using Origin software version 10.5.

## 3. Results and Discussion

### 3.1. Climatic Variables in the Crater Area of Lakes El Sol and La Luna

Figure 3 presents the average climatic variables for Lakes El Sol and La Luna over the study period, including ambient temperature (°C), precipitation, and evaporation (mm). The data correspond to the study period, along with meteorological station records and previously published studies from 2009 to date [23,26].



**Figure 3.** Climatic parameters in the crater area of Lakes El Sol and La Luna (2009–2023).

The analysis of climatic variables indicates that evaporation remained constant (>900 mm) throughout the study period, consistent with observations from 2009.

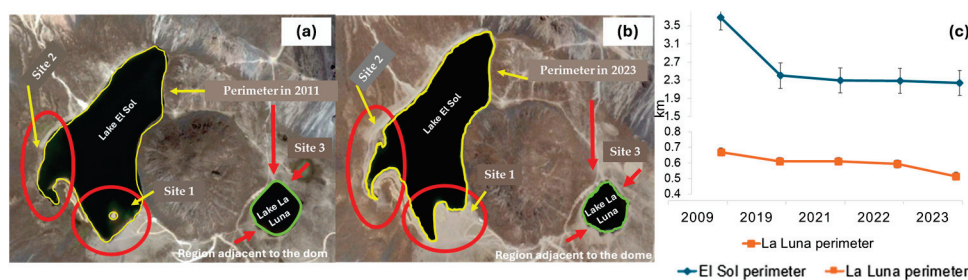
Precipitation levels varied over time. The average precipitation from 2021 to 2023 was 1400 mm, while 2022 and 2009 recorded 1250 mm, and 2019–2020 recorded 1100 mm. This indicates that between 2009 and 2021–2023, precipitation increased by 10–12%.

In turn, the ambient temperature shows fluctuations from 12.9 to 13.3 °C, with maximums in 2019 and 2023 and minimums in 2021–2022. However, from 2009 to 2023, the temperature showed changes to <0.5 °C.

Maximum ambient temperatures correspond to periods of low precipitation, while minimum ambient temperatures coincided with high precipitation. However, these variables did not have a significant influence on water evaporation.

### 3.2. Lake Perimeters

Figure 4a depicts the shape of the hydrological basin form and the lake perimeters based on data from the previous decade (2004–2011 records). Figure 4b illustrates the changes in the hydrological basin shape and lake perimeters observed during the 2021–2023 period.



**Figure 4.** (a) Hydrological basin shape (red circles) and lake perimeters based on data from the previous decade of 2011–2020. (b) Hydrological basin shape (red circles) and lake perimeters based on data from 2021 to 2023. The yellow line in Lake El Sol and the green line in Lake La Luna indicate the change in their size and hydrology at different times, showing the loss of the water column. (c) Lake perimeter data (km) between 2009 and 2023.

Lake El Sol shows significant alterations in its hydrological form, with the most noticeable changes occurring in the circled areas (Site 1 and 2), where a reduction in the lake area is evident. Similarly, Lake La Luna shows a loss in the area in the region adjacent to the dome that separates it from Lake El Sol.

Figure 4c further illustrates the changes in the average lake perimeter over time. In 2009, the recorded perimeters for Lakes El Sol and La Luna were 5 km and 1 km, respectively. However, the data from the 2021–2023 period show a 38% reduction in Lake El Sol's perimeter and a 15% decrease in Lake La Luna's perimeter compared to 2009. The 2021–2023 perimeters indicate that the loss observed between 2009 and 2023 is irreversible.

Seasonal variations in the lakes' perimeters also showed variations ( $p < 0.05$ ); however, perimeter loss was partially recovered during the wet season due to precipitation. The changes in the lakes' perimeters observed between 2009, 2021, and 2023 were linked to climatic variables and/or eutrophication conditions, which are analyzed in later sections. These data indicate a continuing risk of lake perimeter decline in the coming years.

### 3.3. Water Quality and Eutrophication Parameters

Table 2 shows the average water quality concentrations of Lakes El Sol and La Luna during the 2021–2023 seasons. Data from the lakes indicate that there are significant

differences between the lake water parameters ( $p < 0.05$ ). Lake El Sol recorded the highest parameters, implying that it had higher pollution than Lake La Luna. In turn, there were significant seasonal variations in the quality parameters for both lakes ( $p < 0.05$ ), with the highest impact during the dry, cold, and wet seasons.

**Table 2.** Average values of water quality parameters based on seasonal samples from Lakes El Sol and La Luna (2021–2023).

| Lake    | Season  | Ph            | EC ( $\mu\text{S}/\text{cm}$ ) | WT ( $^{\circ}\text{C}$ ) | TS ( $\text{mg}/\text{L}$ ) | Chl-a ( $\text{mg}/\text{m}^3$ ) | SD (m)        | TN ( $\text{mg}/\text{m}^3$ ) | TP ( $\text{mg}/\text{m}^3$ ) |
|---------|---------|---------------|--------------------------------|---------------------------|-----------------------------|----------------------------------|---------------|-------------------------------|-------------------------------|
| El Sol  | Dry     | $7.2 \pm 0.5$ | $49.5 \pm 0.5$                 | $12.5 \pm 0.4$            | $105 \pm 4$                 | $2.5 \pm 0.1$                    | $6.2 \pm 0.5$ | $530 \pm 3$                   | $10 \pm 1$                    |
|         | Wet     | $6.8 \pm 0.1$ | $46.5 \pm 0.4$                 | $11.9 \pm 0.6$            | $153 \pm 5$                 | $3.1 \pm 0.4$                    | $7.9 \pm 0.2$ | $680 \pm 4$                   | $43 \pm 1$                    |
|         | Cold    | $6.5 \pm 0.2$ | $46.0 \pm 1.3$                 | $10.8 \pm 0.9$            | $110 \pm 3$                 | $2.6 \pm 0.1$                    | $5.7 \pm 0.5$ | $450 \pm 2$                   | $66 \pm 2$                    |
|         | Average | $6.8 \pm 0.3$ | $47.0 \pm 1.5$                 | $11.5 \pm 0.3$            | $119 \pm 5$                 | $2.7 \pm 0.3$                    | $6.0 \pm 0.3$ | $520 \pm 1$                   | $40 \pm 1$                    |
| La Luna | Dry     | $6.3 \pm 0.1$ | $12.8 \pm 1.7$                 | $12.0 \pm 0.6$            | $53 \pm 5$                  | $1.2 \pm 0.3$                    | $7.0 \pm 0.3$ | $210 \pm 1$                   | $41 \pm 1$                    |
|         | Wet     | $6.0 \pm 0.2$ | $11.0 \pm 0.6$                 | $11.9 \pm 1.5$            | $85 \pm 8$                  | $2.0 \pm 0.6$                    | $7.5 \pm 0.1$ | $310 \pm 2$                   | $36 \pm 1$                    |
|         | Cold    | $6.2 \pm 0.2$ | $11.9 \pm 0.5$                 | $10.5 \pm 0.7$            | $65 \pm 2$                  | $1.3 \pm 0.4$                    | $7.0 \pm 1.3$ | $260 \pm 1$                   | $57 \pm 1$                    |
|         | Average | $6.2 \pm 0.1$ | $11.9 \pm 0.3$                 | $11.6 \pm 0.1$            | $68 \pm 3$                  | $1.7 \pm 0.1$                    | $7.4 \pm 0.1$ | $230 \pm 1$                   | $45 \pm 1$                    |

Lake El Sol recorded an average pH of  $6.8 \pm 0.3$ , while Lake La Luna had an average pH of  $6.2 \pm 0.1$ , both indicating a neutral pH, characteristic of environmentally healthy bodies of water [30]. However, during the dry season, the pH of Lake El Sol increased to 7.2.

Historical trends in pH levels (Figure 5a) reveal that in past decades, Lake El Sol had recorded values of  $<6$ , while Lake La Luna had pH values of  $<5$  [17,24,35,36]. However, recent data indicate a 20% increase in pH with a trend toward alkalization, which serves as an indicator of lake pollution.

The change in pH in the lake water is associated with factors such as catchment bedrock and soil type, climatic changes, and high algae activity due to excessive levels of nitrogen and phosphorus [37,38].

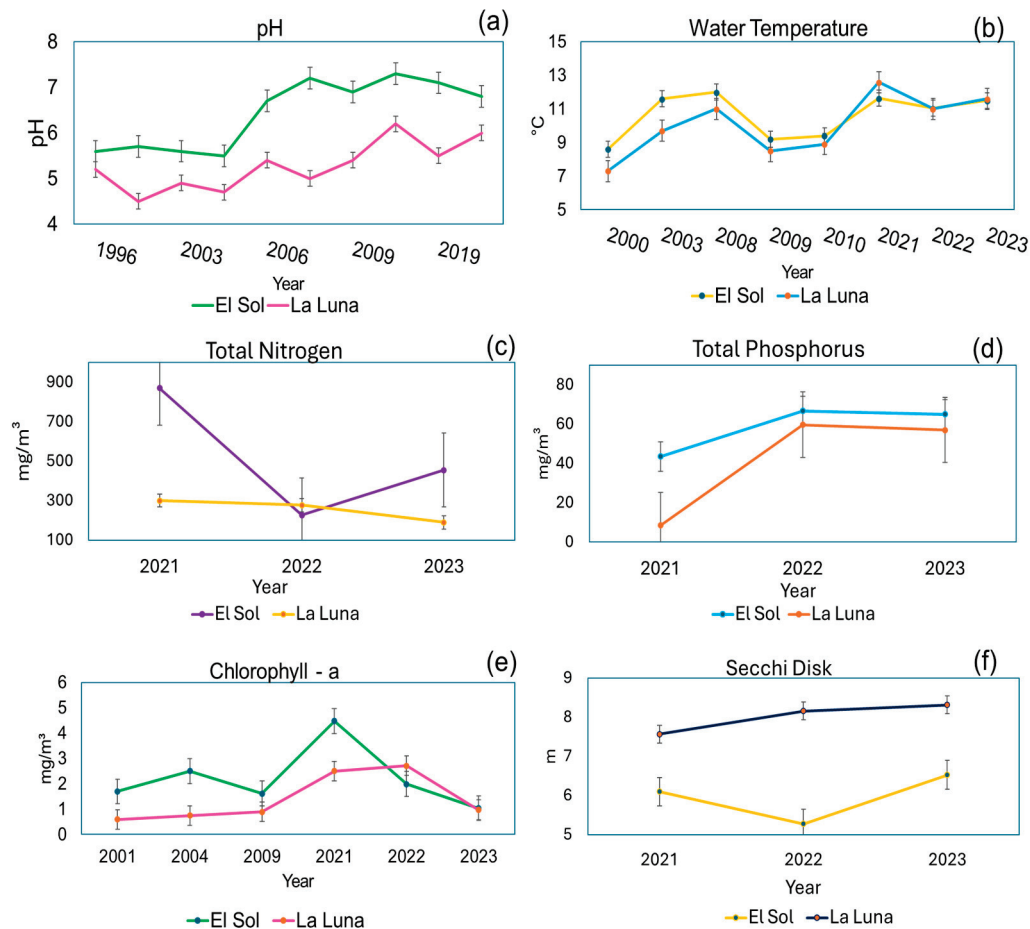
The pH increase observed in the studied lakes aligns with the findings of Ibarra et al. [39], who reported a rise in lake water pH in 2018. Similar trends have been observed in other high mountain lakes. For instance, lakes in the Tatra Mountains of southern Poland recorded pH values of 7.13 (Morskie Oko), 7.23 (Przedni Staw Polski), and 6.49 (Czarny Staw Gąsienicowy) [40]. Likewise, lakes in the Andean plateau showed pH values ranging between 6.8 and 7.5 [41]. Additionally, Muñoz-López and Rivera-Rondón [42] found that eastern mountain lakes in Colombia exhibited a pH range of 4.7–8.5. However, these lakes were not classified as altered water bodies.

Regarding the water temperature (WT) in the lakes, significant changes were observed during the 2021–2023 period, with notable differences between seasons and across years ( $p < 0.05$ ). The average WT remained below  $<12^{\circ}\text{C}$ , with the highest WT recorded during the wet season.

Historical data in Figure 5b indicate that WT has increased by 3–5% over the years, with recorded temperatures of  $<9^{\circ}\text{C}$  in the year 2000. Figure 5b also highlights that Lake La Luna experienced the most significant temperature increases, which were linked to its smaller size. However, on a broader scale, both lakes showed WT alterations compared to past decades.

The electrical conductivity (EC) values showed significant differences between the lakes, with higher concentrations in Lake El Sol. However, seasonal variations did not significantly impact the EC values ( $p < 0.05$ ), indicating that EC may be influenced by

historical pollution. Similarly, the total solids (TS) parameter showed significant variations between lakes and across seasons, with higher concentrations recorded in Lake El Sol and during the wet season ( $p < 0.05$ ). The increased TS levels were associated with pollen, salts, and algae.



**Figure 5.** Historical changes in the quality parameters of lake water. Past decades and the 2021–2023 period. (a) pH, (b) water temperature, (c) total nitrogen, (d) total phosphorus, (e) chlorophyll-a, and (f) Secchi disk.

The total nitrogen (TN) levels, the data in Table 2, and Figure 5c show that TN was higher in Lake El Sol than in La Luna ( $p < 0.05$ ), recording maximum values during the wet season at 680 mg/m<sup>3</sup>, while Lake La Luna recorded 450 mg/m<sup>3</sup>.

The TN levels in Lake El Sol indicated a high presence of algae [38], with a rising concentration trend approaching 1 g/m<sup>3</sup>, while Lake La Luna exhibited a TN trend of 0.5 g/m<sup>3</sup>. However, historical TN data are unavailable.

In comparison with other high mountain lakes, the TN concentrations in Lakes El Sol and La Luna were like those in Lake Tahoe (500 mg/m<sup>3</sup>) [43]. In contrast, lakes in the Andean–Patagonia region and Lake Cavihue recorded lower TN values (80 mg/m<sup>3</sup>) [44], indicating low pollution.

The indices used to assess eutrophication do not include TN (Table 1), but this parameter is important because it defines the concentration of nutrients and ammoniacal nitrogen in the aquatic ecosystem.

Regarding the eutrophic parameter in Table 1, the total phosphorus (TP) values in Lakes El Sol and La Luna showed significant differences across seasons and study years



(Figure 5d). The highest TP concentrations were observed during the cold season. Lake El Sol registered an average TP concentration of  $66 \text{ mg/m}^3$ , and Lake La Luna recorded  $57 \text{ mg/m}^3$ . The most notable TP changes occurred in 2022.

Historical TP data are not available; however, Table 1 shows an average TP range of  $10\text{--}66 \text{ mg/m}^3$ , indicating eutrophic conditions based on the Carlson and OECD indices, with a mesotrophic classification for Lakes El Sol and La Luna.

Other high mountain lakes show varying TP levels. For instance, lakes in the Tatra Mountains of Poland recorded TP values below  $10 \text{ mg/m}^3$ , reflecting clean water conditions. Morskie Oko recorded  $8 \text{ mg/m}^3$ , while Czarny Staw Gąsienicowy recorded  $7 \text{ mg/m}^3$ . However, TP values above  $10 \text{ mg/m}^3$  were identified as promoters of aquatic algae proliferation [40].

Muñoz-López and Rivera-Rondón [42] observed a high TP range of  $19\text{--}20.4 \text{ mg/m}^3$  in lakes in the eastern mountain range of Colombia. Lake Tianchi, located in the mountains of Central Asia, also reported TP fluctuations between  $13$  and  $24 \text{ mg/m}^3$ , while Lake Yeguas reported  $14.4 \text{ mg/m}^3$ , and Lake Virgen reported  $34.3 \text{ mg/m}^3$ . These values, together with high ultraviolet radiation, fostered the development of viruses and bacteria [43].

In studies of Andean lakes, TP levels also varied. Lake Negra recorded  $50 \text{ mg/m}^3$ , while Green and Blue Lakes recorded  $>12 \text{ mg/m}^3$  [41]. During the summer, Lake Tianchi exhibited a eutrophic state of  $<12 \text{ mg/m}^3$  [12].

Regarding water transparency, measured using a Secchi disk (SD), seasonal changes were observed between Lakes El Sol and La Luna. Lake El Sol had an SD range of  $7\text{--}9 \text{ m}$ , while Lake La Luna recorded  $5\text{--}7 \text{ m}$ . The minimum value was observed in March 2021 (dry season) for both lakes. The SD results indicated that the lakes' water transparencies fall within the threshold between clean and polluted water [45].

For the eutrophic parameter Chl-a, the concentrations varied significantly between Lakes El Sol and La Luna, as well as across seasons ( $p < 0.05$ ). Lake El Sol recorded its highest Chl-a values during the wet season, with concentrations below  $<3.1 \text{ mg/m}^3$ , while La Luna had a maximum of below  $<2.0 \text{ mg/m}^3$ .

Figure 5d illustrates Chl-a variations from 2004 to 2021–2023, showing notable increases between 2009 and 2022. La Luna's Chl-a levels increased in 2021–2022, reflecting seasonal variations.

According to the latest report by Alcocer et al. [17], the Chl-a range in Lake El Sol was previously  $0.14\text{--}1.62 \text{ mg/m}^3$ , while in Lake La Luna, it was  $0.13\text{--}0.89 \text{ mg/m}^3$ . This indicates that Chl-a values have doubled over the past 15 years.

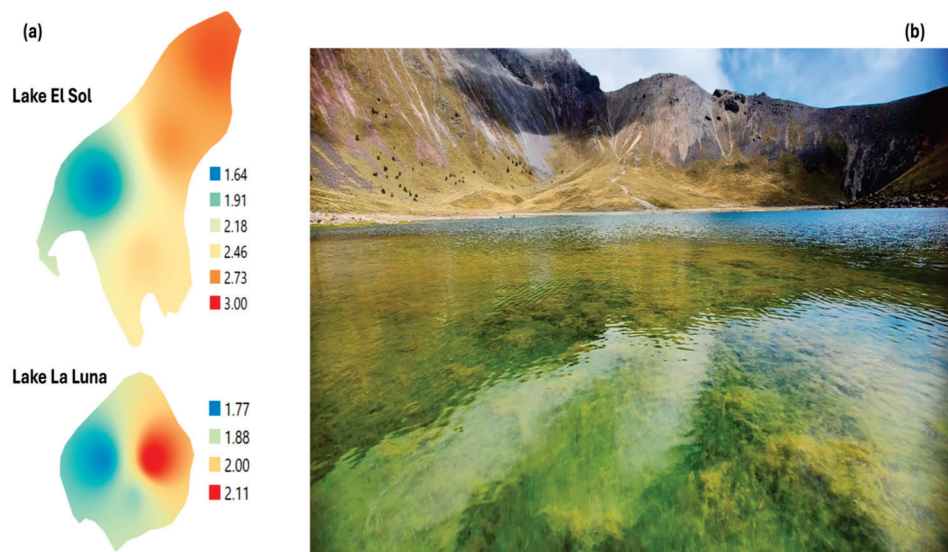
Similar Chl-a ranges have been recorded in Polish lakes, with concentrations between  $2.5$  and  $8.4 \text{ mg/m}^3$ , reflecting eutrophication alterations [40]. Andean lakes also showed high Chl-a ranges, reaching up to  $12.80 \text{ mg/m}^3$  [41].

In contrast, lower Chl-a concentrations were observed in Tianshan Mountain lakes in Central Asia, with values ranging from  $2.20$  to  $4.48 \text{ mg/m}^3$  and an annual average of  $3.3 \text{ mg/m}^3$  [12]. Similarly, Lake Tahoe recorded  $0.5 \text{ mg/L}$  [43], while the Andean–Patagonia region exhibited values of  $0.08 \text{ mg/L}$  [44], indicating minimal lake alterations.

#### 3.4. Trophic Classification of Lakes El Sol and La Luna

The eutrophic index of Carlson and OECD (Table 1) indicates that the current Chl-a values classify Lake El Sol as mesotrophic category, while Lake La Luna falls within the oligotrophic category, transitioning to mesotrophic. This suggests that both lakes are undergoing eutrophic changes in the present period [46].

Figure 6 highlights the primary sites affected by Chl-a alterations in the lakes. In Lake El Sol, the most impacted sites were Site 1 and Site 2, with Chl-a concentrations reaching  $3 \text{ mg/m}^3$ , whereas Site 4 recorded the lowest Chl-a concentration ( $1.6 \text{ mg/m}^3$ ).



**Figure 6.** (a) Distribution of impacted sites in Lakes El Sol and La Luna by Chl-a. (b) Image of Lake El Sol water showing high levels of algae proliferation.

In order, La Luna has the highest values of Chl-a in Sites 1 and 3 ( $2\text{--}1.7 \text{ mg/m}^3$ ); however, Site 2 contains low levels of Chl-a ( $1 \text{ mg/m}^3$ ).

Matching Table 1 with the Carlson index ( $<30$ ), the SD range in Lakes El Sol and La Luna was  $12\text{--}8 \text{ m}$ , classifying both lakes as oligotrophic with a tendency toward mesotrophic conditions. However, according to the OECD index, an SD range of  $12\text{--}6 \text{ m}$  corresponds to the mesotrophic category, indicating that Lakes El Sol and La Luna are experiencing eutrophic modifications, trending toward mesotrophic conditions based on SD values.

Regarding the TP parameter, both the Carlson and OECD indices classify Lakes El Sol and La Luna as experiencing eutrophic conditions.

The study findings indicate that both lakes undergo eutrophic alterations characterized by high Chl-a and TP levels and low SD.

According to the Carlson and OECD indices, both lakes are classified as mesotrophic based on Chl-a and SD values, while TP levels place them in the eutrophic category, confirming ongoing eutrophication processes.

Based on the eutrophication study, the data from Lakes El Sol and La Luna were similar to Lake Shihmen, which recorded TP levels of  $14\text{--}25 \text{ mg/m}^3$ , Chl-a concentrations of  $2.70\text{--}7.26 \text{ mg/m}^3$ , and SD values ranging from  $0.96$  to  $1.93 \text{ m}$ . Lake Shihmen was classified as experiencing eutrophic conditions [47].

Eutrophication studies of Lake Baikal reported TP levels of  $12 \text{ mg/m}^3$  and TN concentrations ranging from  $90$  to  $180 \text{ mg/m}^3$ , indicating algal proliferation. However, based on TP values, the Carlson index described Lake Baikal as oligotrophic with minor mesotrophic elements [48].

Similarly, Poniewozik and Lenard [49] analyzed Lakes Glinki and Gumienek, noting pH alterations ( $8.1$  in Glinki and  $8.4$  in Gumienek) and SD values between  $0.6$  and  $2.9 \text{ m}$ . However, high TN and TP values indicated a eutrophic state. Li et al. [50] applied the Carlson index to assess the trophic state of Lake Dawangtan, revealing seasonal variations

in eutrophication, particularly during summer. Noori et al. [51] recorded alterations in the Sabalan Reservoir, confirming a state of eutrophication, with TN levels ranging from 220 to 300 mg/m<sup>3</sup> and TP values between 110 and 380 mg/m<sup>3</sup>.

### 3.5. Pearson Correlations Between Variables Affecting the Eutrophication Status of Lake El Sol and La Luna

The influence of distinct variables on the eutrophic conditions of Lakes El Sol and La Luna is summarized in Tables 3–6, using the following Pearson correlations: (1) water quality parameters on eutrophic variables, (2) climatic conditions on eutrophic variables, (3) eutrophic variables on lake perimeters, and (4) climatic conditions on lake perimeters.

**Table 3.** Pearson correlation between water quality parameters and eutrophic variables in Lakes El Sol and La Luna.

| Lake    | Water Quality Parameters/Eutrophic Variables | pH    | EC    | TS    | WT    |
|---------|----------------------------------------------|-------|-------|-------|-------|
| El Sol  | Chl-a                                        | 0.91  | 0.86  | 0.48  | 1     |
|         | TN                                           | 1     | 0.83  | 0.86  | 0.83  |
|         | TP                                           | −0.27 | −0.18 | 0.50  | −0.70 |
|         | SD                                           | 0.74  | 0.66  | 0.19  | 0.97  |
| La Luna | Chl-a                                        | 0.15  | −0.86 | 1     | 0.36  |
|         | TN                                           | 0.61  | −1    | 0.92  | −0.14 |
|         | TP                                           | 0.63  | 0.21  | −0.60 | −0.93 |
|         | SD                                           | −0.92 | −0.86 | −0.60 | 0.61  |

**Table 4.** Pearson correlation between climatic conditions on eutrophic variables in Lakes El Sol and La Luna.

| Lake    | Climatic Conditions/Eutrophic Variables | Precipitation | Amb. Temp. | Evaporation |
|---------|-----------------------------------------|---------------|------------|-------------|
| El Sol  | Chl-a                                   | 0.97          | 0.92       | −0.41       |
|         | TN                                      | 0.95          | 0.82       | −0.58       |
|         | TP                                      | −0.46         | −0.90      | −0.42       |
|         | SD                                      | 0.86          | 0.99       | −0.11       |
| La Luna | Chl-a                                   | 0.83          | 0.32       | −0.95       |
|         | TN                                      | 0.45          | −0.18      | −0.98       |
|         | TP                                      | −0.97         | −0.92      | 0.41        |
|         | SD                                      | 0.05          | 0.64       | 0.76        |

**Table 5.** Pearson correlation between eutrophic variables and perimeters of Lakes El Sol and La Luna.

| Lake  | El Sol | La Luna |
|-------|--------|---------|
| Chl-a | −0.15  | 0.54    |
| SD    | 0.85   | 0.03    |
| TN    | 0.81   | 0.87    |
| TP    | 0.11   | −0.68   |

**Table 6.** Pearson correlation analysis between climatic conditions and lake perimeters.

| Lake Perimeter | Ambient Temperature | Precipitation | Evaporation |
|----------------|---------------------|---------------|-------------|
| El Sol         | −0.34               | 0.15          | −0.86       |
| La Luna        | −0.49               | −0.35         | 0.36        |

(1) The Pearson correlations between water quality parameters and eutrophic variables are displayed in Table 3.

