

Article

Comparison of Machine Learning Models in Simulating Glacier Mass Balance: Insights from Maritime and Continental Glaciers in High Mountain Asia

Weiwei Ren ¹, Zhongzheng Zhu ¹, Yingzheng Wang ², Jianbin Su ^{1,*}, Ruijie Zeng ³, Donghai Zheng ¹ and Xin Li ¹

¹ National Tibetan Plateau Data Center (TPDC), State Key Laboratory of Tibetan Plateau Earth System Science, Environment and Resources (TPESER), Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China; renweiwei@itpcas.ac.cn (W.R.); zhuzz@itpcas.ac.cn (Z.Z.); zhengd@itpcas.ac.cn (D.Z.); xinli@itpcas.ac.cn (X.L.)

² College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China; wangyzh20@lzu.edu.cn

³ School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85281, USA; ruijie.zeng.1@asu.edu

* Correspondence: jianbinsu@itpcas.ac.cn

Abstract: Accurately simulating glacier mass balance (GMB) data is crucial for assessing the impacts of climate change on glacier dynamics. Since physical models often face challenges in comprehensively accounting for factors influencing glacial melt and uncertainties in inputs, machine learning (ML) offers a viable alternative due to its robust flexibility and nonlinear fitting capability. However, the effectiveness of ML in modeling GMB data across diverse glacier types within High Mountain Asia has not yet been thoroughly explored. This study addresses this research gap by evaluating ML models used for the simulation of annual glacier-wide GMB data, with a specific focus on comparing maritime glaciers in the Niyang River basin and continental glaciers in the Manas River basin. For this purpose, meteorological predictive factors derived from monthly ERA5-Land datasets, and topographical predictive factors obtained from the Randolph Glacier Inventory, along with target GMB data rooted in geodetic mass balance observations, were employed to drive four selective ML models: the random forest model, the gradient boosting decision tree (GBDT) model, the deep neural network model, and the ordinary least-square linear regression model. The results highlighted that ML models generally exhibit superior performance in the simulation of GMB data for continental glaciers compared to maritime ones. Moreover, among the four ML models, the GBDT model was found to consistently exhibit superior performance with coefficient of determination (R^2) values of 0.72 and 0.67 and root mean squared error (RMSE) values of 0.21 m w.e. and 0.30 m w.e. for glaciers within Manas and Niyang river basins, respectively. Furthermore, this study reveals that topographical and climatic factors differentially influence GMB simulations in maritime and continental glaciers, providing key insights into glacier dynamics in response to climate change. In summary, ML, particularly the GBDT model, demonstrates significant potential in GMB simulation. Moreover, the application of ML can enhance the accuracy of GMB modeling, providing a promising approach to assess the impacts of climate change on glacier dynamics.



Citation: Ren, W.; Zhu, Z.; Wang, Y.; Su, J.; Zeng, R.; Zheng, D.; Li, X. Comparison of Machine Learning Models in Simulating Glacier Mass Balance: Insights from Maritime and Continental Glaciers in High Mountain Asia. *Remote Sens.* **2024**, *16*, 956. <https://doi.org/10.3390/rs16060956>

Academic Editor: Peter Romanov

Received: 23 January 2024

Revised: 5 March 2024

Accepted: 7 March 2024

Published: 8 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: glacier mass balance; machine learning; maritime glaciers; continental glaciers; GBDT

1. Introduction

The glaciers in High Mountain Asia (HMA) serve as vital freshwater reservoirs, playing an essential role in the socio-economic development and eco-environmental sustainability of extensive the HMA region and its surroundings [1,2]. However, most of these glaciers are rapidly contracting and becoming more vulnerable, primarily due to the accelerated

pace of climate warming resulting from human-induced greenhouse gas emissions [3–5]. This rapid glacial melting poses severe threats, including rising sea levels and increased risks of glacial lake outburst floods. Moreover, it is important to note that the dynamics of glacier mass balance (GMB) modelling are intricately linked to climate changes, influencing both the timing and magnitude of transitions in hydrological regimes [6]. Therefore, accurate GMB modeling is imperative in order to assess the impact of climate change on glaciers and provides a vital foundation for elucidating glacier dynamic processes, predicting changes, and effectively managing basin water resources [7–9].