The positive correlations indicate a strong relationship between water quality parameters and eutrophic variables ( $p < 0.05$ ). In Lake El Sol, the Chl-a concentration is primarily influenced by  $WT > pH > EC > TS$ . However, in Lake La Luna, Chl-a is mainly affected by TS.

In addition, TN is influenced by pH, while TP in Lake El Sol is affected by an increase in TS. In Lake La Luna, TP is affected by pH.

The water transparency (SD) in Lake El Sol is significantly affected by  $WT > pH > EC$ , whereas in Lake La Luna, WT is the primary influencing factor.

(2) The Pearson correlations between climatic conditions and eutrophication variables (Table 4) indicate significant correlations in both lakes, with varying degrees of influence ( $p < 0.05$ ).

Chl-a is strongly affected by changes in precipitation and ambient temperature, while evaporation does not affect Chl-a in either lake.

Total nitrogen (TN) is primarily impacted by precipitation, with a lesser effect of ambient temperature.

In terms of TP, climatic variables do not have a significant impact on Lake El Sol. In Lake La Luna, TP is partially affected by evaporation.

The SD in Lake El Sol is influenced by precipitation and ambient temperature, while in Lake La Luna, the SD is affected by ambient temperature and evaporation.

(3) The Pearson correlations between lake perimeter and eutrophication variables (Table 5) show different results for each lake ( $p < 0.05$ ).

Changes in the perimeter of Lake El Sol are strongly influenced by SD and TN, while changes in the perimeter of Lake La Luna are influenced by Chl-a and TN.

(4) The Pearson correlations between climatic variables and lake perimeter are shown in Table 6.

The results show that for both lakes, the correlation between climatic variables and lake perimeter was predominantly negative ( $p < 0$ ), indicating that climatic variables do not influence lake perimeter changes.

According to Pearson correlations, Lakes El Sol and La Luna are experiencing significant stress due to decreasing precipitation; however, the principal factor driving eutrophication is water pollution. Therefore, it is essential to develop conservation strategies and implement measures to reduce contamination and protect the lakes.

#### 4. Conclusions

This environmental study of Lakes El Sol and La Luna revealed that the most significant alterations occurred in Lake El Sol; however, Lake La Luna also exhibited environmental changes.

Compared to previous decades, both lakes showed a perimeter decrease. Lake El Sol showed a 38.3% decline in perimeter, along with changes in its hydrological form. Lake La Luna exhibited a 14.9% perimeter loss.

The water quality analysis of both lakes indicated pollution. Compared to recent decades, pH shifted from acidic conditions to a more alkaline tendency, with a recent range of 6–7.

The water temperature increased by more than 4°C compared to earlier years. Total solids (TS) levels indicated lake pollution, expressed as high total solids, whereas the EC reflected the presence of salts.

Water eutrophication parameters indicated eutrophic conditions, characterized by low SD and elevated levels of Chl-a, TN, and TP.



According to the Carlson index, OECD classification, and eutrophication parameter values (Chl-a, TP, and SD), Lake El Sol transitioned from a mesotrophic to a eutrophic state. Lake La Luna shifted from oligotrophic to mesotrophic conditions. Both lakes exhibited eutrophic changes during the wet season.

The Pearson correlations indicated that eutrophic conditions, including Chl-a, TN, TP, and SD, are strongly influenced by water quality parameters. In Lake El Sol, a symbiotic increase was observed between pH and Chl-a and pH with a reduction in SD. However, TP is principally affected by TS. In turn, Lake La Luna showed that Chl-a levels were affected by TS, while TP levels were primarily influenced by pH.

Among climatic variables, precipitation and ambient temperature were the principal climatic factors impacting eutrophic variables (Chl-a, SD, and TN). However, in Lake La Luna, TP levels were affected by evaporation.

The relationship between climatic variables and lake perimeter indicated that they did not significantly impact the perimeter of the lakes. However, the eutrophic conditions of SD and TN impacted the perimeter of Lake El Sol. The perimeter of Lake La Luna was primarily impacted by Chl-a and TN.

The findings from this study indicate that eutrophic changes in Lakes El Sol and La Luna are strongly linked to water pollution. However, continuous monitoring is necessary, and more research is required to reduce uncertainty around the present analysis. Further studies should be focused on collecting data on eutrophication and water quality parameters, increasing historical climatic data to better understand their impact on eutrophication, and analyzing the predominant species in the lakes that are associated with eutrophication.

As a result of this study, lake conservation and eutrophication management strategies should pay attention to pollution reduction.

**Author Contributions:** Conceptualization, C.M. and F.G.G.-M.; methodology, J.L.E.-C.; software, J.L.E.-C.; validation, C.M. and Y.A.; formal analysis, J.L.E.-C. and H.V.C.-B.; investigation, Y.A., F.G.G.-M. and C.M.; data curation, C.M.; writing—original draft preparation, F.G.G.-M.; writing—review and editing, C.M., J.L.E.-C. and H.V.C.-B.; supervision, C.M.; project administration, C.M.; funding acquisition, C.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** Tecnológico Nacional de México supported this study. Grant numbers TECNM 41222010 and TECNM 41571463.

**Data Availability Statement:** Data are contained within this article.

**Acknowledgments:** The authors thank CONANP for the facilities provided to carry out research in the protected area and Lakes El Sol and la Luna.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Shumka, S.; Shumka, L.; Špoljar, M.; Shuka, L. Evidence of Climate Change and the Conservation Needed to Halt the Further Deterioration of Small Glacial Lakes. *Climate* **2024**, *12*, 124. [CrossRef]
2. Tran, T.N.D.; Tapas, M.R.; Do, S.K.; Etheridge, R.; Lakshmi, V. Investigating the impacts of climate change on hydroclimatic extremes in the tar-Pamlico River basin, North Carolina. *J. Environ. Manag.* **2024**, *363*, 121375. [CrossRef] [PubMed]
3. Guan, G.; Wu, L.; Li, Y.; Ma, B.; Liu, Y.; Zhao, C.; Wang, Z.; Zhao, Y. Evaluating the Impacts of Fertilization and Rainfall on Multi-Form Phosphorus Losses from Agricultural Fields: A Case Study on the North China Plain. *Agronomy* **2024**, *14*, 1922. [CrossRef]
4. Tapas, M.R.; Etheridge, R.; Tran, T.-N.-D.; Finlay, C.G.; Peralta, A.L.; Bell, N.; Xu, Y.; Lakshmi, V. A methodological framework for assessing sea level rise impacts on nitrate loading in coastal agricultural watersheds using SWAT+: A case study of the Tar-Pamlico River basin, North Carolina, USA. *Sci. Total Environ.* **2024**, *951*, 175523. [CrossRef]

5. Tapas, M.R.; Etheridge, R.; Howard, G.; Mair, M. Framework for Stakeholder-Driven Socio-Hydrological Modeling: Conceptual Foundations for Policy Development and Evaluation to Improve Ecosystem Health. In *Hydrology and Hydrologic Modelling*; Pandey, M., Umamahesh, N., Das, J., Pu, J.H., Eds.; HYDRO 2023; Lecture Notes in Civil Engineering; Springer: Singapore, 2025; p. 410. [CrossRef]
6. Chen, J.; Liu, X.; Chen, J.; Jin, H.; Wang, T.; Zhu, W.; Li, L. Underestimated nutrient from aquaculture ponds to Lake Eutrophication: A case study on Taihu Lake Basin. *J. Hydrol.* **2024**, *630*, 130749. [CrossRef]
7. Leng, M.; Wu, X.; Ge, X.; Yang, X.; Huang, Z.; Liu, H.; Zhu, J.; Li, J.; Gong, M.; Sun, Z.; et al. Evaluation of Water Quality and Eutrophication of Typical Lakes in Southeast Hubei, China. *Sustainability* **2024**, *16*, 8964. [CrossRef]
8. Pastorino, P.; Elia, A.C.; Pizzul, E.; Bertoli, M.; Renzi, M.; Prearo, M. The old and the new on threats to high-mountain lakes in the Alps: A comprehensive examination with future research directions. *Ecol. Indic.* **2024**, *160*, 111812. [CrossRef]
9. Kuczynski, A.; Smith, R.G.; Fraser, C.E.; Larned, S.T. Environmental indicators of lake ecosystem health in Aotearoa New Zealand: Current state and trends. *Ecol. Indic.* **2024**, *165*, 112185. [CrossRef]
10. Magyari, E.K.; Szabó, Z.; Falus, G.; Mora, A.; Szalai, Z.; Hamerlik, L.; Tóth, M.; Farkas, Á.; Pomogyi, P.; Somogyi, B.; et al. Large shallow lake response to anthropogenic stressors and climate change: Missing macroinvertebrate recovery after oligotrophication (Lake Balaton, East-Central Europe). *Sci. Total Environ.* **2024**, *946*, 174191. [CrossRef]
11. Zuo, J.; Xiao, P.; Heino, J.; Tan, F.; Soininen, J.; Chen, H.; Yang, J. Eutrophication increases the similarity of cyanobacterial community features in lakes and reservoirs. *Water Res.* **2024**, *250*, 120977. [CrossRef]
12. Han, F.; Liu, T.; Huang, Y.; Zan, C.; Pan, X.; Xu, Z. Response of water quality to climate warming and atmospheric deposition in an alpine lake of Tianshan Mountains, Central Asia. *Ecol. Indic.* **2023**, *147*, 109949. [CrossRef]
13. Han, W.; Zhang, E.; Sun, W.; Lin, Q.; Meng, X.X.; Ni, Z.; Ning, D.; Shen, J. Anthropogenic activities altering the ecosystem in Lake Yamzhog Yumco, southern Qinghai-Tibetan Plateau. *Sci. Total Environ.* **2023**, *904*, 166715. [CrossRef]
14. Poikane, S.; Kelly, M.; Várbiro, G.; Borics, G.; Erős, T.; Hellsten, S.; Kolada, A.; Lukács, B.; Lyche-Solheim, A.; Pahissa, J.; et al. Estimating nutrient thresholds for eutrophication management: Novel insights from understudied lake types. *Sci. Total Environ.* **2022**, *827*, 154242. [CrossRef] [PubMed]
15. Jacquemin, C.; Bertrand, C.; Franquet, E.; Mounier, S.; Misson, B.; Oursel, B.; Cavalli, L. Effects of catchment area and nutrient deposition regime on phytoplankton functionality in alpine lakes. *Sci. Total Environ.* **2019**, *674*, 114–127. [CrossRef]
16. Barba-López, M.d.R.; Filonov, A.; Alcocer, J.; Ladah, L.; Tereshchenko, I. Complex interactions between meteorological parameters and the thermal regime of two tropical, high altitude crater lakes (Nevado de Toluca, México). *Sci. Total Environ.* **2021**, *771*, 145310. [CrossRef]
17. Alcocer, J.; Oseguera, L.A.; Escobar, E.; Peralta, L.; Lugo, A. Phytoplankton Biomass and Water Chemistry in Two High-Mountain Tropical Lakes in central Mexico. *Arct. Antarct. Alp. Res.* **2004**, *36*, 342–346. [CrossRef]
18. Cuna, E.; Caballero, M.; Ruíz, A.C.; Lozano, S.; Alcocer, J. Environmental impacts of Little Ice Age cooling in central Mexico recorded in the sediments of a tropical alpine lake. *J. Paleolimnol.* **2014**, *51*, 11–14. [CrossRef]
19. Filonov, A.; Barba-Lopez, M.d.R.; Tereshchenko, I.; Ladah, L.B.; Pantoja, D.A.; Alcocer, J.; Velazquez-Muñoz, F.A. Thermal regime and water balance of two tropical high-mountain lakes in the Nevado de Toluca Volcano, Mexico. *Water* **2022**, *14*, 1104. [CrossRef]
20. Ibarra-Morales, D.; Alcocer, J.; Oseguera, L.; Sosa-Echeverría, R. Bulk Deposition and Main Ionic Composition in a Remote Tropical Region: Nevado de Toluca, México. *Water Air Soil Pollut.* **2020**, *231*, 413. [CrossRef]
21. Alcocer, J.; Oseguera, L.; Ibarra-Morales, D.; Escobar, E.; García-Cid, L. Responses of Benthic Macroinvertebrate Communities of Two Tropical, High-Mountain Lakes to Climate Change and Deacidification. *Diversity* **2021**, *13*, 243. [CrossRef]
22. Cuna, E.; Alcocer, J.; Gaytan, M.; Caballero, M. Phytoplankton Biodiversity in Two Tropical, High Mountain Lakes in Central Mexico. *Diversity* **2022**, *14*, 42. [CrossRef]
23. Fernández, R.; Alcocer, J.; Oseguera, L.; Zuñiga-Ramos, C.; Vilaclara, G. Phytoplankton Communities' Seasonal Fluctuation in Two Neighboring Tropical High-Mountain Lakes. *Plants* **2024**, *12*, 3021. [CrossRef] [PubMed]
24. Alcocer, J.; Roberson, J.; Oseguera, L.; Lewis, W. Rhythmic episodes of heating and cooling control thermal stratification of two tropical high mountain lakes. *Aquat. Sci.* **2020**, *82*, 58. [CrossRef]
25. García-Solorio, L.; Muro, C.; De La Rosa, I.; Amador-Muñoz, O.; Ponce-Vélez, G. Organochlorine pesticides and polychlorinated biphenyls in high mountain lakes, Mexico. *Environ. Sci. Pollut. Res. Int.* **2022**, *29*, 49291–49308. [CrossRef]
26. Muro, C.; Sandoval, J.; García, F.G.; Andrade, M.; del Diaz-Nava, M.C.; Alvarado, Y. Evaluation of environmental risks by heavy metals pollution and algae accumulations in two alpine lakes. *Chem. Ecol.* **2024**, *41*, 194–214. [CrossRef]
27. Soto, V.; Yoshikawa, K.; Schörghofer, N. Climatic variation in the high mountains of central Mexico: Temperature and precipitation indices at Nevado de Toluca volcano. *Atmósfera* **2020**, *33*, 301–318. [CrossRef]
28. Armienta, M.A.; De la Cruz-Reyna, S.; Macías, J. Chemical characteristics of the crater lakes of Popocatepetl, El Chichon, and Nevado de Toluca volcanoes, Mexico. *J. Volcanol. Geotherm. Res.* **2000**, *97*, 105–125. [CrossRef]

29. Standard Methods Committee of the American Public Health Association; American Water Works Association; Water Environment Federation. 10150 chlorophyll-a. In *Standard Methods for the Examination of Water and Wastewater*, 23rd ed.; Lipps, W.C., Baxter, T.E., Braun-Howland, E., Eds.; APHA Press: Washington DC, USA, 2023. [CrossRef]
30. Mexican Standard NMX-AA-026-SCFI-2010; Water Analysis—Measurement of Total Kjeldahl Nitrogen in Natural, Wastewater and Treated Wastewater. Official Journal of the Federation: Ciudad de México, Mexico, 2010.
31. Mexican Standard NMX-AA-029-SCFI-2001; Water Analysis—Determination of Total Phosphorus in Natural, Wastewater and Treated Wastewater. Official Journal of the Federation: Ciudad de México, Mexico, 2001.
32. Pantoja, D.; Vega-Álvarez, N.; Gasca-Ortiz, T. Trophic state in a tropical lake based on Chlorophyll-a profiler data and Sentinel-2 images: The onset of an algal bloom event. *Water Environ. Res.* **2021**, *93*, 2185–2197. [CrossRef]
33. Lin, J.L.; Karangan, A.; Huang, Y.; Kang, S. Eutrophication factor analysis using Carlson trophic state index (CTSI) towards non-algal impact reservoirs in Taiwan. *Sustain. Environ. Res.* **2022**, *32*, 25. [CrossRef]
34. Carlson, R.E. A trophic state index for lakes. *Limnol. Oceanogr.* **1977**, *22*, 361–369. [CrossRef]
35. Sánchez, M.R.; Reyes, N.; Lugo, A.; Alcocer, J.; Oseguera, L.; Peralta, L. Bacterioplankton variation in two high mountain Mexican lakes. *Nat. Sci.* **2004**, *7*, 69–83.
36. Armienta, M.A.; Vilaclara, G.; De la Cruz, S.; Ramos, S.; Cenicerros, N.; Cruz, O.; Aguayo, A.; Arcega, F. Water chemistry of lakes related to active and inactive Mexican volcanoes. *J. Volcanol. Geotherm. Res.* **2008**, *178*, 249–258. [CrossRef]
37. Moldan, F.; Stadmark, J.; Jutterström, S.; Kronnäs, V.; Blomgren, H.; Cosby, B. MAGIC library—A tool to assess surface water acidification. *Ecol. Indic.* **2020**, *112*, 106038. [CrossRef]
38. Kumar, S.; Shrivastava, S.; Tigga, V.; Lakra, M. Review paper on the role of pH in water quality implications of aquatic life, Human Health, and Environmental Sustainability. *IARJSET* **2023**, *10*, 215–218. [CrossRef]
39. Ibarra-Morales, D.; Silva, R.; Oseguera, L.; Merino, M.; Alcocer, J. Impacts of global change on two tropical, high mountain lakes in Central Mexico. *Sci Total Environ.* **2022**, *852*, 158521. [CrossRef]
40. Sienkiewicz, E.; Gasiński, M. Changes in the Trophic Status of Three Mountain Lakes—Natural or Anthropogenic Process? *Pol. J. Environ. Stud.* **2014**, *23*, 875–892.
41. Ordoñez, O.; Flores, M.; Dib, J.; Paz, A.; Farías, M. Extremophile Culture Collection from Andean Lakes: Extreme Pristine Environments that Host a Wide Diversity of Microorganisms with Tolerance to UV Radiation. *Microb. Ecol.* **2009**, *58*, 461–473. [CrossRef]
42. Muñoz-López, C.L.; Rivera-Rondón, C.A. Diatom response to environmental gradients in the high mountain lakes of the Colombia's Eastern Range. *Aquat. Sci.* **2022**, *84*, 15. [CrossRef]
43. Rose, K.; Williamson, C.; Schladow, G.; Winder, M.; Oris, J. Patterns of spatial and temporal variability of UV transparency in Lake Tahoe, California-Nevada. *J. Geophys. Res-Bioge.* **2009**, *114*, G00D03. [CrossRef]
44. Domagalski, J.; Morway, E.; Álvarez, N.; Hutchins, J.; Rosen, M.; Coats, R. Trends in nitrogen, phosphorus, and sediment concentrations and loads in streams draining to Lake Tahoe, California, Nevada, USA. *Sci. Total Environ.* **2021**, *752*, 1–17. [CrossRef]
45. Diaz, M.; Pedrozo, F.; Reynolds, C.; Temporetti, P. Chemical composition and the nitrogen-regulated trophic state of Patagonian lakes. *Limnologia* **2007**, *37*, 17–27. [CrossRef]
46. Cavallini, E.; Viaroli, P.; Naldi, M.; Saccò, M.; Scibona, A.; Barbieri, E.; Franceschini, S.; Nizzoli, D. Seasonal Variability and Hydrological Patterns Influence the Long-Term Trends of Nutrient Loads in the River Po. *Water* **2024**, *16*, 2628. [CrossRef]
47. Liao, W.; Chen, H.; Peng, M.; Chang, T. Spatio-Temporal Variation of Trophic Status and Water Quality with Water Level Fluctuation in a Reservoir. *Water* **2023**, *15*, 3154. [CrossRef]
48. Domysheva, V.; Vorobyeva, S.; Golobokova, L.; Netsvetaeva, O.; Onischuk, N.; Sakirko, M.; Khuriganovay, O.; Fedotov, A. Assessment of the Current Trophic Status of the Southern Baikal Littoral Site. *Water* **2023**, *15*, 1139. [CrossRef]
49. Poniewozik, M.; Lenard, T. Phytoplankton Composition and Ecological Status of Lakes with Cyanobacteria Dominance. *Int. J. Environ. Res. Public Health* **2022**, *19*, 3832. [CrossRef]
50. Li, R.; Xiao, K.; Zhao, G.; Huang, X.; Li, Z.; Wu, H.; Huang, X.; Pan, Y.; Liang, L. Comprehensive Assessment of Eutrophication and the Mechanisms Driving Phytoplankton Blooms in Multifunctional Reservoirs. *Water* **2024**, *16*, 1752. [CrossRef]
51. Noori, R.; Ansari, E.; Jeong, Y.-W.; Aradpour, S.; Maghrebi, M.; Hosseinzadeh, M.; Bateni, S.M. Hyper-Nutrient Enrichment Status in the Sabalan Lake, Iran. *Water* **2021**, *13*, 2874. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



## Article

# Monitoring and Disaster Assessment of Glacier Lake Outburst in High Mountains Asian Using Multi-Satellites and HEC-RAS: A Case of Kyagar in 2018

Long Jiang <sup>1</sup>, Zhiqiang Lin <sup>1,\*</sup>, Zhenbo Zhou <sup>2</sup>, Hongxin Luo <sup>1</sup>, Jiafeng Zheng <sup>1</sup>, Dongsheng Su <sup>1</sup> and Minhong Song <sup>1</sup>

<sup>1</sup> School of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu 610225, China; 3230101073@stu.cuit.edu.cn (L.J.); 3230101049@stu.cuit.edu.cn (H.L.); zjf1988@cuit.edu.cn (J.Z.); sds@cuit.edu.cn (D.S.); songmh@cuit.edu.cn (M.S.)

<sup>2</sup> Tibet Institute of Plateau Atmospheric and Environmental Science, Meteorological Bureau of Tibet Autonomous Region, Lhasa 850000, China; zhouzb0891@aliyun.com

\* Correspondence: linzq@cuit.edu.cn

**Abstract:** The glaciers in the High Mountain Asia (HMA) region are highly vulnerable to global warming, posing significant threats to downstream populations and infrastructure through glacier lake outburst floods (GLOFs). The monitoring and early warnings of these events are challenging due to sparse observations in these remote regions. To explore reproducing the evolution of GLOFs with sparse observations in situ, this study focuses on the outburst event and corresponding GLOFs in August 2018 caused by the Kyagar Glacier lake, a typical glacier lake of the HMA in the Karakoram, which is known for its frequent outburst events, using a combination of multi-satellite remote sensing data (Sentinel-1 and Sentinel-2) and the HEC-RAS hydrodynamic model. The water depth of the glacier lake and downstream was extracted from satellite data adapted by the Floodwater Depth Elevation Tool (FwDET) as a baseline to compare them with simulations. The elevation-water volume curve was obtained by extrapolation and was applied to calculate the water surface elevation (WSE). The inundation of the downstream of the lake outburst was obtained through flood modeling by incorporating a load elevation-water volume curve and the Digital Elevation Model (DEM) into the hydrodynamic model HEC-RAS. The results showed that the Kyagar glacial lake outburst was rapid and destructive, accompanied by strong currents at the end of each downstream storage ladder. A series of meteorological evaluation indicators showed that HEC-RAS reproduced the medium and low streamflow rates well. This study demonstrated the value of integrating remote sensing and hydrodynamic modeling into GLOF assessments in data-scarce regions, providing insights for disaster risk management and mitigation.

**Keywords:** Kyagar Glacier; Glacier lake; remote sensing; high mountain Asia; HEC-RAS

## 1. Introduction

High-latitude and mountainous regions are characterized by cold climates that have led to the formation of extensive glaciers, serving as critical freshwater sources for agriculture, domestic use, and hydropower in many countries. They serve as a critical resource for both livelihoods and economic activities, especially in arid and semi-arid regions, where the resulting oases from glacier meltwater are often the only sources of water for human settlement [1].

In recent decades, global warming (GW), driven by industrial activities, has led to significant, increasing, and extreme weather events such as heatwaves, droughts, floods, and storms. Glaciers are particularly vulnerable to GW in ecologically sensitive areas, such as high mountains. As a result of climate change, glaciers are rapidly shrinking, with significant loss of mass worldwide, which has exacerbated freshwater shortages [2]. The response of glaciers and their decline in the face of GW varies by region due to their



complex terrains and regionally climatic conditions. For example, glaciers in Central Europe have experienced significant ice mass loss due to the small sizes of glaciers and insufficient snowfall replenishment; many have already disappeared in these regions, and most will disappear by the end of the 21st century [3,4]. However, glaciers in the Tibetan Plateau have exhibited diverse trends, with some regions experiencing expansion and others undergoing rapid shrinkage. In China, the amount of areas covered by glacier lakes has expanded significantly since the 1990s; glacier-fed lakes have expanded approximately six times faster than non-glacier-fed lakes. The rapid expansion of these lakes is closely tied to rising temperatures and the retreating of glaciers [5]. The Tibetan Plateau, with over 40,000 glaciers and more than 16,000 glacier lakes, has undergone dramatic changes. A total of 369 lakes have been classified as potentially dangerous based on the identification index system of glacier lake outbursts [6]. In response to GW, the surface albedo of the Tibetan Plateau had decreased over the past few decades [7,8] and increased in terms of glacier volume loss in the northern and central parts, while the southeastern regions have remained more stable or even experienced slight expansion [9]. Due to the retreating of glaciers and surging, the formation of glacier lakes plays a dual role in both the storing of freshwater and posing significant natural hazards, such as flooding, landslides, and debris flows, which have made them a critical focus of research into glacier-related hazards and disaster risk management.

Research on glacier lakes is inherently complex, as lake evolution varies by region and is influenced by timescales pertaining to glacier surging, which can range from hours to centuries. Under the influence of GW, glacier meltwater is increasingly blocked by glaciers, moraines, or bedrock, leading to the formation of glacier lakes [10]. GLOFs are sudden and catastrophic events, often triggered by glacier movement, ice dam failures, or other dynamic forces [10]. Numerous studies have investigated glacier lake outburst characteristics across various regions. In the Nyainqêntanglha Range of the Tibetan Plateau, two outburst events were analyzed using field measurements, satellite remote sensing, and hydrodynamic modeling to reveal their mechanisms and flow dynamics [11]. In the Doyinongba Basin of southeastern Tibet, remote sensing and numerical simulations were employed to study glacier lakes, showing how rapid debris influx from upstream glaciers triggered these outbursts and helped delineate hazard zones based on flood risk [12]. Given the sudden onset and destructive nature of GLOFs, these events often result in extensive loss of life and property downstream, evolving into debris flows, landslides, and other geohazards [13]. The increase in terms of the amount of glacier lakes and their size has drawn public attention to the growing risks of GLOFs since the 1990s. Populations downstream of these lakes are becoming increasingly vulnerable. On 4 October 2023, a GLOF occurred at South Lhonak Lake in the northwest of Sikkim, India, posing a severe threat to lives and property. By 5 October, at least 18 individuals had been confirmed dead, with an additional 26 injured and 102 missing, including 22 military personnel. The floodwaters also resulted in the destruction of 11 bridges [14]. Recent assessments of GLOF risks, incorporating lake conditions, exposure, and vulnerability data, have indicated that the High Mountain Asia (HMA) regions face the highest GLOF risk, while polar regions and remote areas have the lowest [14,15]. Therefore, monitoring glacier lakes in the HMA region and understanding outburst processes are critical for deepening our understanding of glacier behaviors and assessing glacier-related disaster risks.

Monitoring GLOFs in the HMA region is challenging because of the remote and high-altitude locations of glacier lakes, which make in situ observations both costly and dangerous. Moreover, the sudden nature of GLOF events, combined with the region's complex terrain, hinders data collection pertaining to the outburst process [16,17]. The rapid development of remote sensing technologies, however, has significantly enhanced GLOF research. Recent advancements in remote sensing technologies, such as the Normalized Difference Water Index (NDWI), have enabled the monitoring of glacier lake extents and dynamics [18]. For instance, the NDWI was used to track the maximum lake area of Merzbacher Lake in the years leading up to its outburst [19]. A combination of optical

remote sensing and SAR/InSAR techniques has been used to accurately monitor the spatiotemporal evolution of Jinwuco Glacier Lake in southeastern Tibet, revealing the factors that triggered its outburst [20]. However, satellite observations face limitations, including cloud cover and terrain shadows. In such cases, hydrodynamic models can complement remote sensing techniques by simulating water flow under various outburst scenarios. For example, hydrodynamic modeling was used to assess the impact of the Galong Co Glacier Lake outburst in downstream regions, demonstrating how deeper breach depths lead to greater downstream impacts [21]. Additionally, a combination of 1D and 2D hydrodynamic modeling was used to assess the flood risk posed by the potential outburst of Satopanth Lake in the central Himalayas [22]. Combining remote sensing data with hydrodynamic modeling offers a comprehensive approach to studying GLOFs. Remote sensing provides valuable input parameters for hydrodynamic models, ensuring that the simulated flood dynamics align with actual observations. This integrated approach enables high-resolution assessments of GLOF events, supporting disaster risk reduction in the HMA region. The Kyagar Glacier is a notable example of a glacier with a history of frequent outburst floods. Between 1961 and 2021, the Kyagar Glacier lake experienced 30 outburst floods, with August being the most common month for such events, followed by July [23]. One particularly notable event occurred on 4 September 1961, when a GLOF destroyed a bridge on the Yarkant River, causing direct economic losses of 1 million yuan. Outburst floods in 1982 and 1984 caused significant damage to downstream infrastructure, including the Karqun diversion project [24].

Numerous studies have concentrated on utilizing remote sensing data to investigate the Kyagar Glacier surge and the hazard resulting from the breach of the ice tongue; however, there is a paucity of scholarly work that conducts dynamic analyses of inundation events in the downstream areas. By matching model simulations with real-world observations under different flow scenarios, this research aims to contribute to the assessment of GLOF hazards and provide a method for estimating flood discharge in areas where flow monitoring stations are difficult to establish. Section 2 describes the study area, data sources, core methodologies, and operational techniques, with a particular emphasis on depth extraction methods based on satellite data. Additionally, this section outlines the fundamental principles and specific parameter settings employed in GLOF modeling. Section 3.1 presents the water depth data obtained through direct inversion using remote sensing data, establishing a foundational dataset for the GLOF hydrodynamic model. Section 3.2 illustrates the hydrodynamic model's representation of the GLOF event, examines the hazard characteristics revealed by this model, and provides an evaluation based on remote sensing data. Chapter 4 discusses the study's limitations and considers future applications in hazard assessment, disaster prevention, and early warning systems for glacial lake outbursts. Finally, Chapter 5 summarizes the key findings of this study.

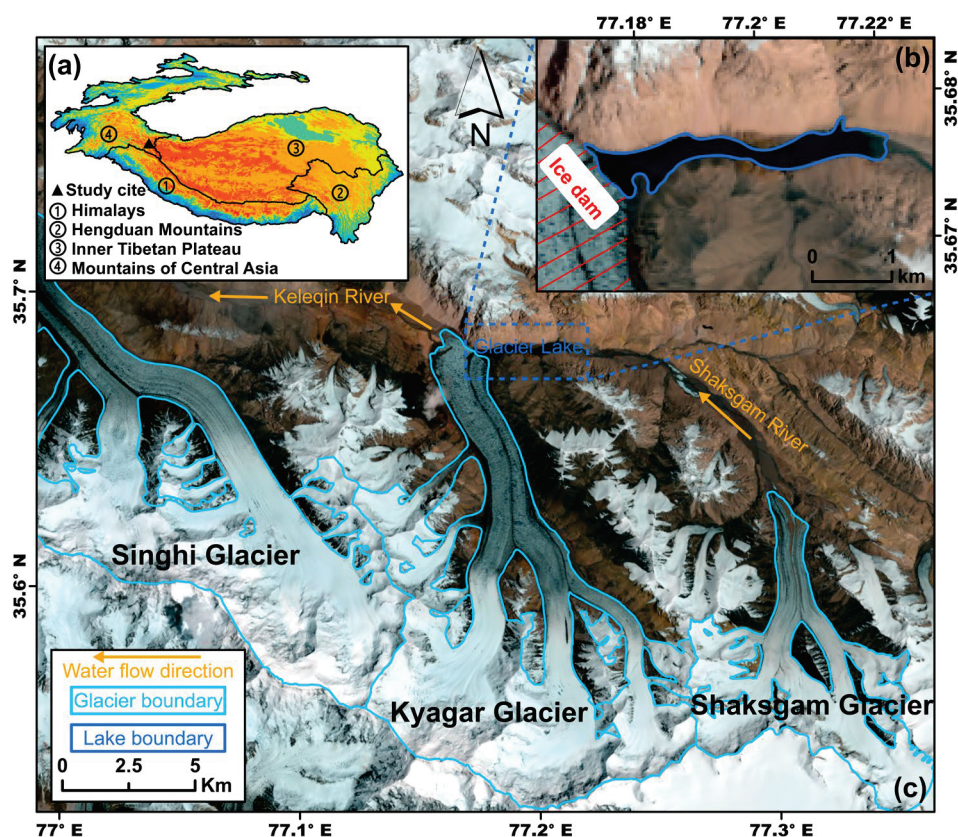
## 2. Materials and Methodology

### 2.1. Study Site

Kyagar Glacier is located on the northern slopes of the Karakoram Range and next to Shaksgam Valley, playing a significant role in the Yarkant River's flow path [25]. Due to its high altitude and cold climate, this area is ideal for glacier formation and movement. The slow northward movement of the Kyagar Glacier results in the accumulation of an ice tongue on its northeastern side within the Shaksgam Valley [26]. This ice tongue, with a height ranging from 50 to 150 m and a width of about 1.5 km, plays the role of a reservoir dam and blocks the valley's riverbed, forming a glacier lake with a storage capacity exceeding 230 MCM (million cubic meters) [27].

The coordinates of the glacier lake are approximately 35.67°–35.68°N and 77.17°–77.22°E (Figure 1a). The glacier tongue extends 8 km, with a width of 1.5 km. The tongue surface is rugged, with a significant spatial span and an average slope of more than 2°. These factors contribute to the formation of a natural ice dam that blocks the valley's riverbed and enables the accumulation of water on the eastern side of the glacier tongue [25,28]. During

non-outburst periods, the ice tongue acts as a partial barrier. The glacier's complex system of crevasses and fluvial channels, located under the northern side of the ice tongue, acts as a drainage pathway, connecting the glacier lake to the downstream Keleqin River [28]. When the hydrological pressure in these crevasses reaches its peak, large-scale drainage events occur, which were the primary mechanism behind outburst events before 2017 [26]. After 2017, the increased mass of the Kyagar Glacier tongue enhanced its blocking ability, resulting in the 5th highest of all records of WSE in 2018 of the Kyagar Glacier lake [29]. The thickening of the ice tongue overwhelmed existing drainage channels, causing them to close. Meanwhile, erosion on the northern side of the ice tongue created a stable flow channel, which became one of the primary drainage pathways during the 2018 outburst [26].

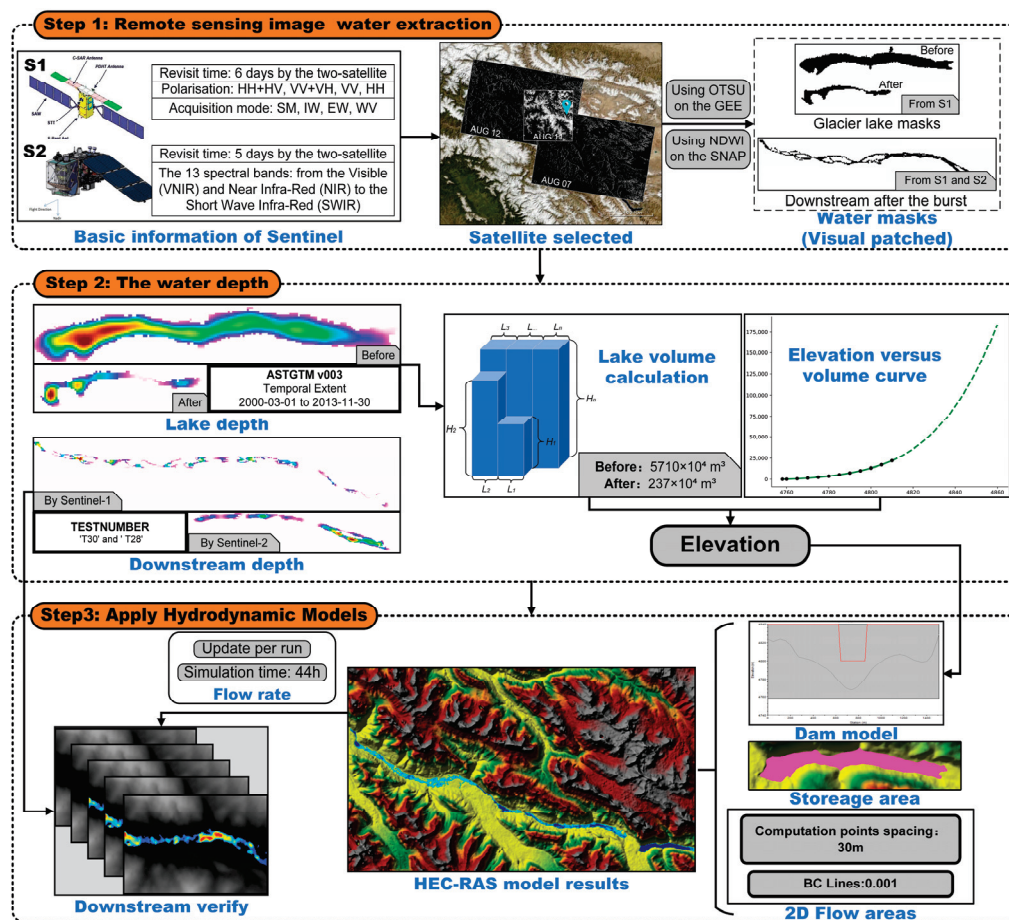


**Figure 1.** Kyagar Glacier lake. (a) The location of the lake in the HMA region. The black triangle represents the Kyagar Glacier. The background was made with DEM using the Shuttle Radar Topography Mission (SRTM) and HMA boundary [30]. (b,c) Geographic location of the Kyagar Glacier and lake. The image is a false-color composite based on Landsat 8 Level-2 surface reflectance data acquired on 12 July 2018, using bands 5, 4, and 3. The glacier boundary data were from the National Tibetan Plateau/Third Pole Environment Data Center. <https://cstr.cn/18406.11.glacier.001.2013.db> (accessed on 26 October 2024) [31,32].

## 2.2. Datasets

This study used a combination of tools, including Google Earth Engine (GEE, <https://code.earthengine.google.com/>) (accessed on 26 October 2024) and the Sentinel Application Platform (SNAP), to process the Sentinel-1 (S1) and Sentinel-2 (S2) satellite data. These data were used to extract the water masks and calculate lake volume and streamflow. The HEC-RAS model was then employed to simulate the event. The flowchart (Figure 2) illustrates the process of obtaining and analyzing remote sensing data, as well as the steps for hydrodynamic modeling.





**Figure 2.** Flowchart of data processing of remote sensing and hydrodynamic modeling process. The satellite parameters information was from Sentinel Online-SentiWiki (<https://sentiwiki.copernicus.eu/web/s1-mission>) (accessed on 25 September 2024). S1 and S2 images credits: TAS-I and EADS Astrium.

We focused on the Kyagar Glacier outburst event that occurred on 10 August 2018, drawing from the GLOF inventory for the High Mountain Asia (HMA) region. For data collection, we used satellite data collected before and after the event, including Sentinel-1 and Sentinel-2 data from 7 August to 12 August 2018. The S1 dataset used in the GEE is represented by COPERNICUS/S1\_GRD, which includes Ground Range Detected (GRD) scenes processed through the Sentinel-1 Toolbox to yield calibrated, ortho-rectified products. For this study, we selected the Interferometric Wide (IW) swath acquisition mode and dual polarization (HH&HV) configurations [33]. S2 data were sourced from a wide-swath, high-resolution, multispectral imaging mission. To streamline the workflow, we used Sentinel-2 Level-2A (L2A) products, which were ortho-rectified and provided Bottom-of-Atmosphere (BOA) reflectance [34]. Digital Elevation Model (DEM) data were acquired from the Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model V3 (ASTGTM v003), jointly developed by NASA and the Japan Aerospace Exploration Agency (JAXA), covering the region from  $35^{\circ}\text{N}$  to  $36^{\circ}\text{N}$  and  $76^{\circ}\text{E}$  to  $77^{\circ}\text{E}$ . In preliminary tests, alternative DEM datasets, including ALOS (12.5 m), SRTM30M, and NASADEM (30 m), were assessed. However, the selected DEM proved to be the most suitable, as other datasets failed to accurately represent water depth. These inaccuracies were largely due to extensive downstream data gaps and, in some instances, inconsistent glacier water depth patterns relative to the DEM data, occasionally even producing inverse depth representations.



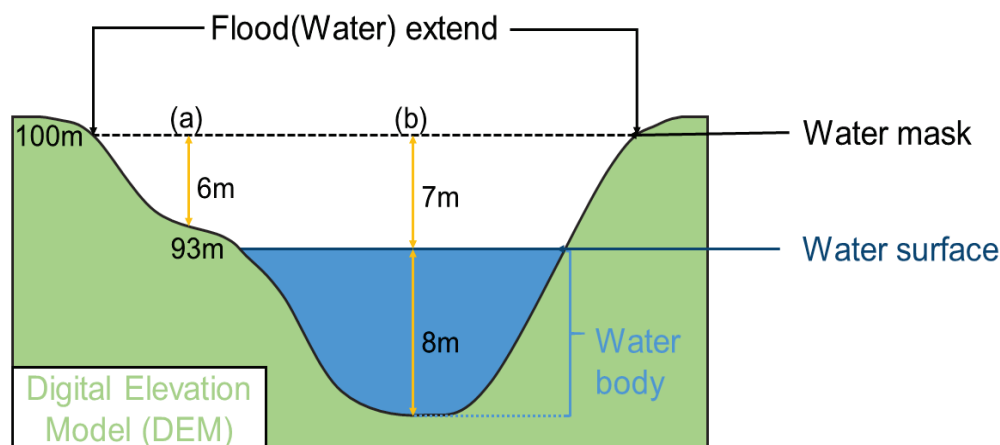
### 2.3. Approach of GLOF Detection in Sparse Observation Region

#### 2.3.1. Water Mask Derived from Remote Sensing Images

This study utilized data provided by the European Space Agency's (ESA) Sentinel-1 and Sentinel-2 satellites (Figure 2). Sentinel-1, equipped with a Synthetic Aperture Radar (SAR), was chosen for its capability to penetrate cloud cover and provide reliable observations in poor weather conditions, as well as its frequent revisit periods. Dual-polarization datasets from 7 August and 12 August were processed using Google Earth Engine (GEE), a platform developed in collaboration with Google, Carnegie Mellon University, and the United States Geological Survey (USGS), to extract water bodies from glacier lake images and downstream water features using OTSU-based hydrological extraction methods [35]. Sentinel-2 data were processed using SNAP V11.00 software. The Normalized Difference Water Index (NDWI), derived from surface reflectance data of bands 3 and 8, was used to extract water bodies [36]. This index effectively reduces interference from vegetation and soil, proving particularly successful in extracting large lake reservoirs [37]. Results with an NDWI value greater than 0 were selected, and visual corrections were applied using ArcGIS V10.2 to refine the extracted water body data (Figure 2).

#### 2.3.2. Water Depth Estimation

The FwDET calculates water depth by subtracting the topographic elevation of each grid cell within the flooded area from the flood surface elevation (above mean sea level, a.m.s.l.) derived from a DEM [38]. The core of the FwDET algorithm is the identification and application of the elevation of the boundary between water and land. The algorithm procedure [39] is as follows: (1) converting the water mask to a line layer, which is the boundary of water; (2) identifying from the boundary the grid cells that are the same size as the DEM; (3) extracting the elevation values from the DEM for these grid cells at same location (referred to as boundary grid cells elevation, BCE); (4) assigning the BCE to each grid cell within the flood domain from their nearest grid cells; and (5) calculating water depth by subtracting the elevation of DEM from BCE at each grid cell, as shown in Figure 3.



**Figure 3.** This is a cross-section diagram illustrating the principle of the FwDET calculation of flood depth. For example, if the imported water mask has an elevation of 100 m, the tool computes the water depth below each grid cell within that mask. In perennial rivers, the calculated depth tends to be underestimated, as measuring instruments for satellites usually capture the WSE rather than the riverbed, resulting in a reference plane for DEMs at the water surface. The point of (a) and (b) is separately land and waterbody in none flooding time. The actual water depth at point (a) and (b) is 6 m and 15 m, respectively.

FwDET v2.1 introduces a boundary grid-smoothing and filtering process, using a box  $5 \times 5$  grid for boundary elevation smoothing and multiple iteration loops to refine outputs [40–42]. In FwDET, both the extracted water body imagery and the selected DEM are crucial inputs, as well as specific settings for the Slope Filter (SF) and Iterations Option

(IO). Inputting appropriate SF and IO can repair the sharp transition area and streak area in the output result to a certain extent.

Custom parameter configurations (T01 to T31), those developed by the algorithm's owner and consisting of various SF and IO combinations, were tested iteratively (Table 1). "CustoNumber" (CtN) refers to customizable parameter sets, which were used to compare different presets. This study integrates the extracted water masks with 30-m resolution ASTER GDEM data and applies the FwDET v2.1 to obtain optimized depth results for both lakes and downstream rivers (see Step 2 in Figure 2). By multiplying each pixel's depth with its corresponding resolution and summing the resulting values across all pixels, the lake volumes before and after the outburst were calculated. The difference between the before and after outburst lake volumes, divided by the time elapsed between satellite image captures, yielded an estimated average outburst discharge.

**Table 1.** Different combinations of slope filter and iterations option.

| CtN | SF  | IO | CtN | SF  | IO | CtN | SF    | IO    |
|-----|-----|----|-----|-----|----|-----|-------|-------|
| T01 | 0.5 | 1  | T11 | 1.5 | 1  | T21 | 2.5   | 1     |
| T02 | 0.5 | 2  | T12 | 1.5 | 2  | T22 | 2.5   | 2     |
| T03 | 0.5 | 3  | T13 | 1.5 | 3  | T23 | 2.5   | 3     |
| T04 | 0.5 | 4  | T14 | 1.5 | 4  | T24 | 2.5   | 4     |
| T05 | 0.5 | 5  | T15 | 1.5 | 5  | T25 | 2.5   | 5     |
| T06 | 1.0 | 1  | T16 | 2.0 | 1  | T26 | FALSE | 1     |
| T07 | 1.0 | 2  | T17 | 2.0 | 2  | T27 | FALSE | 2     |
| T08 | 1.0 | 3  | T18 | 2.0 | 3  | T28 | FALSE | 3     |
| T09 | 1.0 | 4  | T19 | 2.0 | 4  | T29 | FALSE | 4     |
| T10 | 1.0 | 5  | T20 | 2.0 | 5  | T30 | FALSE | 5     |
|     |     |    |     |     |    | T31 | FALSE | FALSE |

### 2.3.3. GLOF Modeling with HEC-RAS

The Hydrologic Engineering Centers River Analysis System (HEC-RAS), developed by the US Army Corps of Engineers, is widely used for river hydrology and hydraulic modeling. It supports two-dimensional hydrodynamic analysis by solving either the full Saint-Venant equations or the simplified 2D diffusion wave equations. HEC-RAS has been extensively applied in the hazard assessment of GLOFs in steep terrains, such as the Himalayas, offering a promising approach for regional simulations of dam-break scenarios [22,43]. For this study, three key elements were modeled: Storage Area (SA), SA/2D Connections, and the 2D Flow Area (Step 3 in Figure 2).

First of all, the glacial lake upstream of the dam was modeled as a storage area, replacing the reservoir's function. We achieved quantitative control over the storage area by inputting 100 numerical points of the elevation-water volume curve. The initiation and termination of the outburst event could be regulated by adjusting the WSE at the onset and cessation of the breach taken from the remote sensing data. Due to the lack of field observations in the study area, empirical equations were used to estimate the flood hydrograph and breach development, which provided upstream boundary conditions for the storage area.

Secondly, the dam was modeled as SA/2D Connections by us, which represented the breach development during the outburst. Natural dam breaches often have complex, hard-to-measure shapes, being triangular, trapezoidal, or rectangular. Based on prior observations, this study used a trapezoidal shape [44]. Empirical equations were applied to estimate two key breach parameters: the average breach width ( $B_{ave}$ ) and breach development time ( $T_f$ ). As a result of considerations for minimizing the input of uncertain parameters, this study opted to employ the Von Thun and Gillette equation [45].

$$B_{ave} = 2.5h_w + C_b \quad (1)$$

$$\begin{cases} T_f = \frac{B_{ave}}{4h_w} (\text{Erosion resistant}) \\ T_f = \frac{B_{ave}}{4h_w + 61.0} (\text{Prone to erosion}) \end{cases} \quad (2)$$

where  $h_w$  is lake depth above the bottom of breach (m),  $C_b$  is a coefficient, which is a function of reservoir size (Table 2).

**Table 2.** Values of  $C_b$  are based on reservoir size.

| Reservoir Size (m <sup>3</sup> )        | $C_b$ (m) |
|-----------------------------------------|-----------|
| $<1.23 \times 10^6$                     | 6.1       |
| $1.23 \times 10^6$ – $6.17 \times 10^6$ | 18.3      |
| $6.17 \times 10^6$ – $1.23 \times 10^7$ | 42.7      |
| $>1.23 \times 10^7$                     | 54.9      |

The development time of the breach was calculated using Equation (2) using the estimated average breach width. It is worth noting that the breach development time equation differentiates between dams made of “erosion-resistant” and “erosion-prone” materials. In the study by Von Thun and Gillette, it was recommended to consider well-structured, erosion-resistant dams as the upper bound and poorly structured, easily erodible dams as the lower limitation of the calculations, respectively. The triggering mechanism for a dam breach was considered as overtopping. The computational results indicated that the width of the breach at the base was approximately 135 m, with side slopes of the trapezoidal breach at 0.5, and the breach formation time was around 1.05 h.

Finally, a 2D flow area was defined downstream of the dam, extending from the dam to 92,279.5 m and was subdivided into grids of equal size at  $30 \times 30$  m. Then, we computed 2D flow area hydraulic tables to preprocess the 2D flow area’s computational cells and faces into detailed tables based on the underlying terrain data to make sure the 2D hydrodynamic calculations were performed for each scenario. The total simulation time was 44 h, including a computation interval of 10 s and mapping output intervals set to 1 min. Four Boundary Condition Lines (BC Lines) were artificially set along the downstream segment to guide the flow in the correct east-west direction with a friction slope ( $S_f$  in Manning’s formula [46]; the slope of the energy grade line) of 0.001 applied to the downstream boundary conditions.

$$Q = VA = \left(\frac{1}{n}\right) AR^{\frac{2}{3}} \sqrt{S_f} \quad (3)$$

where  $Q$  is the flow rate (m<sup>3</sup>/s),  $V$  is velocity (m/s),  $A$  is the flow area (m<sup>2</sup>),  $n$  is Manning’s Roughness Coefficient, and  $R$  is hydraulic radius (m).

#### 2.4. Evaluation Metrics

In the field of numerical weather forecasting, prediction methods for precipitation primarily rely on satellite cloud imagery and radar echo extrapolation. The objective evaluation of forecast products is a crucial step in the operational workflow. Some widely used metrics have been utilized to evaluate the performance of radar-estimated rainfall and severe precipitation events [47–49], such as the Threat Score (TS), also known as the Critical Success Index (CSI), the False Alarm Ratio (FAR), and the Probability of Detection (POD) [50].

$$TS = \frac{TP}{TP + FP + FN} \times 100\% \quad (4)$$

$$POD = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

$$FAR = \frac{FP}{TP + FP} \times 100\% \quad (6)$$

where refers to the number of correctly predicted grid (or station) points, FP represents the number of false alarms, FN denotes missed forecasts, and TN indicates correctly predicted non-rainy points, as defined in the contingency table in Table 3. The waterbody is defined as the region with a water depth greater than 0.1 m in HEC-RAS simulations and detected as water from satellite images. TS measures the ratio of correct water cover in the simulation of HEC-RAS (successes) against observations of the water mask identified through remote sensing (including false alarms and missed events). POD reflects the proportion of true positives among the predicted positives, with higher POD values indicating a higher probability of correct simulation and thus greater reliability. FAR, on the other hand, measures the ratio of false positives to all simulated samples.

**Table 3.** Confusion matrix of rainfall classification.

| Is waterbody in observation of<br>satellite images | Is Waterbody in Simulated Output of HEC-RAS |                                           |                                           |
|----------------------------------------------------|---------------------------------------------|-------------------------------------------|-------------------------------------------|
|                                                    | True                                        |                                           | False                                     |
|                                                    | True<br>False                               | TP (True Positive)<br>FN (False Negative) | FP (False Positive)<br>TN (True Negative) |

### 3. Results

#### 3.1. Depth and Flow Estimation from Multiple Remote Sensing

##### 3.1.1. Changes in Upstream Glacier Lake

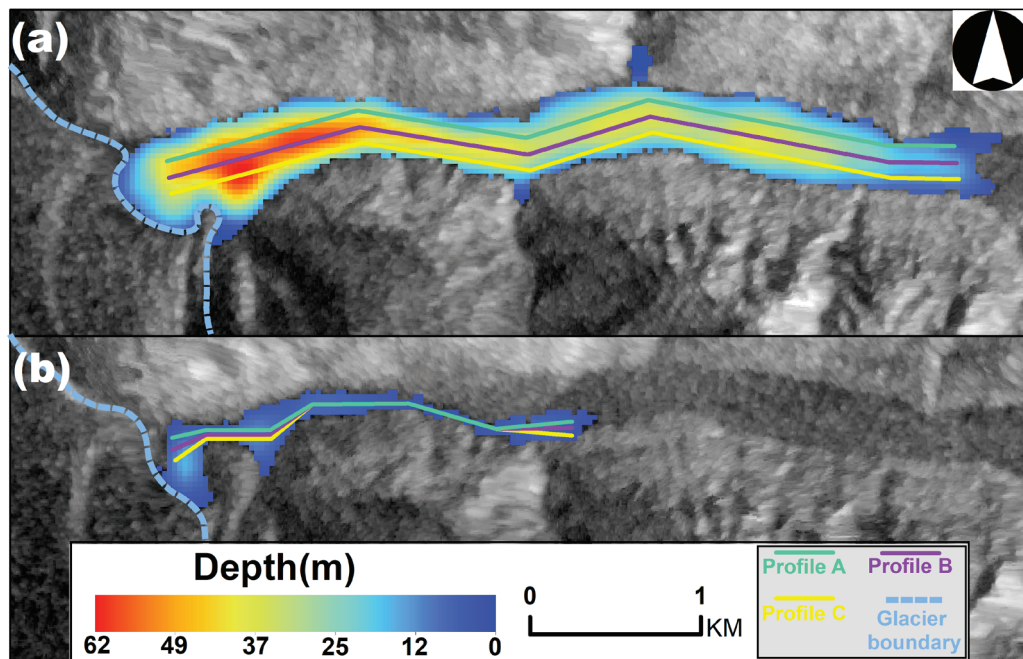
Though previous studies have recommended 10 iterations and a 0.5% slope threshold as generally effective for most DEMs, these settings were suboptimal for our calculations [40].

We assumed that the WSE of the lake was consistent throughout and drew three polylines named Profile A–Profile C covering the east and west sides of the lake (Figure 4). The values of water depth and terrain along these polylines were then extracted. Under the assumption of a uniform WSE, the rates of change in water depth along the polylines were expected to be inversely proportional to the rates of change in terrain. We calculated the percentage of instances where these changes were inversely related, with a higher percentage indicating a more accurate representation of the lake’s calculations; otherwise, it was the opposite. In this study, we analyzed the two sets of inversion results before and after the lake outburst and calculated the inversely related percentage along six selected polylines (Figure 5). In the before scenario, it can be observed that the calculations for T28–T30 show better agreement, with the calculated lake area closely matching the results of water body extraction. In the after scenario, Profile C stands out as the primary polyline with a higher proportion of opposite slopes and less fluctuation, showing relative consistency and a higher success rate for T24–T30 when compared to the other two profiles. Based on these findings, the results from T30 and T28 were selected to represent the lake depths before and after the glacier lake outburst in the subsequent sections.

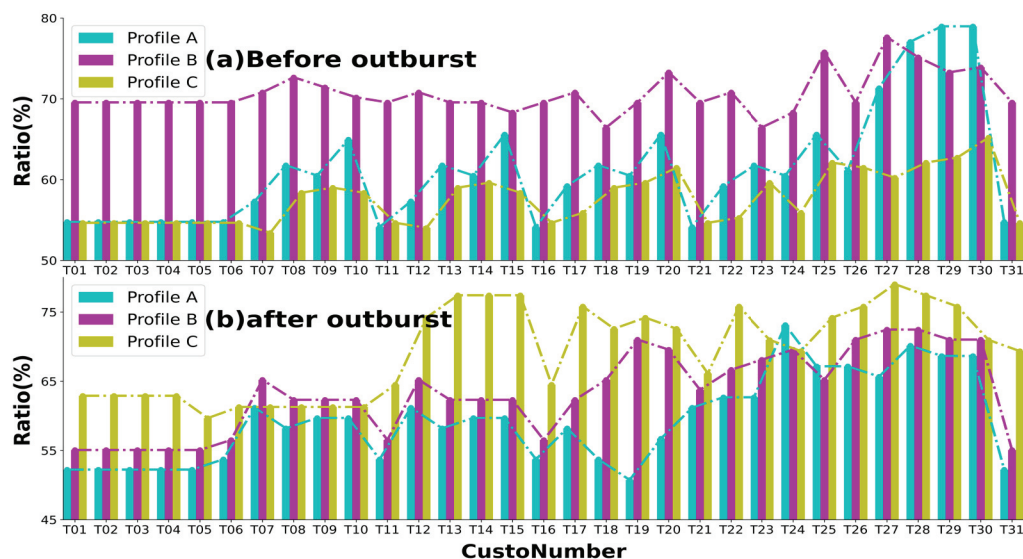
Before the outburst, the glacier lake expanded due to glacial meltwater and rainfall, reaching a surface area of 2.25 km<sup>2</sup> and a capacity of 57.10 MCM, slightly below the 2017 record. The before outburst imagery (Figure 4a) shows deeper areas close to the dam, with depth decreasing from west to east. This configuration channels water from the Shaksgam Valley to the lake, priming the lake for a rapid and violent outburst. At this stage, the maximum water depth was 61.78 m, with an average depth of 27.01 m. In the early morning of August 10, a large amount of lake water gushed out through the dam, but due to the limited water level and the flow channel, the lake could not completely pour out. Following the outburst (Figure 4b), only minimal residual water remained, with a maximum depth of just 15.08 m and an average depth of 5.90 m.

With this depth data, the total lake volume was computed by summing the products of the pixel depth and the pixel area, obtaining an estimated mean discharge rate of 126.6 m<sup>3</sup>/s over 432,476 s.





**Figure 4.** FwDET-generated water depth maps for the Kyagar Glacier lake. Background: Sentinel-1 dual-polarization images acquired on 7 August (before outburst) and 12 August (after outburst), 2018. (a) Before the outburst on 7 August 2018 by Sentinel-1. (b) After the outburst on 12 August 2018 by Sentinel-1.

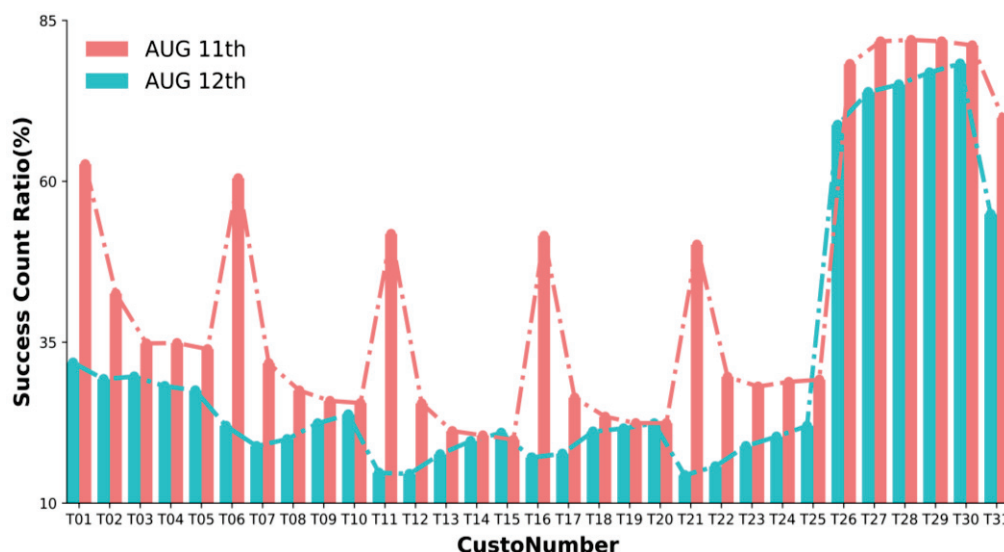


**Figure 5.** Slope filter and iteration setting combinations in FwDET and their respective success rates for computation results.

### 3.1.2. Downstream Flooding Process

Downstream of the glacier is different from the lake, resulting in difficulty and inauthenticity for the FwDET inversion of the downstream. In this section, we calculated the percentage of the water flow area obtained from inversion results under varying parameter conditions relative to the water mask abstracts (visual patches) from the remote sensing data (Step 1 in Figure 2). According to the success count ratio (Figure 6), for T1–T25, the success rate of 12 August remained below 35%, while 11 August showed a maximum success rate of approximately 60% and a minimum of 23%. However, when the parameter combinations reached T26–T30, the success rate rose sharply, with the average success rate for both the 11 August and 12 August exceeding 70%. This indicates that

inversion results are more reliable with larger parameter combinations. Accordingly, this study utilized the depth results calculated under the T30 combination as reference data for subsequent analyses.



**Figure 6.** Area percentage of FwDET-derived results under varied parameter combinations, based on Sentinel-2 (11 August) and Sentinel-1 (12 August) data for downstream rivers that were abstracted.

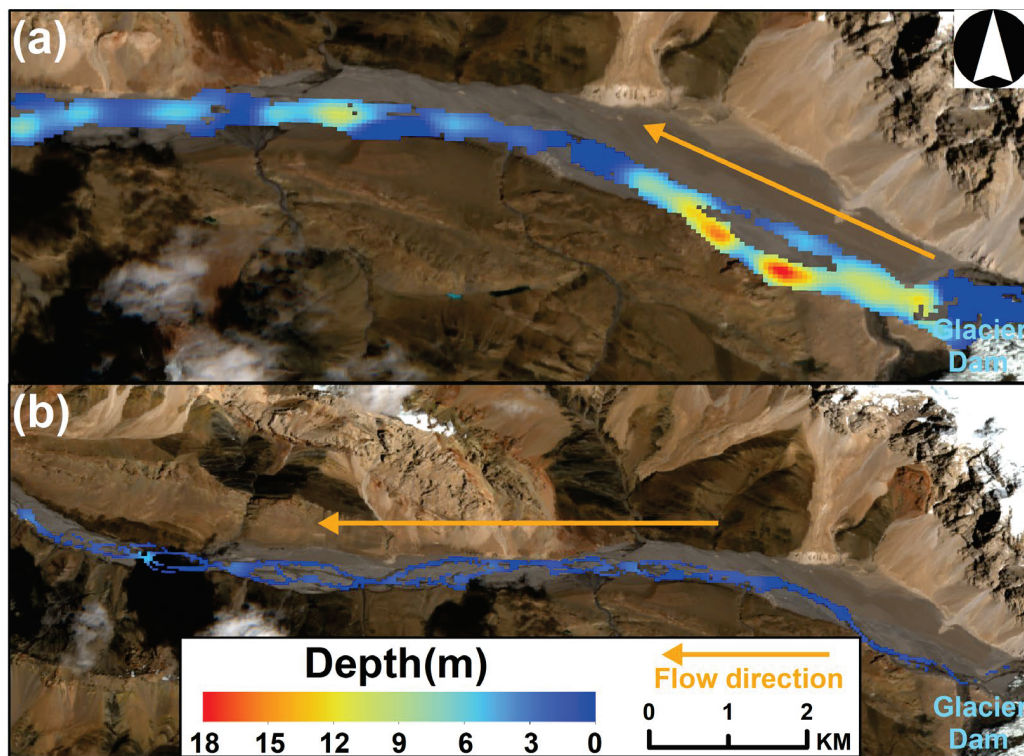
Based on the depth maps of the downstream after the outburst (Figure 7) and remote sensing data, several key observations can be made. First, the downstream region exhibits an overall west-to-east decrease in elevation, a feature that prevents backflow from the lake during discharge events. Second, there is a substantial elevation drop between the dam crest and downstream areas, with a height difference exceeding 50 m. This significant drop in height facilitates the rapid discharge of lake water during dam breach events, leading to intensified flood flows. Lastly, the downstream river network is complex, with segments that become disconnected during periods outside the peak flood. These characteristics are evident for 12 August (Figure 7b), where the peak discharge phase has already passed, resulting in markedly reduced flow, consequently creating a sparse and shallow river network near the glacier tongue.

### 3.1.3. Elevation-Water Volume Curve

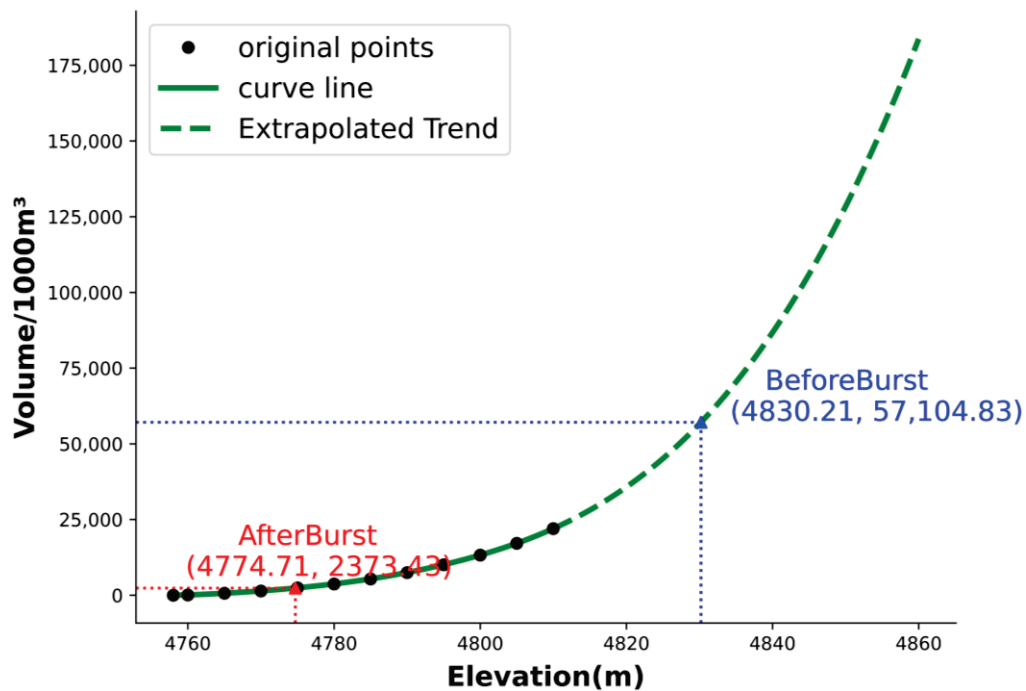
Building on the lake depth data obtained from the water masks, we calculated the lake volume to estimate the average discharge rate. Beyond direct discharge calculation, the lake volume also enables us to derive corresponding WSEs, achieved by constructing a storage capacity curve. This curve represents the relationship between the WSE ( $x$ -axis) and the lake volume ( $y$ -axis). In previous research, the lake area and volume at different elevations were estimated by using contour analysis and a parabolic fitting method based on GIS lake outlines and a recent DEM [51]. In this study, we extended these findings by interpolating 12 elevation-volume data points using Python V3.11. To ensure the extrapolated curve encompassed lake volumes before and after the outburst, we generated 1000 points uniformly between 4758 m and 4860 m and applied cubic spline interpolation to create a continuous, smooth curve. This approach maintained consistent trends in the curve for  $x$ -values above and below 4810 m, ensuring continuity in volume estimations.

The elevation-water volume (Figure 8) indicates that the volume of the lake before the outburst was approximately 57.1 MCM at a WSE of 4830.2 m, and the volume of the lake after the outburst was approximately 2.4 MCM at a WSE of 4774.7 m, closely aligning with the mean values derived from combined depth and DEM data, for which the WSE before the burst was 4830.1 m, while after it was 4784.8 m. The error between the inversion

result and the storage capacity curve after the collapse was 0.2%. Thus, we obtained the elevation-water volume curve and WSEs for both before and after the outburst.



**Figure 7.** FwDET-generated water depth maps for the downstream after the Kyagar Glacier lake outburst. Background: True-color composite image of Sentinel-2 satellite data based on B4, B3, and B2 bands. (a) Represents 11 August 2018 by Sentinel-2. (b) Represents 12 August 2018 by Sentinel-1.



**Figure 8.** Elevation-water volume curve for the lake derived via interpolation.

In fact, due to the distinctive structure of the Kyagar glacier dam, which is a mix of permafrost and ice layers rather than a tightly sealed dam, the lake undergoes continuous



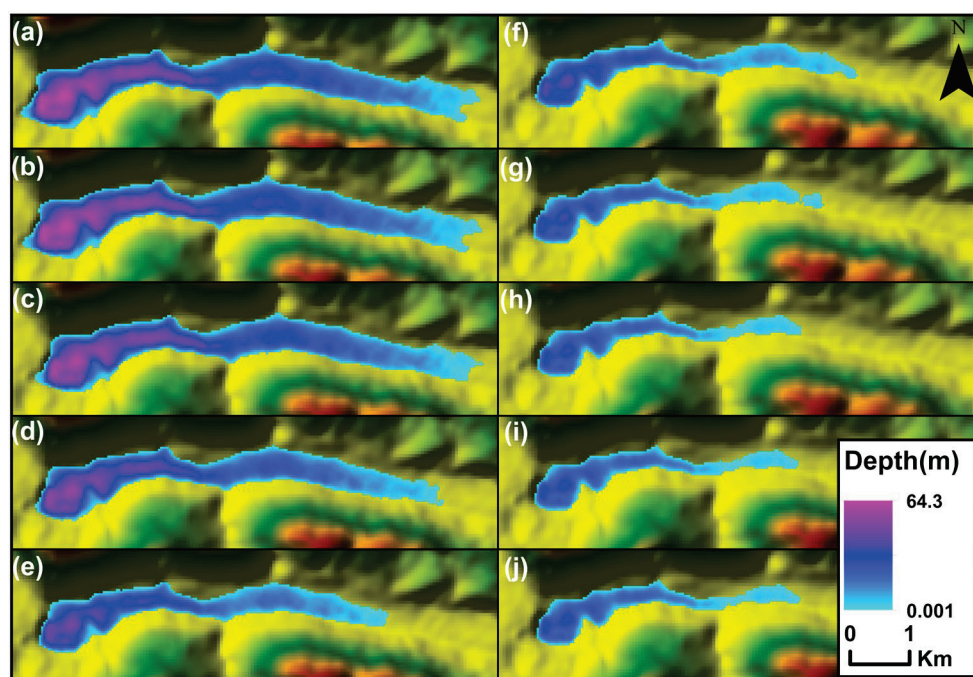
drainage even outside breach events, especially when nearing the dam's maximum holding capacity. This causes a gradual movement in the lake water from east to west, flowing from high to low areas. Consequently, the WSE across the lake is not uniformly distributed but varies spatially. The WSE is an average value.

### 3.2. GLOF Process Modeling by HEC-RAS

#### 3.2.1. Hydrograph Modeling

To accurately estimate downstream flooding, precise estimates of the breach location, size, and development time are essential, though these parameters often introduce substantial uncertainty in the analysis. Due to limited field observations, this study utilized regression equations of HEC-RAS for breach modeling, built by Von Thun and Gillette [45]. To simulate the storage area in the model, we randomly selected 100 points within the 4748–4860 m range from the elevation-water volume curve. Given the complex structure of the glacier dam and Kyagar Glacier's surge mechanism, we set the final bottom elevation of the breach to 4800 m, which was 0.3% higher than the WSE from the combination of depth and DEM. This higher-than-average WSE made sure the water was directed downstream while facilitating breach progression. This adjustment ensured the simulated outcomes were similar to the after-outburst imagery attained through remote sensing (Figure 4b).

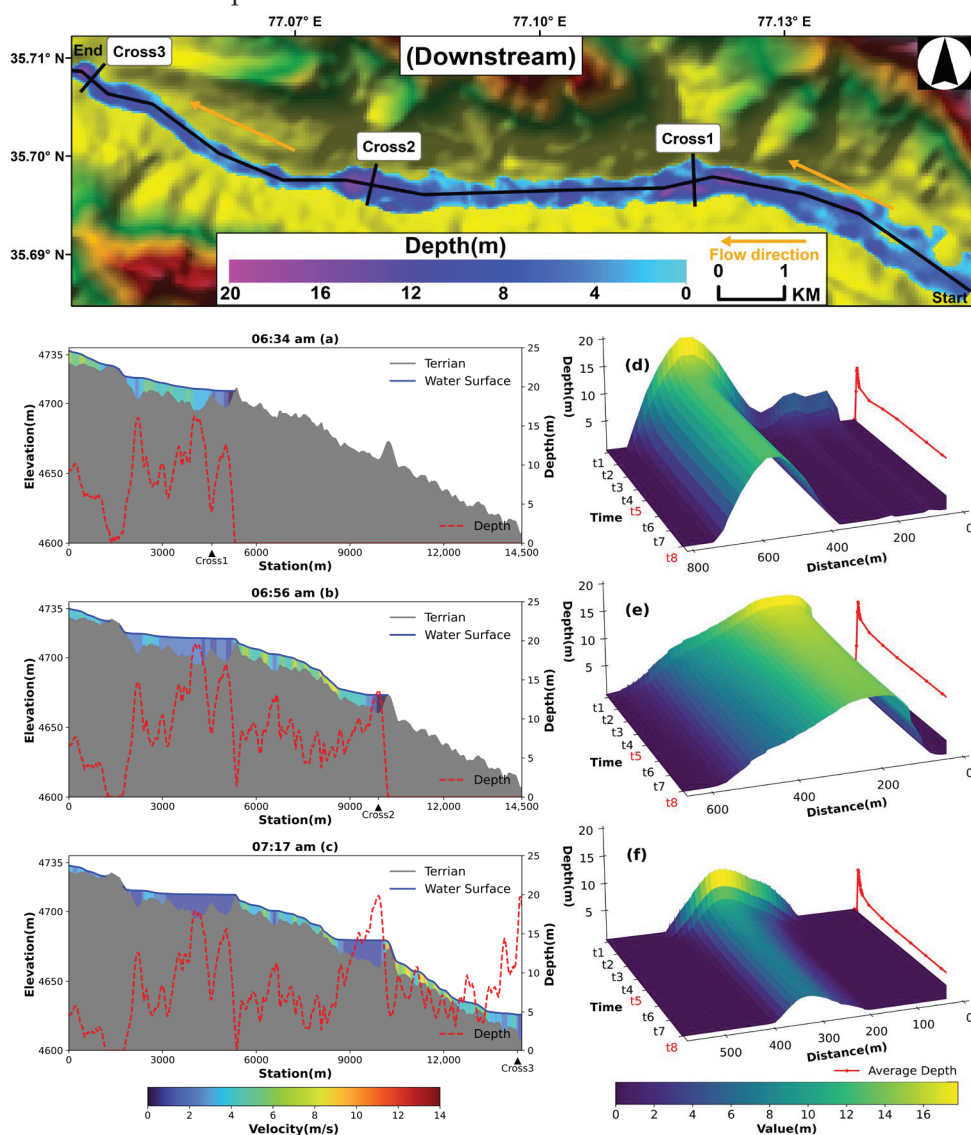
Figure 9 illustrates the model-simulated lake coverage area and water depth. A few outlier regions with unreasonable extreme water presence were excluded from the analysis. In Figure 9a, the before-outburst lake area, which was simulated by the HEC-RAS model based on the elevation-water volume curve, closely matches the before-breach lake area (Figure 4a) derived from remote sensing data. Moreover, from the perspective of depth observation, both demonstrate a similar pattern in terms of greater depth in the west and shallower waters in the east. Following the burst, the lake area rapidly decreased, with a corresponding sharp decline in WSE, continuing until approximately 0900 am on 10 August 2018. During the simulation, it was evident that, by the third hour, lake dynamics began to stabilize, with only minor fluctuations in lake area and water volume. As observed in Figure 9j, during the model's final equilibrium phase, the simulated lake area was slightly more than the lake area gained from remote sensing data (Figure 4b).



**Figure 9.** Images of glacier lake evolution over time as simulated by HEC-RAS. (a–e) The lake image times on 10 August 2018 at 6:00 a.m., 6:10 a.m., 6:20 a.m., 6:30 a.m., and 6:45 a.m. (f–j) The lake image times on 10 August 2018 at 7:00 a.m., 7:30 a.m., 8:00 a.m., 8:30 a.m., and 9:00 a.m.



The outburst process of the Kyagar Glacier lake involved the rapid discharge of a substantial water volume. Profiles revealed a stepped structure in the downstream flow path: 0–5314 m as the first step, 5314–10,203 m as the second step, and 10,203–14,852 m as the third step. During the outburst, the rapid flow along each step initiated abrupt increases in velocity, with water accumulating until it reached the elevation of the next step. In the first step of water storage, the velocity was relatively high when the flow initially entered, causing the rapid rise in WSE (Figure 10a–c). However, during the subsequent process of water delivery, the velocity gradually decelerated until it reached a stable WSE. The second and third steps had higher drops and more turbulent water flows compared to the first step, and there existed a rough terrain section of riverbed. The main water storage areas of the second and third steps were still located at the end of each step, while the areas with the highest velocities were the rough riverbed areas. For each step, when the water storage exceeded the bearing capacity of the step, due to the drainage effect, the water storage height of the step gradually decreased until it was eventually level with the terrain that divided the steps.



**Figure 10.** Inundation maps for downstream regions of the glacier dam over time, as simulated by HEC-RAS. The inundation diagram from HEC-RAS, which is a union of the maximum inundation depths in each part during modeling. Start-end indicates the longitudinal line of the downstream river taken along the flow direction. (a–c) The results of the outburst model transform with time and

distance along the start-end, while cross1-3 represent the locations where three-dimensional views of water depth change over time at different cross-sections. (d–f) The variations of water depth along the cross1-3 with time at different points. Each time interval (t1–t8) represents a snapshot within the overall simulation, with t5 (11 August 5:36 a.m.) and t8 (12 August 0:58 a.m.) matching satellite data acquisition times. The average depth represents the average rate of water depth on the crosses.

To investigate depth variations across main storage areas in each step, we constructed three cross-sectional profiles nearly perpendicular to the start-end line. We created a 3D visual map of water depth changes over time from the cross-sections. These panels (Figure 10) revealed that the river is widest in the first step, with an average width of 670 m (Figure 10d), and average widths of 610 m (Figure 10e) and 200 m (Figure 10f) in the second and third steps, respectively. At the end of the first step and the beginning of the second, uneven terrain results in a complex water network with multiple maxima. This effect, caused by extended low-lying areas downstream of the glacier tongue, enhances water storage and flow during non-outburst events and serves as a stable conduit. The confinement of water to these lowlands under simulated outburst conditions leads to prolonged retention in specific grid cells due to limited model resolution, creating regions of slower flow and even potentially permanent standing water.

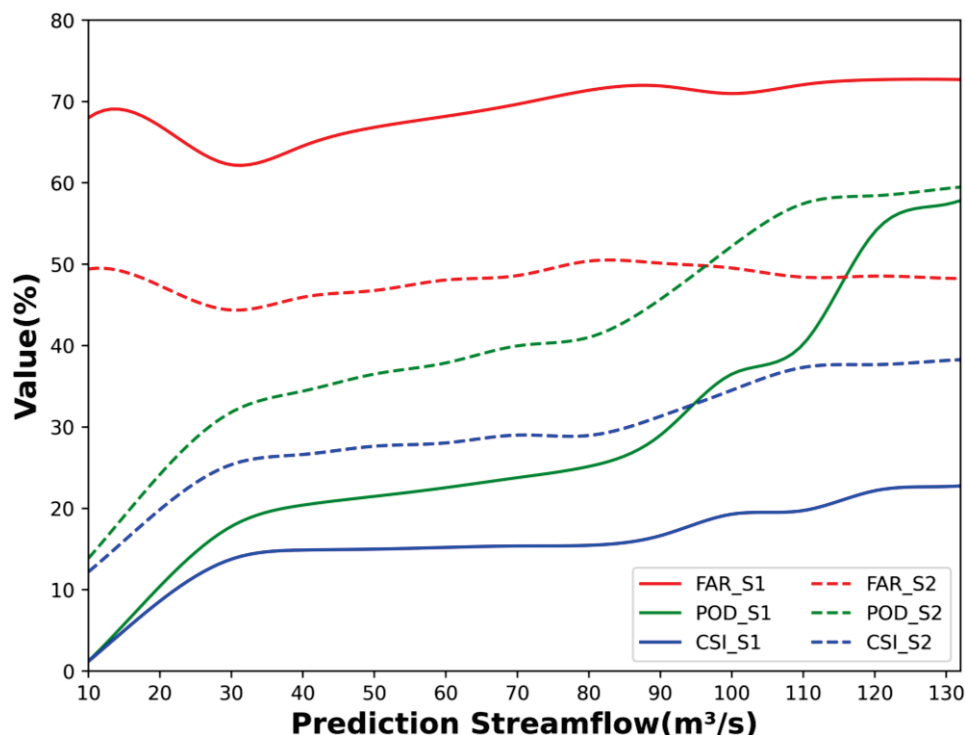
### 3.2.2. Optimal Parameters Assessment

To evaluate the accuracy of the simulated inundation depths compared to the actual flood data, a binary threshold of 0.1 m was used to distinguish flooded from non-flooded areas. Meteorological scoring functions were introduced: POD represents the proportion of correctly predicted flooded areas relative to the total actual flooded areas, CSI reflects the ratio of correctly predicted flooded areas to the total number of predicted and observed flooded areas, and FAR quantifies the percentage of false alarms, or areas incorrectly predicted as flooded [47,50].

### 3.2.3. Evaluation Against Remote Sensing

In Figure 11, the evaluation metrics reveal that POD and CSI generally increase as flow rates approach  $130 \text{ m}^3/\text{s}$ , indicating the model's reliability within moderate flow scenarios. Beyond  $130 \text{ m}^3/\text{s}$ , the changes in metric values stabilize. Initial increases in POD and CSI were due to the low discharge flow, which limited the spread and extent of inundation within the simulation timeframe, resulting in higher omission rates. With the rising of the streamflow, the modeling outcomes improved in matching observed width and extent, enhancing hit rates. At higher streamflow, modeling outcome performance stabilized for two reasons: (1) The inundation simulation area corresponds with available remote sensing data, which are limited to portions of the river, constraining excessive false alarms. Increasing discharge causes water to flow further downstream, restricting its interaction with comparable satellite observations. (2) Downstream water-holding capacity limits further accumulation within the model's domain (Figure 10), requiring water to spread laterally. Comparing the model results for 11 and 12 August at lower flow rates suggests a better model fit to 11 August data, which represent the immediate after-outburst period. The rapid flow in the early stages of the outburst explains the wider and deeper river observed in the 11 August satellite image, correlating closely with simulated conditions around 5:36 am on 11 August. Consequently, the model achieves higher POD and CSI, with lower FAR on this date.

Overall, we conclude that the Kyagar Glacier lake outburst model is best suited for moderate flow rates, with an upper limit near  $132 \text{ m}^3/\text{s}$ . The estimated average discharge rate of  $126.6 \text{ m}^3/\text{s}$  calculated from before and after outburst lake capacities is thus considered reliable. This model is less suitable for extreme flood events, as indicated by historical data showing that, out of 20 outbursts from 1959 to 2020, 8 exceeded  $130 \text{ m}^3/\text{s}$ , with peak flows reaching up to  $6000 \text{ m}^3/\text{s}$  [23,52].



**Figure 11.** Model evaluation based on POD, CSI, and FAR metrics across varying average flow rates. S1 represents downstream depth inversions using FwDET based on Sentinel-1 imagery (12 August), while S2 reflects Sentinel-2 data (11 August). We excluded the results where the flow rate was greater than 132 m<sup>3</sup>/s. This is because an overly large average flow rate would exceed the total water capacity of this glacier lake.

#### 4. Discussions

The burst of this particular glacier dam occurred in the early morning of 10 August 2018. We used the satellite data obtained on 7 August 2018 as a reference for the lake image before the outburst. Due to the lack of data, we could not determine how different the lake image (Figure 4a) extracted on August 7 was from the real lake. Therefore, the conclusion of this article is based to a certain extent on the hypothesis that the lake surface did not change until the burst event. The water extent extracted from S1 data has limitations. However, with the gradual development of machine learning, the use of optical sensing satellites with higher revisit rates for water mask extraction has made it possible to obtain better lake images [53]. This would be an approach worth considering for other GLOFs.

This study utilized satellite data from various sources to monitor and characterize glacial lake structures, a foundational step for effective GLOF analysis. However, the rapid outburst of the Kyagar Glacier lake and the swift downstream flow presented challenges in capturing detailed observations through satellite imagery. The Kyagar Glacier lake exhibits a distinct cyclical filling-and-bursting pattern, with similar outburst events. This study focused on a single case of such an event; however, due to the cyclic nature of these disasters, multiple cases are often necessary to substantiate experimental results, highlighting a limitation of the study's generalizability.

Extracting accurate hydrological images from satellite data remains challenging, as the automated recognition of water bodies can underestimate their actual extent. The Kyagar Glacier lake formed due to the Kyagar Glacier's ice tongue, which facilitates water accumulation. Complex glacier surging behaviors and physical mechanisms can result in ice mixed with sediment near the water surface. This mixture reduces SAR signal reflectance, potentially causing the misclassification of water areas. For such cases, incorporating data from other SAR or optical remote sensing satellites can improve image correction. This study primarily employed S1 and S2 data, with limited use of the Landsat

and MODIS satellite series. Future studies could explore different regions and integrate higher-frequency observation satellites to improve data extraction accuracy.

Studies have indicated that the formation and outburst of the Kyagar Glacier lake are closely related to its unique geographical setting and local climate variations. Therefore, changes in water depth are not solely due to upstream outbursts. Outburst events typically occur in the summer, a period marked by high temperatures and mountain snowmelt. This season brings additional inflows from both sides of the valley, along with surrounding rockfalls, which can increase downstream water depth. We did not account for WSE changes due to non-upstream factors or depth reduction caused by deviations in the shallow water wave equations in the model. Furthermore, infiltration processes along downstream channels complicate accurate water depth predictions, adding aspects that are beyond the scope of hydrodynamic modeling. Long-term upstream erosion, coupled with floating ice and sediment transport, has resulted in relatively flat terrain with intricate flow pathways between the two glaciers. The formation of physical characteristics downstream still warrants further investigation.

The accuracy of water depth estimates derived using FwDET, lake volume calculations, and HEC-RAS dam breach simulations depends significantly on the choice of DEM. High-resolution DEMs enhance terrain detail, supporting more precise depth estimation. DEMs with greater vertical resolution can better capture local elevation changes, thus improving depth accuracy. DEMs are also crucial input data for HEC-RAS models, directly influencing key simulation outcomes, including water flow paths, channel morphology, flow distribution, and flood inundation extent. Counterintuitively, when using the ALOS DEM with a 12.5 m resolution in preliminary tests, the derived water depth estimations were inferior, achieving less than 82.1% coverage compared to the results presented in this study, which was more than 98.9% (Figure 7). This discrepancy underscores the importance of selecting an appropriate DEM dataset, which warrants further investigation.

In our model development process, we observed a distinctive characteristic related to flood events. Specifically, during weaker flood events associated with lower flow rates, the downstream water extent is generally shorter, and thus the impact on downstream communities is minimal. In contrast, during stronger flood events with higher flow rates (defined here as events with an average flow rate exceeding  $132 \text{ m}^3/\text{s}$ ) our model indicates a ‘stabilization’ phenomenon in downstream water flow at around  $130 \text{ m}^3/\text{s}$ . At this threshold, the water extent reaches a point where it no longer expands significantly, suggesting that the model may be approaching an upper limit where the storage capacity within each grid cell reaches equilibrium. What causes this phenomenon, and how can the model be optimized to address it? These questions represent crucial areas for further exploration, particularly as this approach is applied to a broader range of case studies in the future.

## 5. Conclusions

Using remote sensing data from the Sentinel-1 and Sentinel-2 satellites, this study successfully delineated the water extent of the Kyagar Glacier lake before and after the outburst. By applying the FwDET, we derived water depth and volume changes, estimating an average outburst discharge of approximately  $126.6 \text{ m}^3/\text{s}$ . This estimated discharge provides a valuable reference for future hydrological analysis and model development related to glacial lake outburst floods (GLOFs). Additionally, understanding lake volume changes further enhances our knowledge of glacial lake dynamics. These results were integrated into a comprehensive GLOF model using HEC-RAS, combining FwDET-derived depth data with hydrodynamic simulations to assess downstream impacts of the outburst at different stages. From the initial breach until 05:36 am on 12 August 2018, the model estimated an average discharge between  $120\text{--}140 \text{ m}^3/\text{s}$ . Given that the latter range includes the estimated  $126.6 \text{ m}^3/\text{s}$ . The findings indicate that the Kyagar Glacier lake outburst was both rapid and destructive. The integration of remote sensing data and hydrodynamic modeling provided a thorough evaluation of the inundation extent, depth, and flow velocity downstream of the ice dam, offering insights valuable for future disaster management



and early warning systems. This study demonstrates an effective multi-tool approach to GLOF hazard assessment, underscoring the extensive potential of remote sensing and hydrodynamic modeling for disaster simulation.

**Author Contributions:** Conceptualization, L.J. and Z.L.; methodology, L.J.; software, L.J.; validation, L.J. and Z.L.; formal analysis, Z.L. and Z.Z.; investigation, L.J. and H.L.; resources, J.Z.; data curation, D.S.; writing—original draft preparation, L.J. and M.S.; writing—review and editing, L.J. and Z.L.; visualization, L.J. and Z.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Second Tibetan Plateau Scientific Expedition and Research (STEP) Program (Grant No. 2019QZKK0103), the National Natural Science Foundation of China (Grant No. 42165006 and No. 42030611), and the Key R&D Plans of Tibet Autonomous Region in 2022 (Grandt No. XZ202201ZY0008G).

**Data Availability Statement:** The Sentinel satellite data are available from <https://dataspace.copernicus.eu/> (accessed on 27 October 2024). The ASTGTM v003 are available from <https://search.earthdata.nasa.gov/> (accessed on 27 October 2024).

**Acknowledgments:** The authors thank the European Space Agency (ESA) satellite missions for providing satellite data support for this study, as well as NASA and the Japan Aerospace Exploration Agency (JAXA) for jointly developing and making publicly available digital elevation models. The glacier data set is provided by National Tibetan Plateau/Third Pole Environment Data Center (<http://data.tpdc.ac.cn> (accessed on 26 October 2024)). We would also thank many anonymous contributors for their technical and financial support. The authors thank Planet Labs for Planet and RapidEye datasets, Japan Aerospace Exploration Agency (JAXA) and European Space Agency (ESA) for ALOS-1 PALSAR-1 and ALOS-2 PALSAR-2 data, and Qinghai-Tibet Plateau ground monitoring station for metrological data. We also thank the editor and anonymous reviewers for their helpful comments and suggestions.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Ding, Y.K.; Feng, Y.Q.; Chen, K.; Zhang, X.C. Analysis of spatial and temporal changes in vegetation cover and its drivers in the aksu river basin, china. *Sci. Rep.* **2024**, *14*, 15. [CrossRef] [PubMed]
2. Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2022—Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2023.
3. Hugonnet, R.; McNabb, R.; Berthier, E.; Menounos, B.; Nuth, C.; Girod, L.; Farinotti, D.; Huss, M.; Dussaillant, I.; Brun, F.; et al. Accelerated global glacier mass loss in the early twenty-first century. *Nature* **2021**, *592*, 726. [CrossRef] [PubMed]
4. Zemp, M.; Huss, M.; Thibert, E.; Eckert, N.; McNabb, R.; Huber, J.; Barandun, M.; Machguth, H.; Nussbaumer, S.U.; Gärtner-Roer, I.; et al. Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* **2019**, *568*, 382–386. [CrossRef] [PubMed]
5. Yin, Y.; Wang, X.; Liu, S. Characteristics and influence factors of the glacial lake changes in china from 1990 to 2020. *J. Lake Sci.* **2023**, *35*, 358–367.
6. Tang, M.; Chen, H.; Zhao, H.; Yang, W.; Deng, W. Identification, development law and risk assessment of the hidden dangers of glacial lake outburst disasters on the qinghai-tibet plateau. *Geol. Bull. China* **2023**, *42*, 730–742.
7. Liu, X.; Chen, B. Climatic warming in the tibetan plateau during recent decades. *Int. J. Climatol. J. R. Meteorol. Soc.* **2000**, *20*, 1729–1742. [CrossRef]
8. Rangwala, I.; Miller, J.R.; Xu, M. Warming in the tibetan plateau: Possible influences of the changes in surface water vapor. *Geophys. Res. Lett.* **2009**, *36*, L06703. [CrossRef]
9. Tang, S.; Vlug, A.; Piao, S.; Li, F.; Wang, T.; Krinner, G.; Li, L.Z.; Wang, X.; Wu, G.; Li, Y. Regional and tele-connected impacts of the tibetan plateau surface darkening. *Nat. Commun.* **2023**, *14*, 32. [CrossRef]
10. Zhang, G.; Carrivick, J.L.; Emmer, A.; Shugar, D.H.; Veh, G.; Wang, X.; Labeledz, C.; Mergili, M.; Mölg, N.; Huss, M.; et al. Characteristics and changes of glacial lakes and outburst floods. *Nat. Rev. Earth Environ.* **2024**, *5*, 447–462. [CrossRef]
11. Peng, M.; Wang, X.; Zhang, G.; Veh, G.; Sattar, A.; Chen, W.; Allen, S. Cascading hazards from two recent glacial lake outburst floods in the nyainqentanglha range, tibetan plateau. *J. Hydrol.* **2023**, *626*, 130155. [CrossRef]
12. Chen, L.; Fan, X.; Xiong, J.; Wang, X.; Dou, X. Hazard assessment of glacial lake outbursts in the doyingongba basin, southeastern tibetan plateau. *Bull. Geol. Sci. Technol.* **2023**, *42*, 258–266.

13. Ren, S.; Jiang, L.; Zhai, S.; Ye, X.; Guo, J. Study on erosion evolution characteristics of glacial lake break flood based on dynamic process. *J. Glaciol. Geocryol.* **2023**, *45*, 1300–1313.
14. Yu, Y.; Li, B.; Li, Y.; Jiang, W. Retrospective analysis of glacial lake outburst flood (glof) using ai earth insar and optical images: A case study of south lhonak lake, sikkim. *Remote Sens.* **2024**, *16*, 2307. [CrossRef]
15. Taylor, C.; Robinson, T.R.; Dunning, S.; Rachel Carr, J.; Westoby, M. Glacial lake outburst floods threaten millions globally. *Nat. Commun.* **2023**, *14*, 487. [CrossRef] [PubMed]
16. Emmer, A.; Allen, S.K.; Carey, M.; Frey, H.; Huggel, C.; Korup, O.; Mergili, M.; Sattar, A.; Veh, G.; Chen, T.Y.; et al. Progress and challenges in glacial lake outburst flood research (2017–2021): A research community perspective. *Nat. Hazards Earth Syst. Sci. Discuss.* **2022**, *22*, 3041–3061. [CrossRef]
17. Luo, H.; Lin, Z.; Chen, H.; Xiao, D.; Chen, G.; Su, D. Assessments of multiple precipitation products and application in hydrodynamic simulations: A case of casualty-inducing mountain torrents in Sichuan, southwest China. *J. Flood Risk Manag.* **2024**, *17*, e13016. [CrossRef]
18. Su, L.-f.; Li, Z.; Gao, F.; Yu, M. A review of remote sensing image water extraction. *Remote Sens. Land Resour.* **2021**, *33*, 9–19.
19. Gu, C.; Li, S.; Liu, M.; Hu, K.; Wang, P. Monitoring glacier lake outburst flood (glof) of lake Merzbacher using dense Chinese high-resolution satellite images. *Remote Sens.* **2023**, *15*, 1941. [CrossRef]
20. Gao, Y.; Yang, W.; Guo, R.; Jiang, L. Remote sensing monitoring and analysis of Jinwuco lateral moraine landslide-glacial lake outburst in southeast Tibet. *Remote Sens.* **2023**, *15*, 1475. [CrossRef]
21. Yang, L.; Lu, Z.; Ouyang, C.; Zhao, C.; Hu, X.; Zhang, Q. Glacial lake outburst flood monitoring and modeling through integrating multiple remote sensing methods and hec-ras. *Remote Sens.* **2023**, *15*, 5327. [CrossRef]
22. Sattar, A.; Goswami, A.; Kulkarni, A.V. Application of 1d and 2d hydrodynamic modeling to study glacial lake outburst flood (glof) and its impact on a hydropower station in central Himalaya. *Nat. Hazards* **2019**, *97*, 535–553. [CrossRef]
23. Tunyaz, N.; Mijit, M.; Weiye, M.; Mijit, M.; Zhang, A.; Liu, Y. Variation characteristics of kyagar glacial lake outburst floods in the yarkand river basin from 1961 to 2021. *J. Glaciol. Geocryol.* **2023**, *45*, 1288–1299.
24. Jingrong, W. Investigation and analysis on the causes of sudden glacial floods in the yarkand river of xinjiang. *Bull. Soiland Water Conserv.* **1990**, *10*, 33–38.
25. Haemmig, C.; Huss, M.; Keusen, H.; Hess, J.; Wegmüller, U.; Ao, Z.; Kulubayi, W. Hazard assessment of glacial lake outburst floods from Kyagar glacier, Karakoram mountains, China. *Ann. Glaciol.* **2014**, *55*, 34–44. [CrossRef]
26. Gao, Y.; Wang, J.; Liu, S.; Yao, X.; Qi, M.; Liang, P.; Xie, F.; Mu, J.; Ma, X. Monitoring dynamics of kyagar glacier surge and repeated draining of ice-dammed lake using multi-source remote sensing. *Sci. Total Environ.* **2024**, *928*, 172467. [CrossRef] [PubMed]
27. Feng, Q. Characteristics of glacier outburst flood in the yarkant river, karakorum mountains. *GeoJournal* **1991**, *25*, 255–263. [CrossRef]
28. Round, V.; Leinss, S.; Huss, M.; Haemmig, C.; Hajnsek, I. Surge dynamics and lake outbursts of Kyagar glacier, Karakoram. *Cryosphere* **2017**, *11*, 723–739. [CrossRef]
29. Li, G.; Lv, M.; Quincey, D.J.; Taylor, L.S.; Li, X.; Yan, S.; Sun, Y.; Guo, H. Characterizing the surge behaviour and associated ice-dammed lake evolution of the Kyagar glacier in the Karakoram. *Cryosphere* **2023**, *17*, 2891–2907. [CrossRef]
30. Liu, J.; Milne, R.I.; Zhu, G.-F.; Spicer, R.A.; Wambulwa, M.C.; Wu, Z.-Y.; Boufford, D.E.; Luo, Y.-H.; Provan, J.; Yi, T.-S.; et al. Name and scale matter: Clarifying the geography of tibetan plateau and adjacent mountain regions. *Glob. Planet. Chang.* **2022**, *215*, 103893. [CrossRef]
31. Guo, W.; Liu, S.; Xu, J.; Wu, L.; Shangguan, D.; Yao, X.; Wei, J.; Bao, W.; Yu, P.; Liu, Q.; et al. The second Chinese glacier inventory: Data, methods and results. *J. Glaciol.* **2015**, *61*, 357–372. [CrossRef]
32. Liu, S.; Yao, X.; Guo, W.; Xu, J.; Shangguan, D.; Wei, J.; Bao, W.; Wu, L. The contemporary glaciers in china based on the second chinese glacier inventory. *Acta Geogr. Sin.* **2015**, *70*, 3–16.
33. Torres, R.; Snoeijs, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.; Floury, N.; Brown, M. Gmes sentinel-1 mission. *Remote Sens. Environ.* **2012**, *120*, 9–24. [CrossRef]
34. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]
35. Feng, Q.; Wang, Q.; Huang, H.; Wang, Z. Improved otsu algorithm for water extraction based on sentinel-1 dual-polarization data. *Remote Sens. Technol. Appl.* **2022**, *37*, 1003–1011.
36. Gao, B.-C. Normalized difference water index for remote sensing of vegetation liquid water from space. In *Imaging Spectrometry*; SPIE: Bellingham, WA, USA, 1995; pp. 225–236.
37. Du, Y.; Zhang, Y.; Ling, F.; Wang, Q.; Li, W.; Li, X. Water bodies' mapping from sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the swir band. *Remote Sens.* **2016**, *8*, 354. [CrossRef]
38. Cohen, S.; Raney, A.; Munasinghe, D.; Loftis, J.D.; Molthan, A.; Bell, J.; Rogers, L.; Galantowicz, J.; Brakenridge, G.R.; Kettner, A.J.; et al. The floodwater depth estimation tool (fwdet v2. 0) for improved remote sensing analysis of coastal flooding. *Nat. Hazards Earth Syst. Sci.* **2019**, *19*, 2053–2065. [CrossRef]
39. Cohen, S.; Brakenridge, G.R.; Kettner, A.; Bates, B.; Nelson, J.; McDonald, R.; Huang, Y.F.; Munasinghe, D.; Zhang, J. Estimating floodwater depths from flood inundation maps and topography. *JAWRA J. Am. Water Resour. Assoc.* **2018**, *54*, 847–858. [CrossRef]

40. Peter, B.G.; Cohen, S.; Lucey, R.; Munasinghe, D.; Raney, A.; Brakenridge, G.R. Google earth engine implementation of the floodwater depth estimation tool (fwdet-gee) for rapid and large scale flood analysis. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 1501005. [CrossRef]
41. Parida, B.R.; Tripathi, G.; Pandey, A.C.; Kumar, A. Estimating floodwater depth using sar-derived flood inundation maps and geomorphic model in kosi river basin (india). *Geocarto Int.* **2022**, *37*, 4336–4360. [CrossRef]
42. Cohen, S.; Peter, B.G.; Haag, A.; Munasinghe, D.; Moragoda, N.; Narayanan, A.; May, S. Sensitivity of remote sensing floodwater depth calculation to boundary filtering and digital elevation model selections. *Remote Sens.* **2022**, *14*, 5313. [CrossRef]
43. Sattar, A.; Haritashya, U.K.; Kargel, J.S.; Leonard, G.J.; Shugar, D.H.; Chase, D.V. Modeling lake outburst and downstream hazard assessment of the lower barun glacial lake, nepal himalaya. *J. Hydrol.* **2021**, *598*, 126208. [CrossRef]
44. Leoul, A.; Kassahun, N. Dam breach analysis using hec-ras and hec-georas: The case of kesem kebena dam. *Open J. Mod. Hydrol.* **2019**, *9*, 113–142. [CrossRef]
45. Thun, L.J.V.; Gillette, D.R. *Guidance on Breach Parameters*; U.S. Department of the Interior, Bureau of Reclamation: Washington, DC, USA, 1990.
46. Limerinos, J.T. *Determination of the Manning Coefficient from Measured Bed Roughness in Natural Channels*; US Government Printing Office: Washington, DC, USA, 1970.
47. Tang, W.; Zhou, Q.; Liu, X.; Zhu, W.; Mao, X. Anlysis on verification of national severe convective weather categorical forecasts. *J. Meteorol. Res.* **2017**, *43*, 67–76.
48. Casati, B.; Wilson, L.; Stephenson, D.; Nurmi, P.; Ghelli, A.; Pocerlich, M.; Damrath, U.; Ebert, E.; Brown, B.; Mason, S. Forecast verification: Current status and future directions. *Meteorol. Appl. J. Forecast. Pract. Appl. Train. Tech. Model.* **2008**, *15*, 3–18. [CrossRef]
49. Lin, Z.; Yao, X.; Du, J.; Zhou, Z. Refined evaluation of satellite precipitation products against rain gauge observations along the sichuan—Tibet railway. *J. Meteorol. Res.* **2022**, *36*, 779–797. [CrossRef]
50. Cao, C.; Chen, K.; Chen, Y.; Liu, J.; Chen, J.; He, J.; Chen, X. A novel method of error evaluation for radar nowcasting based on shape matching. *J. Meteorol. Res.* **2017**, *43*, 987–997.
51. Peng, L.; Ao, Z.; Yuan, B.; Jiang, H. Study on volume interpolation calculation of kyagar glacier lake based on dem. *Yellow River* **2015**, *37*, 21–23.
52. Zhang, Z.; Zhao, J.; Liu, S.; Zhang, Q.; Jiang, Z.; Xu, Y.; Su, H. Characterization of three surges of the kyagar glacier, karakoram. *Remote Sens.* **2023**, *15*, 2113. [CrossRef]
53. Wang, G.; Hu, Y.; Zhang, S.; Ru, Y.; Chen, K.; Wu, M. Water identification from the gf-1 satellite image based on the deep convolutional neural networks. *Natl. Remote Sens. Bull.* **2022**, *26*, 2304–2316. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## Article

# Identifying Alpine Lakes with Shoreline Features

Zhimin Hu <sup>1</sup>, Min Feng <sup>2,3,\*</sup>, Yijie Sui <sup>2</sup>, Dezhao Yan <sup>2,4</sup>, Kuo Zhang <sup>2,4</sup>, Jinhao Xu <sup>2</sup>, Rui Liu <sup>1</sup> and Earina Sthapit <sup>5</sup><sup>1</sup> College of Geography and Tourism, Chongqing Normal University, Chongqing 401331, China<sup>2</sup> National Tibetan Plateau Data Center, State Key Laboratory of Tibetan Plateau Earth System, Environment and Resources, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China<sup>3</sup> Academy of Plateau Science and Sustainability, Qinghai Normal University, Xining 810016, China<sup>4</sup> University of Chinese Academy of Sciences, Beijing 100049, China<sup>5</sup> Central Department of Hydrology and Meteorology, Tribhuvan University, Kathmandu 44600, Nepal

\* Correspondence: mfeng@itpcas.ac.cn

**Abstract:** Alpine lakes located in high-altitude mountainous regions act as vital sentinels of environmental change. Remote-sensing-based identification of these lakes is crucial for understanding their response to climate variations and for assessing associated disaster risks. However, the complex terrain and weather conditions in these areas pose significant challenges to accurate detection. This paper proposes a method that leverages the high precision of deep learning for small lake and lake boundary extraction combined with deep learning to eliminate noise and errors in the identification results. Using Sentinel-2 data, we accurately identified and delineated alpine lakes in the eastern Himalayas. A total of 2123 lakes were detected, with an average lake area of 0.035 km<sup>2</sup>. Notably, 76% of these lakes had areas smaller than 0.01 km<sup>2</sup>. The slope data is crucial for the lake classification model in eliminating shadow noise. The accuracy of the proposed lake classification model reached 97.7%. In the identification of small alpine lakes, the recognition rate of this method was 96.4%, significantly surpassing that of traditional deep learning approaches. Additionally, this method effectively eliminated most shadow noise present in water body detection results obtained through machine learning techniques.

**Keywords:** alpine lake; Sentinel-2 Imagary; deep learning; ConvNeXt; shoreling feature

## 1. Introduction

Alpine lakes, located in mountainous regions with steep slopes, low temperatures, high solar radiation, intense ultraviolet radiation, and prolonged snow and ice coverage, are highly sensitive to environmental and climate changes, serving as indicators of global environmental shifts [1]. In the Himalayas, approximately 15% of the area is covered by glaciers [2]. Due to ongoing global warming, the region has witnessed accelerated glacier retreat and permafrost melting, particularly affecting alpine lakes within glacier-covered areas. These lakes are responding more rapidly to glacier retreat and climate change, resulting in accelerated lake expansion, with both their areas and numbers increasing rapidly [3–5]. Such rapid changes significantly heighten the risk of glacial lake outburst floods (GLOFs) and landslides [6–9], posing severe threats to downstream populations and their property [10]. The Himalayan region is now one of the most vulnerable areas to GLOFs globally [11,12]. Therefore, accurately identifying alpine lakes is crucial for understanding the impacts of global climate change on these lakes and for improving disaster prediction efforts [13].

Remote sensing technology currently serves as the primary method for large-scale alpine lake surveys, and it has attracted significant attention for its use in identifying alpine lakes [14,15]. However, the substantial terrain relief and frequent cloud cover in alpine regions result in extensive mountain and cloud shadows, which appear very similar



to water bodies in remote sensing images [16], severely affecting the accuracy of water body detection. Additionally, the limited ice-free season in alpine lake distribution areas, coupled with interference from clouds and snow, makes it challenging to distinguish alpine lakes from seasonal snowmelt. Moreover, within the same region, lakes can exhibit varying spectral characteristics due to differences in water turbidity, composition, and depth, further complicating the accurate identification of alpine lakes [17].

Currently, the methods for extracting alpine lakes from remote sensing data include visual interpretation, water indices, machine learning, and deep learning. Visual interpretation relies on expert knowledge to identify alpine lakes. Although it achieves high accuracy, it is limited by economic and time constraints, making it feasible only for small areas [18,19]. Water indices are commonly used indicators for remote sensing of water bodies and have been effectively applied on large scales, even globally. For instance, the Normalized Difference Water Index (NDWI) [20] highlights the distinction between water bodies and vegetation, performing well in densely vegetated areas. The Modified Normalized Difference Water Index (MNDWI) [21] offers some differentiation for shadows and is therefore widely used in the extraction of alpine lakes. The Automated Water Extraction Index (AWEI) [22] enhances the detection of dark features, allowing for water body extraction in areas with low reflectance, making it suitable for shadow-prone regions. Despite the widespread use of these water indices for extracting alpine lakes, a single index is often insufficient to effectively identify lakes across different environments and regions. This underscores the need for more generalized models to improve the accuracy of alpine lake recognition.

Machine learning methods for water body extraction can effectively integrate multi-band and index information to automatically capture the spectral characteristics of land features. This reduces background noise interference, enhances generalization capability, and improves the accuracy and reliability of alpine lake extraction. Common methods for water body extraction include Random Forest, Decision Trees, and Support Vector Machines, which have been successfully applied in regions such as the Himalayas [23]. However, traditional machine learning methods struggle to capture complex environmental features, making it difficult to exclude interferences such as shadows and snow, which necessitates substantial post-processing [8,24,25]. The potential impacts of these errors on downstream applications are significant. For example, in hydrological modeling, shadows affect the spatial heterogeneity of solar radiation in mountainous areas, significantly influencing the spatial distribution of ground temperature, evapotranspiration, snowmelt, and glacier mass balance. Shadows may reduce the quantity and quality of information obtained from these various properties, while also increasing errors and uncertainty in subsequent analyses [26–29].

Deep learning offers a breakthrough in accurately identifying alpine lakes by capturing complex shapes and textures and other features. Compared to traditional pixel-based remote sensing analysis, deep learning has gained increasing attention due to its powerful semantic representation capabilities, advanced feature representation, and robustness in establishing spatial relationships between pixels. ConvNeXt, a modern convolutional neural network architecture [30], significantly enhances the ability to capture environmental features and target land cover with large-scale differences by incorporating certain characteristics from Transformer models. This advancement greatly improves classification accuracy and is well-suited for handling large-scale and complex image classification tasks [31,32].

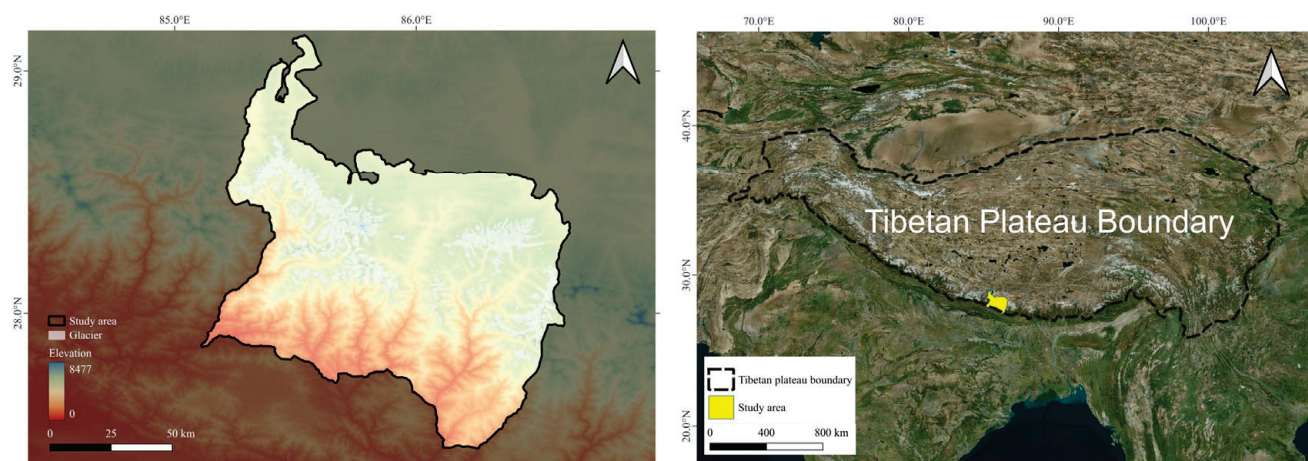
Compared to traditional machine learning methods, deep learning-based identification of alpine lakes has shown a significant improvement in accuracy [33–35]. However, pixel classification methods in deep learning struggle to extract small lakes, particularly those with outlines occupying only a few pixels, as they often fail to adequately capture the contextual information of the surrounding environment. This limitation makes it challenging for the model to accurately distinguish between water bodies and the background, negatively impacting recognition accuracy. Water bodies, such as lakes, are geographically

continuous; however, in deep learning segmentation models, images must be divided into multiple tiles due to computational constraints or the receptive fields of neurons. The tile sizes are typically limited to  $512 \times 512$  or  $256 \times 256$  pixels, resulting in entire lakes being split into several tiles for input. This segmentation disrupts the continuity of the lakes and increases the difficulty of recognition [36]. Consequently, deep learning methods perform suboptimally in the large-scale identification of alpine lakes. Currently, machine learning and deep learning techniques for identifying alpine lakes are widely applied. Compared to using machine learning or deep learning alone, combining the two methods leverages the strengths of each. Machine learning excels in the sensitivity of identifying individual pixels, making it effective for detecting small lakes, while deep learning more effectively captures the surrounding environmental context of targets, achieving higher classification accuracy [37,38]. Therefore, integrating machine learning with deep learning can enable more efficient and accurate identification of a greater number of alpine lakes.

Considering the rapid growth in the number and area of alpine lakes, many of which remain unidentified, this study proposes a two-stage alpine lake identification method that combines a spectral water mapping approach with the strengths of deep learning in capturing the complex environmental characteristics around lakes. First, Sentinel-2 data (European Space Agency, Paris, France) from ice-free seasons between 2019 and 2023 are used to extract the preliminary water coverage area using the RF algorithm, incorporating water indices and visible bands. Building on this, the ConvNeXt model is used to classify and validate alpine lakes versus non-alpine lakes, leveraging the surrounding environmental and textural information. This method applies to alpine lakes of varying sizes and accurately identifies the full extent of alpine lakes. It contributes not only to understanding the spatial distribution and characteristics of alpine lakes but also provides crucial data and scientific support for identifying potentially hazardous lakes, assessing GLOF risks, and estimating alpine lake water storage.

## 2. Study Area

The eastern Himalayas ( $85.1^\circ$  to  $86.6^\circ$ E) extend from east to west, with an average elevation of 3971 m (Figure 1). The region contains 142 continental glaciers, covering a total area of 203 km<sup>2</sup>, most of which are slowly retreating [39]. The area experiences a tropical monsoon climate, with an average annual temperature of  $-2.8^\circ\text{C}$ , and shows a trend of gradual warming year by year [40,41]. Influenced by the Indian Ocean monsoon, the southern part of the Eastern Himalayas receives abundant precipitation and is densely vegetated, while the northern part is dry, with arid conditions. The region experiences cloud cover for most of the year, with clear skies approximately half of the time. Snow cover persists for extended periods, and lakes remain icebound for long durations, with perennial ice and snow coverage. Compared to the Western Himalayas, the Eastern Himalayan region hosts larger and more numerous alpine lakes and exhibits more severe glacial loss [42]. Additionally, the warming rate in this region surpasses the global average, making it one of the areas most susceptible to the impacts of climate change [43]. The landscape in this area is fragile and highly susceptible to natural disasters. In recent years, as the climate has warmed and glaciers have gradually melted, the number of alpine lakes in the region has increased sharply, significantly raising the risk of glacial lake outburst floods (GLOFs). This region has become one of the most frequent and densely occurring areas for GLOFs [44,45].



**Figure 1.** Geographic extent and elevation distribution of the study area.

### 3. Data and Methods

#### 3.1. Data

The Sentinel series of satellites was developed by the European Space Agency (ESA). The Sentinel-2 series is a multispectral high-resolution imaging mission for land monitoring, operated by two polar orbiting satellites that can revisit the same area every five days. The onboard multispectral instrument (MSI) records 13 spectral bands ranging from the visible spectrum to near-infrared (NIR) and short-wave infrared (SWIR), with spatial resolutions ranging from 10 m to 60 m. This study utilized all 10-m resolution Sentinel-2 bands from July to October 2019–2023 (late summer to early autumn), specifically bands B2 (Blue), B3 (Green), B4 (Red), B8 (NIR), and the 20-m resolution B11 (SWIR1) band to calculate various water indices (NDVI, NDWI, MNDWI) [46] to represent the characteristics of water bodies and the surrounding environments of alpine lakes. For the Sentinel-2 SWIR1 band, the à trous wavelet transform (ATWT) algorithm was employed for pan-sharpening, resampling the 20-m resolution to 10-m resolution [47]. Compared to other seasons, the proportion of clouds, fog, and seasonal snowfall in the images during this period is relatively low. The formulas for calculating the above water indices are as follows:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad (2)$$

$$MNDWI = \frac{Green - SWIR1}{Green + SWIR1} \quad (3)$$

The ALOS PALSAR Radiometric Terrain Correction (RTC) Digital Elevation Model (DEM) dataset was developed by the Alaska Satellite Facility (ASF, Fairbanks, AK, USA) and provides elevation information at resolutions of 12.5 m and 30 m. In this study, we resampled the 12.5-m resolution dataset to 10-m resolution using bilinear interpolation to match the spatial resolution of the Sentinel images. Although interpolation artifacts may occur during the resampling process, our evaluation and validation indicate that these artifacts have a minimal impact on the study results, thereby ensuring the reliability and validity of the data. Subsequently, we utilized ArcGIS (Esri, Redlands, CA, USA, version 10.6) to calculate the terrain slope values in the study area based on the ALOS DEM. The resulting terrain slope data can also provide useful information for distinguishing between water and non-water bodies [48,49].

### 3.2. Methods

#### 3.2.1. Water Body Coverage Extraction

In the study area, we randomly selected 100,000 sample points, which were classified into water and non-water bodies through manual interpretation. Among these, 70% (70,000 points) were used for training the model, while 30% (30,000 points) were utilized to validate the model's accuracy, ensuring that the data accurately represented both the training and validation datasets.

Using the method proposed by Sui, a water body classification model was constructed. This model utilized the data values from bands B4, B3, and B2, along with NDVI, NDWI, and MNDWI for each sample point, and input these values into a random forest classification model to differentiate between water and non-water pixels. The resulting water body classification model was applied to Sentinel-2 images for each year, predicting the presence of water on a pixel-by-pixel basis, yielding a water probability value for each pixel, and generating annual water probability maps. To ensure the stability of the detection, only areas with a water probability exceeding 50% throughout the year were classified as water bodies.

This method was repeatedly applied from 2019 to 2023, generating annual water coverage for each year. Finally, by integrating these annual datasets, the maximum water extent between 2019 and 2023 was determined, resulting in comprehensive final water coverage. The final water coverage was converted from raster format to vector format using the ArcGIS (version 10.6) raster-to-vector conversion tool, yielding a total of 25,326 polygons.

#### 3.2.2. Identification of Alpine Lakes

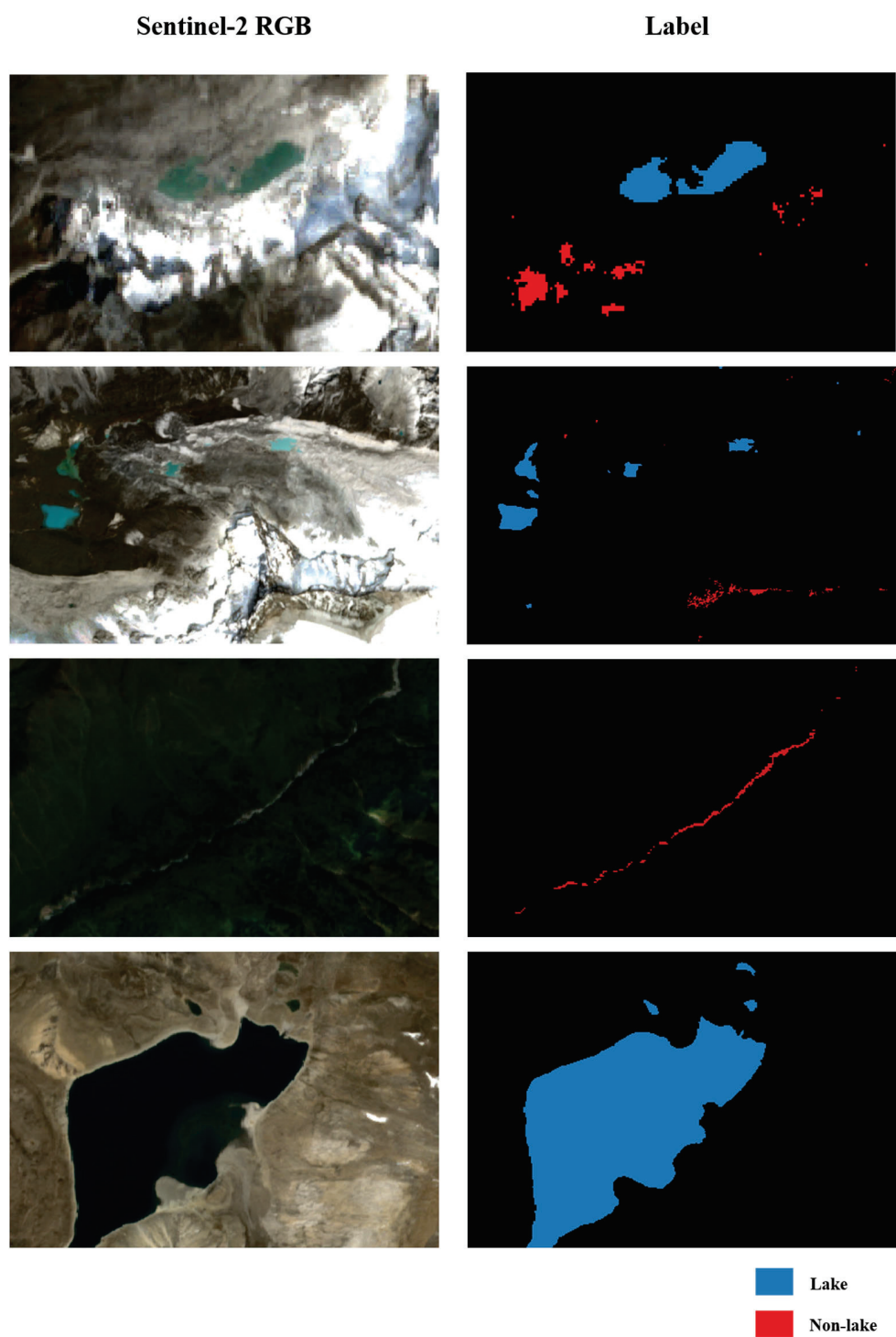
##### a. ConvNeXt classification model

Due to the significant topographical variations in mountainous areas and the perennial snow cover, mountain and cloud shadows can closely resemble water body indices, causing substantial interference in the water body identification model. This has resulted in numerous non-lake noise artifacts, such as snow, shadows, and rivers, within the generated water body coverage.

In this study, the second step focuses on utilizing the surrounding environmental information of water body coverage to classify lakes and non-lake areas. Firstly, we randomly selected 3000 water body coverage polygons. By comparing these with visible light images from Sentinel-2, we manually annotated all lake and non-lake areas. Lakes were classified as positive samples, while non-lakes were categorized as negative samples (Figure 2).

To accurately extract environmental information surrounding individual lakes, we compared the minimum distances between all non-adjacent lake samples and found the shortest distance to be 50 m. Therefore, we extracted Sentinel-2 RGB imagery, slope values, and MNDWI within a 50-m buffer zone around the water body coverage. The Sentinel-2 RGB and MNDWI images were median composites derived from 2019–2023 data. These images contain the water body coverage along with surrounding environmental imagery, water body indices, and topographical information. The resulting images were resampled to  $512 \times 512$  pixels to ensure data consistency. For areas outside the buffer zone, no data values were reassigned to 0, allowing for proper model training without impacting its performance. Data augmentation techniques, such as rotation, flipping, and scaling, were applied to increase the diversity of the training data, improving the model's generalization ability and preventing overfitting. In total, 1250 positive sample images and 4750 negative sample images were obtained. Of these, 70% from each category were used for model training, while 30% were used to validate the model's accuracy (Table 1).





**Figure 2.** Sentinel-2 RGB true-color imagery and corresponding labels for classification.

**Table 1.** Number of the total, positive, and negative samples used for training and validation in the study area.

|          | Training | Validation | Total |
|----------|----------|------------|-------|
| Positive | 875      | 375        | 1250  |
| Negative | 3325     | 1425       | 4076  |
| Total    | 4950     | 1050       | 6000  |

Simultaneously, to assess the impact of the aforementioned data on the model's ability to identify lakes, the Sentinel-2 RGB imagery, slope values, and MNDWI were combined in various ways and used as input data for model classification. As shown in Table 2 below:

**Table 2.** Data input combination table for classification model.

| Train Data | Input             |
|------------|-------------------|
| Input1     | RGB               |
| Input2     | RGB, Slope        |
| Input3     | RGB, MNDWI        |
| Input4     | RGB, Slope, MNDWI |

This method selects the ConvNeXt classification model as the core for the lake classification task, primarily due to its outstanding performance in complex environments, especially its enhanced adaptability in handling diverse images and transformation uncertainties. Compared to other convolutional neural network architectures, ConvNeXt strikes a good balance between performance and computational efficiency, effectively capturing detailed information without sacrificing speed. Other models, such as DeepLabV3, enhance the ability to capture contextual information through the use of zero convolution, but they struggle with precision when dealing with complex boundaries and small targets [50]. Additionally, the extensive use of zero convolutions can lead to higher computational and storage demands [51,52]. From a macro design perspective, ConvNeXt features four main characteristics. First, it utilizes a four-stage architecture with a computation ratio adjusted to (3, 3, 9, 3), representing the number of blocks in each stage, which is considered the optimal distribution of computations. Second, ConvNeXt begins the network with non-overlapping  $4 \times 4$  convolutions, aggressively downsampling the input images. Third, following the strategy proposed by ResNeXt, ConvNeXt employs depthwise convolutions and  $1 \times 1$  convolutions to separate mixed spatial and channel information. Fourth, inspired by Transformer blocks, ConvNeXt incorporates a reverse bottleneck structure and reintroduces large-sized convolutions.

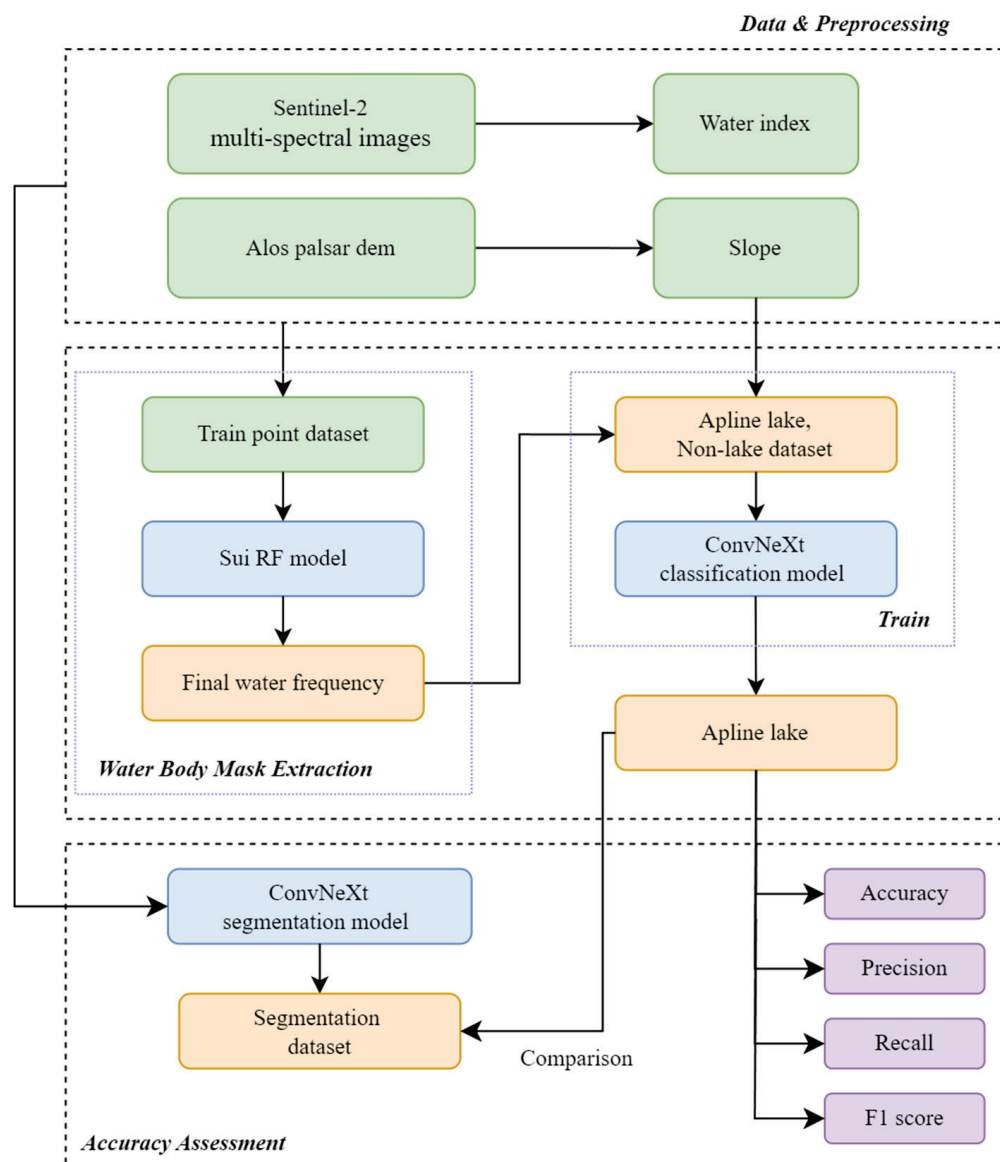
At the micro-design level, ConvNeXt utilizes fewer activation functions and normalization layers, similar to Transformer models. Additionally, ConvNeXt enhances performance by replacing ReLU with GELU and substituting batch normalization (BN) with layer normalization (LN).

In the basic settings for model training, we configured 32 epochs with a batch size of 8, meaning each batch contained 8 samples. The number of steps per epoch was calculated by dividing the total number of samples by the batch size and rounding up using the ceiling function, ensuring the entire dataset was traversed in each epoch. The initial learning rate was fixed at 0.001 to mitigate experimental randomness. The learning rate followed an exponential decay strategy, decreasing exponentially with the current training steps. As the steps increased, the learning rate gradually decreased to adapt to different training phases. The chosen optimizer was Adam, with weight decay set to 0.001 as a regularization method to prevent overfitting. Cross-entropy loss was used as the loss function; given that the number of negative samples significantly exceeded that of positive samples (with a ratio of 1:4), sample weights were assigned to adjust the model's training process and mitigate the impact of class imbalance. During training, two primary evaluation metrics were utilized: loss and mean intersection over union (MIoU). PyTorch served as the deep learning framework, and the network was trained until convergence across 8 GPUs (NVIDIA Tesla, NVIDIA Corporation, Santa Clara, CA, USA).

#### b. ConvNeXt segmentation model

At the same time, to compare the effectiveness of the alpine lakes extracted using the methods of this study with those obtained through deep learning segmentation algorithms, the RGB imagery and other mentioned bands for the entire study area were cropped into  $512 \times 512$ -pixel blocks. A total of 3254 blocks that overlap with positive samples of all lakes

were randomly divided into training and testing sets in a 7:3 ratio for input data into the segmentation models. The segmentation model selected was the ConvNeXt segmentation model, with the same general parameters as those used in the classification algorithm of this method. (Figure 3).



**Figure 3.** Flow diagram of the methods in the study.

### 3.2.3. Accuracy Assessment

Standard deep learning statistical metrics were calculated to assess the accuracy of the model in extracting the number of lakes. True positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs) were determined by comparing the predictions with reference data. Accuracy, precision, recall, and F1 score are standard metrics for evaluating the performance of segmentation models. Accuracy refers to the proportion of correctly predicted samples to the total number of samples. Precision indicates the ratio of predicted lakes to the actual lakes. Recall measures the proportion of true lakes correctly predicted by the model against the total number of lakes. The F1 Score is determined by

the harmonic mean of precision and recall and is often regarded as a comprehensive metric for assessing model performance. Their definitions are as follows:

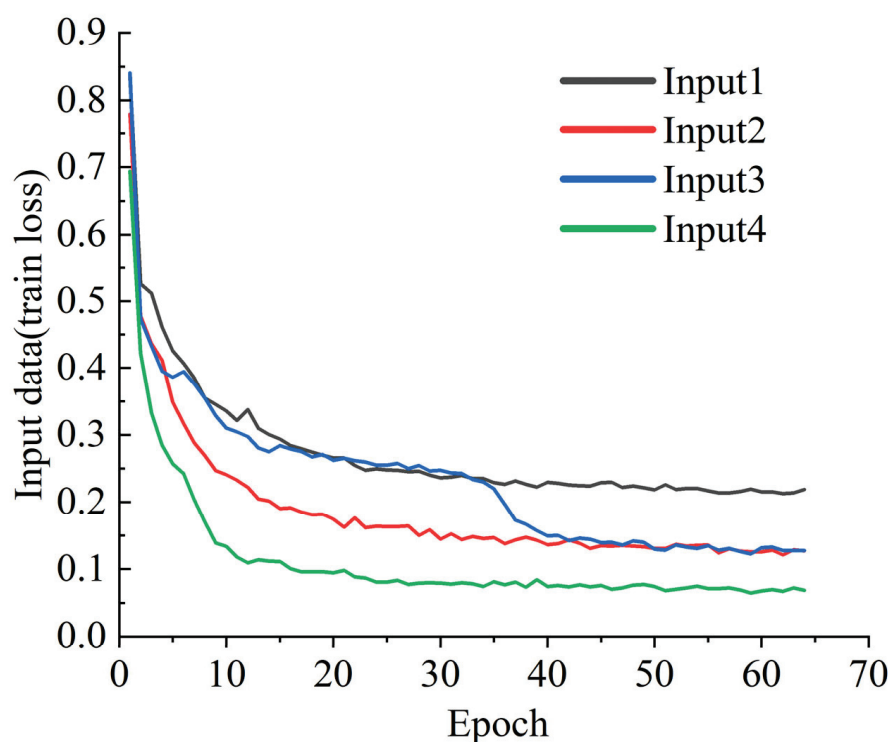
$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

The training loss for alpine lake classification models with different inputs is shown in Figure 4. After 64 iterations, all training loss curves stabilized, indicating that each model had achieved the expected fitting performance. For the Input1 configuration, the minimum loss value was 0.212, which is significantly higher than the minimum loss values of the other three input combinations. For Input2 and Input3, prior to 37 iterations, the loss of Input3 was consistently higher than that of Input2, but after 37 iterations, their loss values converged, with Input2 reaching a minimum loss of 0.121 and Input3 a minimum of 0.123. The Input4 configuration performed the best, achieving the lowest loss of 0.064 after 61 iterations.



**Figure 4.** Training loss for different input data configurations. Input1 consists of Sentinel-2 RGB images; Input2 combines Sentinel-2 RGB images with slope data; Input3 includes Sentinel-2 RGB images with MNDWI; and Input4 integrates Sentinel-2 RGB images with both slope and MNDWI data.

In terms of comparing the accuracy, precision, recall, and F1 score across the different input data combinations, Input1 yielded the lowest values across all metrics, likely because it only uses RGB imagery as input. When slope or MNDWI data was added, all metrics improved, with Input2 outperforming Input3, likely due to the slope data aiding in distinguishing between water bodies and shadows. Input4 produces the best results, with the



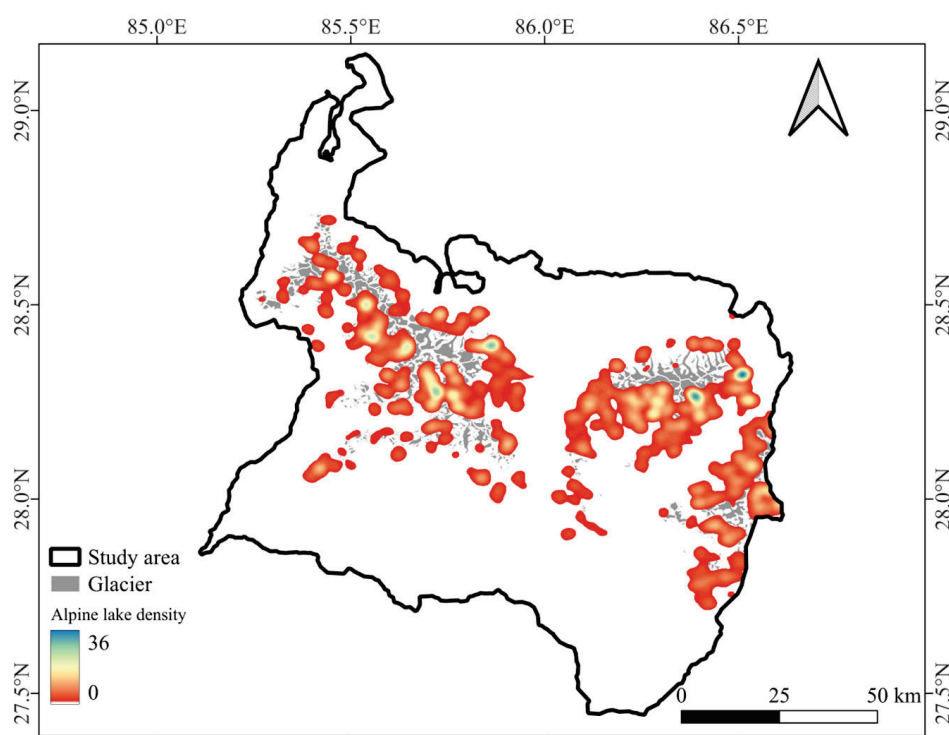
highest values for all metrics, as follows: 97.7% accuracy, 97.7% precision, 96% recall, and 96.8% F1 score (Table 3). The average training time per epoch for Input1 was 449 s, and for Input4 was 886 s. This suggests that as the diversity of input data increases, the accuracy of the alpine lake classification model also improves. The final model trained using the Input4 combination was applied to water coverage across the study area, and by repeating the same steps, the alpine lakes throughout the entire region were identified.

**Table 3.** Performance evaluation of input combinations.

| Input  | Accuracy | Precision | Recall | F1 Score | Time (s) |
|--------|----------|-----------|--------|----------|----------|
| Input1 | 0.812    | 0.747     | 0.792  | 0.762    | 449      |
| Input2 | 0.935    | 0.959     | 0.950  | 0.954    | 596      |
| Input3 | 0.882    | 0.829     | 0.868  | 0.845    | 516      |
| Input4 | 0.977    | 0.977     | 0.960  | 0.968    | 886      |

#### 4. Results

In the study area, a total of 2123 alpine lakes were identified using this method, covering an area of 75.263 km<sup>2</sup>, with an average area of 0.035 km<sup>2</sup> and an average elevation of 5120 m, resulting in a total water storage capacity of 2.937 km<sup>3</sup> [53] (Figure 5). Notably, a large proportion of these alpine lakes (76%) have an area smaller than 0.01 km<sup>2</sup>, categorizing them as small water bodies [54]. Although they are numerous, their total area amounts to only 4.062 km<sup>2</sup>, accounting for just 5% of the total area of all alpine lakes in the study area. Lakes larger than 1 km<sup>2</sup> are relatively rare, with only 14 such lakes identified; however, their combined area reaches 37.588 km<sup>2</sup>, which constitutes half (50%) of the total area of all alpine lakes. (Figure 6) The overall distribution of these lakes spans elevations between 3800 and 6000 m. Most alpine lakes (approximately 55%) are situated between 5000 and 5500 m, covering an area of 57.61 km<sup>2</sup>, which is 77% of the total alpine lake area. Among these, 866 lakes are classified as small alpine lakes, representing up to 74%. Additionally, there are 605 alpine lakes (28%) located at elevations between 4500 and 5000 m.



**Figure 5.** Distribution density map of alpine lakes in the study area.

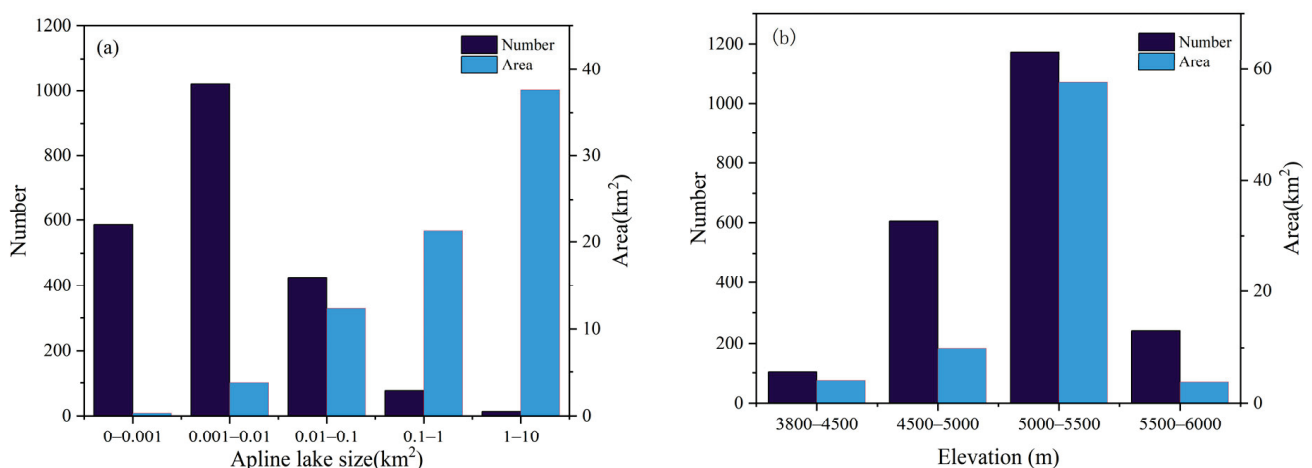


Figure 6. (a) Distribution of alpine lakes by size, (b) distribution of alpine lakes by elevation.

## 5. Discussion

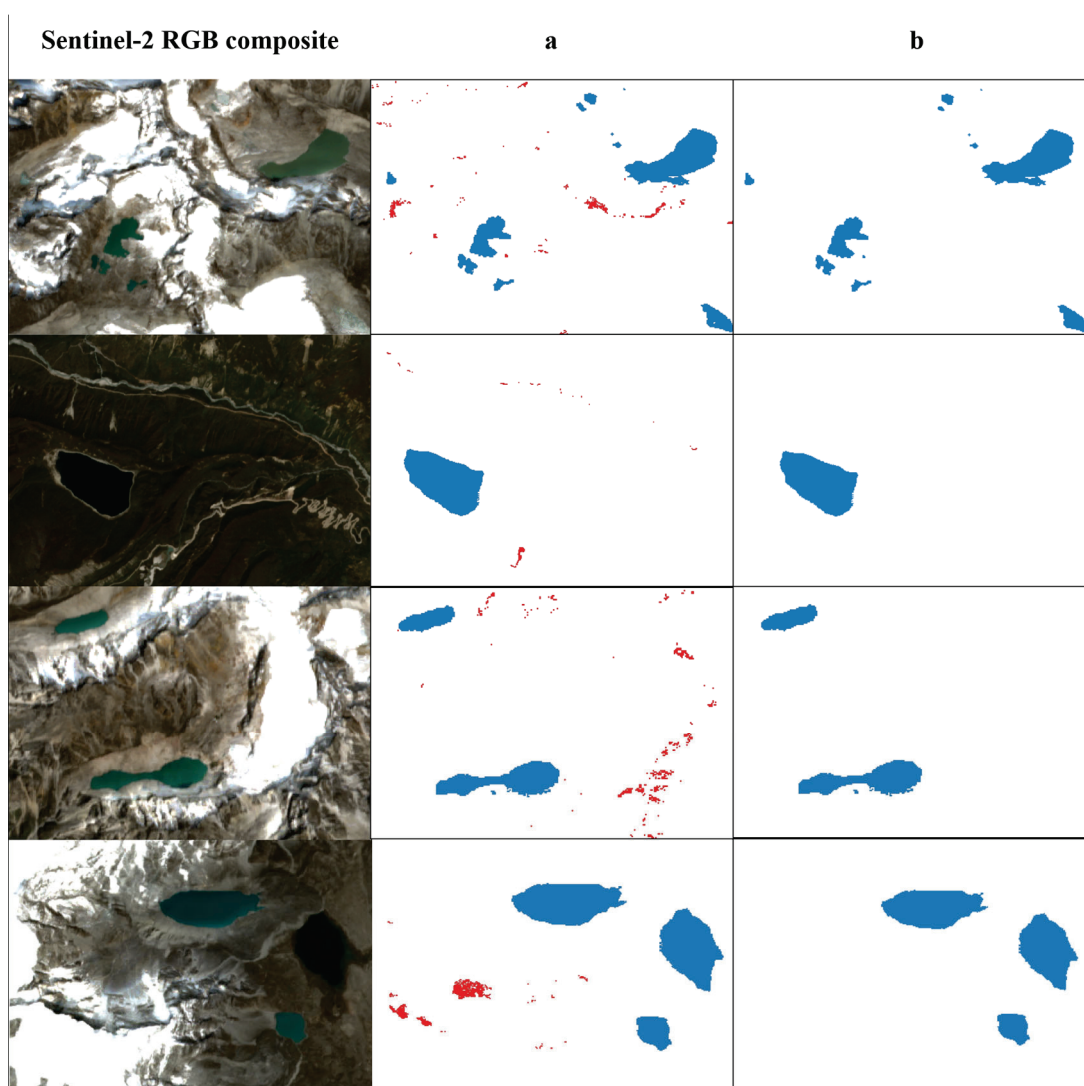
### 5.1. Improving Alpine Lake Mapping Accuracy

In this study area, using the ConvNeXt segmentation model alone, 1532 alpine lakes were identified, with a total area of 69.87 km<sup>2</sup>. However, both the number and area of alpine lakes extracted by this method were less than those identified using the two-step extraction method proposed in this study. Through the RF model, 25,326 water bodies were identified. After visual interpretation and comparison with the RGB and MNDWI bands of the Sentinel-2 imagery, it was found that many of the coverage represented non-lake water bodies such as mountain shadows, snow cover, and fragmented river channels. This indicates that in complex alpine environments, although the RF model can extract water bodies, the results are prone to errors and cannot accurately represent alpine lake information. In contrast, our method classifies all water body coverage using a classification model, effectively eliminating non-alpine water bodies, such as shadows and snow, and accurately identifying alpine lakes, as shown in Figure 7.

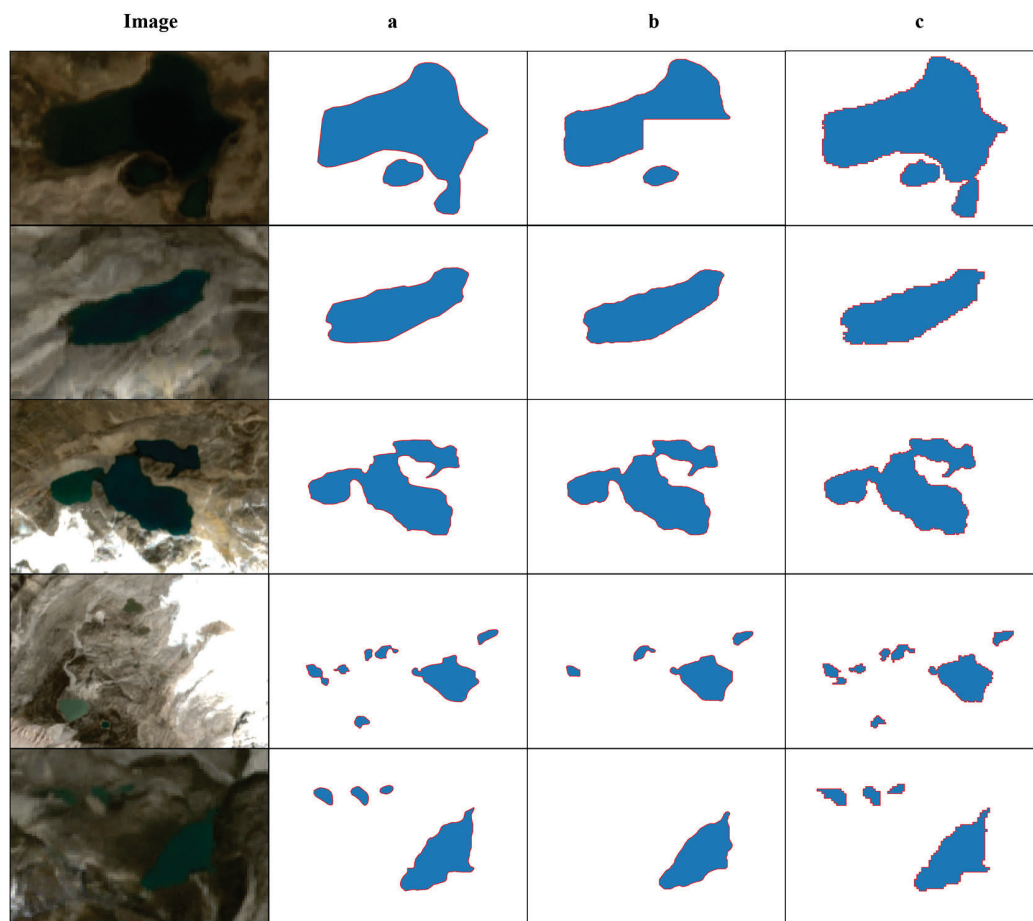
Taking the Sentinel-2 RGB imagery as the base map, five random locations were selected to demonstrate the base imagery, the visually interpreted results, and the alpine lake predictions from both methods to evaluate the consistency between the predicted results and the actual imagery (Figure 8). It can be observed that the RF method performs poorly in error control, resulting in numerous misidentified non-alpine lakes. Compared to the RF model, our method significantly reduces the misidentification of non-lake water bodies. Although the RF model shows certain advantages in large-scale water body extraction, its ability to identify alpine lakes in complex terrain is limited, as it is easily affected by mountain shadows, snow cover, and fragmented river channels, leading to a high error rate. In contrast, our method excels in this aspect, effectively reducing errors and ensuring the reliability of the identification results. The ConvNeXt segmentation model performs well in identifying alpine lakes larger than 0.1 km<sup>2</sup>, delineating their boundaries, though its precision is still lower than that of our method. Furthermore, these methods perform poorly in identifying small lakes, particularly those smaller than 0.01 km<sup>2</sup>, resulting in significant omissions. For alpine lakes larger than 0.1 km<sup>2</sup>, a comparison of boundaries with manually interpreted results revealed an average boundary difference of 2 pixels. For lakes smaller than 0.1 km<sup>2</sup>, the average boundary difference was 1 pixel. The method proposed in this study demonstrates strong performance in identifying alpine lakes of varying sizes, with higher precision and reliability in delineating lake morphology. Due to the significant seasonal influence on alpine lakes in mountainous regions, particularly from snow cover and ice formation, satellite observations of surface water can be hindered. Therefore, to minimize the effects of seasonal variability, this study exclusively selected images collected in late summer and early autumn. By combining the advantages of deep learning with traditional image processing techniques, this method achieves comprehensive

detection of alpine lakes. Notably, in complex alpine terrains, this method demonstrates greater robustness in overcoming shadows, snow cover, and other interfering factors. Compared to the extraction results of a single model, our method more effectively separates lakes from non-lake water bodies, reduces misidentification rates, improves the recognition of small alpine lakes, and enhances the overall detection accuracy of alpine lakes.

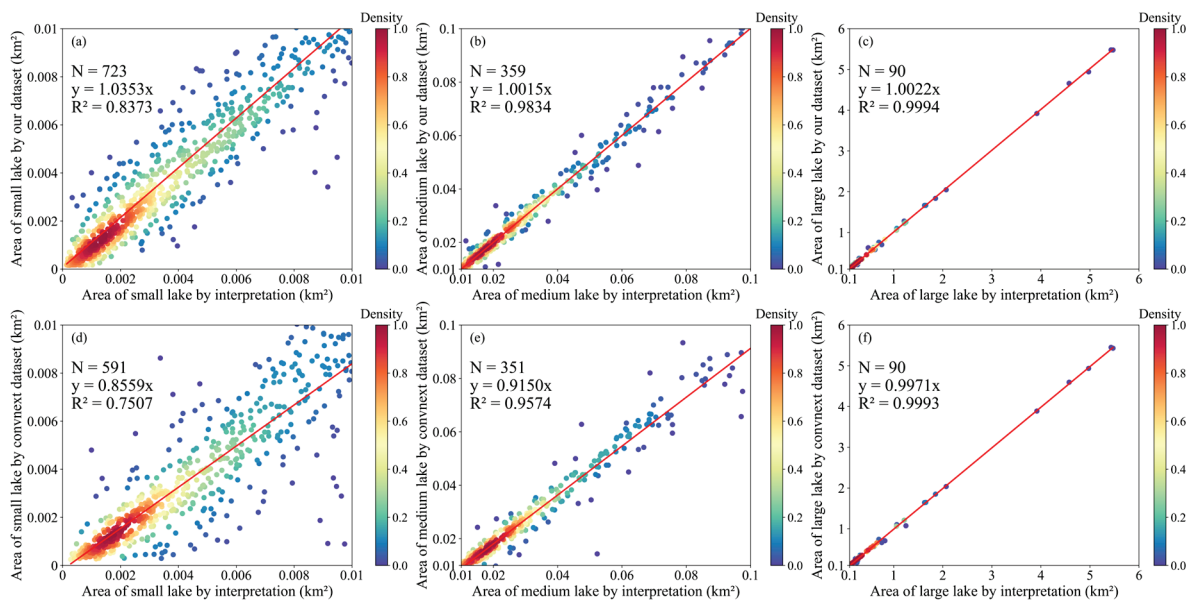
In this study area, the proposed method demonstrated higher accuracy in identifying alpine lakes of different sizes, including small lakes (less than 0.01 km<sup>2</sup>), medium-sized lakes (0.01 to 0.1 km<sup>2</sup>), and large lakes (greater than 0.1 km<sup>2</sup>) (Figure 9). Among the visually interpreted dataset of small alpine lakes, 750 lakes were randomly selected for area comparison with the results from the two methods. Our method identified 723 small lakes, achieving a recognition rate of 96.4%, which is significantly higher than the 591 lakes identified by ConvNeXt. The R<sup>2</sup> value of our method was 0.837, also higher than the 0.751 of ConvNeXt.



**Figure 7.** Sentinel-2 RGB composite displaying alpine lakes; (a) shows RF method results; (b) shows results from the proposed method. Blue indicates alpine lakes and red represents noise.



**Figure 8.** Sentinel-2 RGB composite displaying alpine lakes, with columns labeled as follows: (a) for the manual interpretation results, (b) for the ConvNeXt model segmentation results, and (c) for the results obtained from the proposed method.



**Figure 9.** The area comparison between the dataset from this method and the manually interpreted water body dataset. (a–c) Comparison of classification results with manual interpretation for small, medium, and large alpine lakes. (d–f) Comparison of segmentation results with manual interpretation for small, medium, and large alpine lakes.



In the visually interpreted dataset of medium-sized alpine lakes, 370 lakes were randomly selected for area comparison with the results from the two methods. Our method identified 359 lakes, with a recognition rate of 97%, higher than the 351 lakes identified by ConvNeXt. Moreover, the  $R^2$  value of our method was 0.983, outperforming ConvNeXt's 0.957. For large lakes, both methods performed excellently, achieving high recognition accuracy. In the dataset of 90 visually interpreted large alpine lakes, both methods successfully identified the lakes with minimal area discrepancies. Overall, the proposed method exhibited higher accuracy in identifying alpine lakes of all sizes, particularly excelling in the identification and delineation of small alpine lakes. In contrast, the ConvNeXt segmentation method was less effective in recognizing small alpine lakes, though it performed well in identifying medium and large lakes. Due to the RF method's high rate of misidentification in complex alpine terrains, it was excluded from this comparative analysis.

## 5.2. Limitations And Future Directions

The results of this study exhibit a high degree of consistency with the interpreted data; however, some errors and inaccuracies persist. The complex climate of the Himalayan region, characterized by persistent cloud cover, results in significant variability in image quality. Consequently, we utilized summer fusion images from Sentinel-2 acquired between 2019 and 2023, which limited our ability to observe the interannual dynamics of alpine lakes. Future research should consider integrating high-spatiotemporal resolution optical satellite data with radar satellite data that can penetrate cloud cover, thereby enhancing the accuracy of identifying the geographical features and interannual dynamics of alpine lakes.

In this study, we utilized the ConvNeXt classification model for water body identification, achieving robust performance in distinguishing aquatic features from various noise artifacts. Nevertheless, machine learning and deep learning approaches often necessitate extensive annotated datasets for training, demanding considerable time and human resources. Moreover, optimizing model parameters to effectively navigate complex geographic environments remains a significant challenge. Future research should further explore the classification of alpine lake types, such as glacial lakes, thermokarst lakes, and non-glacial lakes, while also developing a global dataset of alpine lakes with interannual temporal resolution. This initiative will enhance our understanding of the distribution and dynamics of alpine lakes, providing critical insights for identifying potentially hazardous lakes and assessing the impacts of global climate change on these vulnerable ecosystems.

## 6. Conclusions

This study presents an innovative method that combines machine learning and deep learning for the accurate extraction of alpine lakes. The approach first employs a RF model to delineate water bodies, and then utilizes the environmental features surrounding these water bodies to leverage the powerful classification capabilities of deep learning, effectively eliminating misidentified areas. This method successfully addresses the impact of noise, such as mountainous shadows, on lake extraction, allowing for the effective delineation of numerous small alpine lakes in mountainous regions. Compared to traditional machine learning methods, this approach minimizes the impact of interference from mountain shadows, snow cover, and river channels on the identification of alpine lakes, while retaining the capability to accurately identify small alpine lakes. It addresses the issues of missed and incorrect identifications often associated with deep learning segmentation models. The lake classification model in this method utilizes Sentinel-2 data and slope information as input, achieving an accuracy of 97.7%, a precision of 97.7%, a recall of 96%, and an F1 score of 96.8%. The identification accuracy for alpine lakes in mountainous areas reached 96.9%, which is 14.6% higher than that of the ConvNeXt segmentation model. This demonstrates the method's effectiveness in precisely characterizing alpine lakes and highlights its significant potential for mapping these water bodies.

**Author Contributions:** Conceptualization, Z.H. and M.F.; methodology, Z.H. and Y.S.; writing—original draft preparation, Z.H.; writing—review and editing, M.F. and Y.S.; formal analysis, Z.H.; validation, K.Z., J.X. and E.S.; visualization, D.Y. and R.L.; funding acquisition, M.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key Research and Development Program of China [Grant number: 2023YFF0725005], the National Natural Science Foundation of China [Grant number: 42171140], Science and Technology Department of Tibet Program [Grant number: XZ202301ZY0035G], and the TPESER Youth Innovation Key Program [Grant number: TPESER-QNCX2022ZD-04].

**Data Availability Statement:** Data will be made available on request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Parker, B.R.; Vinebrooke, R.D.; Schindler, D.W. Recent Climate Extremes Alter Alpine Lake Ecosystems. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 12927–12931. [CrossRef] [PubMed]
2. Song, C. Remote Sensing of Alpine Lake Water Environment Changes on the Tibetan Plateau and Surroundings: A Review. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 26–37. [CrossRef]
3. Yao, T.; Pu, J.; Lu, A.; Wang, Y.; Yu, W. Recent Glacial Retreat and Its Impact on Hydrological Processes on the Tibetan Plateau, China, and Surrounding Regions. *Arct. Antarct. Alp. Res.* **2007**, *39*, 642–650. [CrossRef]
4. Carrivick, J.L.; Tweed, F.S.; Sutherland, J.L.; Mallalieu, J. Toward Numerical Modeling of Interactions Between Ice-Marginal Proglacial Lakes and Glaciers. *Front. Earth Sci.* **2020**, *8*, 577068. [CrossRef]
5. Ye, Q.; Zhu, L.; Zheng, H.; Naruse, R.; Zhang, X.; Kang, S. Glacier and Lake Variations in the Yamzhog Yumco Basin, Southern Tibetan Plateau, from 1980 to 2000 Using Remote-Sensing and GIS Technologies. *J. Glaciol.* **2007**, *53*, 673–676. [CrossRef]
6. Sattar, A.; Haritashya, U.K.; Kargel, J.S.; Karki, A. Transition of a Small Himalayan Glacier Lake Outburst Flood to a Giant Transborder Flood and Debris Flow. *Sci. Rep.* **2022**, *12*, 12421. [CrossRef]
7. Zhang, T.; Wang, W.; An, B. Heterogeneous Changes in Global Glacial Lakes under Coupled Climate Warming and Glacier Thinning. *Commun. Earth Environ.* **2024**, *5*, 374. [CrossRef]
8. Shugar, D.H.; Burr, A.; Haritashya, U.K.; Kargel, J.S.; Watson, C.S.; Kennedy, M.C.; Bevington, A.R.; Betts, R.A.; Harrison, S.; Strattman, K. Rapid Worldwide Growth of Glacial Lakes since 1990. *Nat. Clim. Chang.* **2020**, *10*, 939–945. [CrossRef]
9. Haeberli, W.; Schaub, Y.; Huggel, C. Increasing Risks Related to Landslides from Degrading Permafrost into New Lakes in De-Glaciating Mountain Ranges. *Geomorphology* **2017**, *293*, 405–417. [CrossRef]
10. Das, S.; Kar, N.S.; Bandyopadhyay, S. Glacial Lake Outburst Flood at Kedarnath, Indian Himalaya: A Study Using Digital Elevation Models and Satellite Images. *Nat. Hazards* **2015**, *77*, 769–786. [CrossRef]
11. Wang, X.; Liu, S.; Ding, Y.; Guo, W.; Jiang, Z.; Lin, J.; Han, Y. An Approach for Estimating the Breach Probabilities of Moraine-Dammed Lakes in the Chinese Himalayas Using Remote-Sensing Data. *Nat. Hazards Earth Syst. Sci.* **2012**, *12*, 3109–3122. [CrossRef]
12. Fujita, K.; Sakai, A.; Nuimura, T.; Yamaguchi, S.; Sharma, R.R. Recent Changes in Imja Glacial Lake and Its Damming Moraine in the Nepal Himalaya Revealed by in Situ Surveys and Multi-Temporal ASTER Imagery. *Environ. Res. Lett.* **2009**, *4*, 045205. [CrossRef]
13. Sattar, A.; Goswami, A.; Kulkarni, A.V.; Emmer, A.; Haritashya, U.K.; Allen, S.; Frey, H.; Huggel, C. Future Glacial Lake Outburst Flood (GLOF) Hazard of the South Lhonak Lake, Sikkim Himalaya. *Geomorphology* **2021**, *388*, 107783. [CrossRef]
14. Quincey, D.J.; Richardson, S.D.; Luckman, A.; Lucas, R.M.; Reynolds, J.M.; Hambrey, M.J.; Glasser, N.F. Quincey Early Recognition of Glacial Lake Hazards in the Himalaya Using Remote Sensing Datasets. *Glob. Planet. Chang.* **2007**, *56*, 137–152. [CrossRef]
15. Luo, D.-L.; Jin, H.-J.; Du, H.-Q.; Li, C.; Ma, Q.; Duan, S.-Q.; Li, G.-S. Variation of Alpine Lakes from 1986 to 2019 in the Headwater Area of the Yellow River, Tibetan Plateau Using Google Earth Engine. *Adv. Clim. Chang. Res.* **2020**, *11*, 11–21. [CrossRef]
16. Giles, P. Remote Sensing and Cast Shadows in Mountainous Terrain. *Photogramm. Eng. Remote Sens.* **2001**, *67*, 833–839.
17. Fisher, A.; Flood, N.; Danaher, T. Comparing Landsat Water Index Methods for Automated Water Classification in Eastern Australia. *Remote Sens. Environ.* **2016**, *175*, 167–182. [CrossRef]
18. Schepaschenko, D.; See, L.; Lesiv, M.; Bastin, J.-F.; Mollicone, D.; Tsendbazar, N.-E.; Bastin, L.; McCallum, I.; Laso Bayas, J.C.; Baklanov, A.; et al. Recent Advances in Forest Observation with Visual Interpretation of Very High-Resolution Imagery. *Surv. Geophys.* **2019**, *40*, 839–862. [CrossRef]
19. Winsvold, S.H.; Käbb, A.; Nuth, C. Regional Glacier Mapping Using Optical Satellite Data Time Series. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3698–3711. [CrossRef]
20. McFeeters, S.K. The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]
21. Xu, H. Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [CrossRef]

22. Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S.R. Automated Water Extraction Index: A New Technique for Surface Water Mapping Using Landsat Imagery. *Remote Sens. Environ.* **2014**, *140*, 23–35. [CrossRef]
23. Chen, F.; Zhang, M.; Tian, B.; Li, Z. Extraction of Glacial Lake Outlines in Tibet Plateau Using Landsat 8 Imagery and Google Earth Engine. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4002–4009. [CrossRef]
24. Nie, Y.; Liu, Q.; Liu, S. Glacial Lake Expansion in the Central Himalayas by Landsat Images, 1990–2010. *PLoS ONE* **2013**, *8*, e83973. [CrossRef]
25. Zhang, G.; Bolch, T.; Yao, T.; Rounce, D.R.; Chen, W.; Veh, G.; King, O.; Allen, S.K.; Wang, M.; Wang, W. Underestimated Mass Loss from Lake-Terminating Glaciers in the Greater Himalaya. *Nat. Geosci.* **2023**, *16*, 333–338. [CrossRef]
26. Zhang, Y.L.; Li, X.; Cheng, G.D.; Jin, H.J.; Yang, D.W.; Flerchinger, G.N.; Chang, X.L.; Wang, X.; Liang, J. Influences of Topographic Shadows on the Thermal and Hydrological Processes in a Cold Region Mountainous Watershed in Northwest China. *J. Adv. Model. Earth Syst.* **2018**, *10*, 1439–1457. [CrossRef]
27. Gruber, S.; Fleiner, R.; Guegan, E.; Panday, P.; Schmid, M.-O.; Stumm, D.; Wester, P.; Zhang, Y.; Zhao, L. Review Article: Inferring Permafrost and Permafrost Thaw in the Mountains of the Hindu Kush Himalaya Region. *Cryosphere* **2017**, *11*, 81–99. [CrossRef]
28. Harris, C.; Arenson, L.; Christiansen, H.; Etzelmüller, B.; Frauenfelder, R.; Gruber, S.; Haeberli, W.; Hauck, C.; Hoelzle, M.; Humlum, O.; et al. Permafrost and Climate in Europe: Monitoring and Modelling Thermal, Geomorphological and Geotechnical Responses. *Earth-Sci. Rev.* **2009**, *92*, 117–171. [CrossRef]
29. Omar, P.J.; Kumar, V. Land Surface Temperature Retrieval from TIRS Data and Its Relationship with Land Surface Indices. *Arab. J. Geosci.* **2021**, *14*, 1897. [CrossRef]
30. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Curran Associates, Inc.: New York, NY, USA, 2017; Volume 30.
31. Li, J.; Wang, C.; Huang, B.; Zhou, Z. ConvNeXt-Backbone HoVerNet for Nuclei Segmentation and Classification. *arXiv* **2022**, arXiv:2202.13560.
32. Zhu, Y.; Yuan, K.; Zhong, W.; Xu, L. Spatial-Spectral ConvNeXt for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 5453–5463. [CrossRef]
33. Bai, Y.; Wu, W.; Yang, Z.; Yu, J.; Zhao, B.; Liu, X.; Yang, H.; Mas, E.; Koshimura, S. Enhancement of Detecting Permanent Water and Temporary Water in Flood Disasters by Fusing Sentinel-1 and Sentinel-2 Imagery Using Deep Learning Algorithms: Demonstration of Sen1Floods11 Benchmark Datasets. *Remote Sens.* **2021**, *13*, 2220. [CrossRef]
34. Chen, F. Comparing Methods for Segmenting Supra-Glacial Lakes and Surface Features in the Mount Everest Region of the Himalayas Using Chinese GaoFen-3 SAR Images. *Remote Sens.* **2021**, *13*, 2429. [CrossRef]
35. Wang, G.; Wu, M.; Wei, X.; Song, H. Water Identification from High-Resolution Remote Sensing Images Based on Multidimensional Densely Connected Convolutional Neural Networks. *Remote Sens.* **2020**, *12*, 795. [CrossRef]
36. Li, Y.; Dang, B.; Zhang, Y.; Du, Z. Water Body Classification from High-Resolution Optical Remote Sensing Imagery: Achievements and Perspectives. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 306–327. [CrossRef]
37. Sui, Y.; Feng, M.; Wang, C.; Li, X. A High-Resolution Inland Surface Water Body Dataset for the Tundra and Boreal Forests of North America. *Earth Syst. Sci. Data* **2022**, *14*, 3349–3363. [CrossRef]
38. Xu, J.; Feng, M.; Sui, Y.; Yan, D.; Zhang, K.; Shi, K. Identifying Alpine Lakes in the Eastern Himalayas Using Deep Learning. *Water* **2023**, *15*, 229. [CrossRef]
39. Liu, H.; Chen, Y.; Ye, Z.; Li, Y.; Zhang, Q. Climate Change in the Himalayas: Current State of Knowledge. *Sci. Rep.* **2019**, *9*, 16277. [CrossRef]
40. Wang, B.; Bao, Q.; Hoskins, B.; Wu, G.; Liu, Y. Tibetan Plateau Warming and Precipitation Changes in East Asia. *Geophys. Res. Lett.* **2008**, *35*. [CrossRef]
41. Xu, Z.X.; Gong, T.L.; Li, J.Y. Decadal Trend of Climate in the Tibetan Plateau—Regional Temperature and Precipitation. *Hydrol. Process.* **2008**, *22*, 3056–3065. [CrossRef]
42. Gardelle, J.; Arnaud, Y.; Berthier, E. Contrasted Evolution of Glacial Lakes along the Hindu Kush Himalaya Mountain Range between 1990 and 2009. *Glob. Planet. Chang.* **2011**, *75*, 47–55. [CrossRef]
43. Shrestha, U.B.; Gautam, S.; Bawa, K.S. Widespread Climate Change in the Himalayas and Associated Changes in Local Ecosystems. *PLoS ONE* **2012**, *7*, e36741. [CrossRef] [PubMed]
44. Zhang, G.; Yao, T.; Xie, H.; Wang, W.; Yang, W. An Inventory of Glacial Lakes in the Third Pole Region and Their Changes in Response to Global Warming. *Glob. Planet. Chang.* **2015**, *131*, 148–157. [CrossRef]
45. Carrivick, J.L.; Tweed, F.S. A Global Assessment of the Societal Impacts of Glacier Outburst Floods. *Glob. Planet. Chang.* **2016**, *144*, 1–16. [CrossRef]
46. Barredo, J.I.; Demicheli, L.; Lavalle, C.; Kasanko, M.; McCormick, N. Modelling Future Urban Scenarios in Developing Countries: An Application Case Study in Lagos, Nigeria. *Environ. Plan. B* **2004**, *31*, 65–84. [CrossRef]
47. Du, J.; Kimball, J.S.; Jones, L.A.; Watts, J.D. Implementation of Satellite Based Fractional Water Cover Indices in the Pan-Arctic Region Using AMSR-E and MODIS. *Remote Sens. Environ.* **2016**, *184*, 469–481. [CrossRef]
48. Li, J.; Sheng, Y. An Automated Scheme for Glacial Lake Dynamics Mapping Using Landsat Imagery and Digital Elevation Models: A Case Study in the Himalayas. *Int. J. Remote Sens.* **2012**, *33*, 5194–5213. [CrossRef]

49. Mitkari, K.V.; Arora, M.K.; Tiwari, R.K. Extraction of Glacial Lakes in Gangotri Glacier Using Object-Based Image Analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 5275–5283. [CrossRef]
50. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef]
51. Zheng, J. ConvNeXt-Mask2Former: A Semantic Segmentation Model for Land Classification in Remote Sensing Images. In Proceedings of the 2024 5th International Conference on Computer Vision, Image and Deep Learning (CVIDL), Zhuhai, China, 19–21 April 2024; pp. 676–682.
52. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. In Proceedings of the International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.
53. Zhang, G.; Carrivick, J.L.; Emmer, A.; Shugar, D.H.; Veh, G.; Wang, X.; Labedz, C.; Mergili, M.; Mölg, N.; Huss, M.; et al. Characteristics and Changes of Glacial Lakes and Outburst Floods. *Nat. Rev. Earth Environ.* **2024**, *5*, 447–462. [CrossRef]
54. Mullen, A.L.; Watts, J.D.; Rogers, B.M.; Carroll, M.L.; Elder, C.D.; Noomah, J.; Williams, Z.; Caraballo-Vega, J.A.; Bredder, A.; Rickenbaugh, E.; et al. Using High-Resolution Satellite Imagery and Deep Learning to Track Dynamic Seasonality in Small Water Bodies. *Geophys. Res. Lett.* **2023**, *50*, e2022GL102327. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



MDPI AG  
Grosspeteranlage 5  
4052 Basel  
Switzerland  
Tel.: +41 61 683 77 34

MDPI Books Editorial Office  
E-mail: [books@mdpi.com](mailto:books@mdpi.com)  
[www.mdpi.com/books](http://www.mdpi.com/books)



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Topic Editors. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editors and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.





Academic Open  
Access Publishing

[mdpi.com](https://mdpi.com)

ISBN 978-3-7258-8844-3