Current simulations of GMB data predominantly rely on process-based physical models, encompassing simple temperature index-based models [10] and more sophisticated energy balance models [11]. While energy balance models can theoretically offer reliable GMB estimations, their accuracy heavily relies on the reliability of input data and the effectiveness of parameter calibrations. Meanwhile, their complex structure demands substantial computing resources and time, constraining their widespread application over large areas. On the other hand, temperature index-based models, establishing empirical relationships between temperature and glacial melt data [10], provide a more straightforward approach suitable for large-scale areas. However, their exclusive dependence on temperature and precipitation inputs may oversimplify glacial processes. Moreover, a fundamental assumption in most temperature index-based models is the constancy of degree-day factors (DDFs). However, this assumption has been challenged by studies, such as by Ismail et al. [12], which indicate a decreasing trend in DDFs, particularly at higher elevations. This study further highlights the sensitivity of DDFs to environmental changes, particularly factors like solar radiation and albedo. Another limitation in applying process-based physical models to GMB simulations in the HMA region is the severe shortage of meteorological observation data [13], which further leads to high uncertainty in satellite-based and reanalyzed products. Such uncertainties can significantly impact the accuracy of GMB simulations, particularly in large-scale river basins. For instance, Lutz et al. [14] suggested that glacial runoff contributed 40.6% of the total runoff in the Upper Indus River basin. Contrastingly, using the same model in the same basin, Khanal et al. [15] reported that glacial runoff accounted for only 5.1%. These stark discrepancies are often exacerbated by uncertainties in meteorological inputs, highlighting the formidable challenges in physical model-based GMB simulation. Therefore, there is a pressing need for new techniques to better understand the complex dynamics of glaciers and their responses to climate change.

Due to robust non-linear fitting capabilities, machine learning (ML) also has shown significant potential in GMB simulation, even driven by biased input data [16]. For instance, Ren et al. [17] demonstrated that ML models outperformed physical models used to detect runoff simulations. However, the application of ML in GMB simulation is still in its early stages, primarily due to the limited availability of GMB samples. Traditionally, GMB samples rely heavily on direct in situ observations, which can be time-consuming and labor-intensive, resulting in limited sample acquisition. With advancements in remote sensing technology, the availability of GMB data derived from remote sensing observations has rapidly increased in recent years. For example, Hugonnet et al. [18] reconstructed global GMB data spanning nearly two decades using ASTER satellite remote sensing data, thereby providing new opportunities for ML applications in GMB simulation. Initially, ML studies used multiple linear regression to estimate GMB data based on temperature and precipitation [19]. Subsequently, non-linear ML algorithms have been employed. For instance, Bolibar et al. [20] utilized deep neural networks (DNNs) to simulate glacier-wide mass balance in French Alpine glaciers, and Anilkumar et al. [21] employed random forest (RF) and gradient boosting decision tree (GBDT) models to simulate point GMB data in the Alps, both achieving promising results. However, applications of ML in the HMA region have not yet been reported, despite its critical importance in elucidating the impact of climate change on GMB simulations.

Previous research has highlighted significant spatial heterogeneity in GMB changes across the HMA region [22–24], reflecting the complexity of glacial melting across diverse

geographical locales. These variations are influenced by various factors, including glacier types (continental, subcontinental, and maritime), climatic conditions, albedo, and glacier internal temperature, among others [25,26]. Notably, glacier type and its associated climatic environment play a crucial role in determining the rate of glacial retreat. For example, Hugonnet et al. [18] observed that glaciers in the maritime climate of Southeast Tibet experience a higher rate of negative mass loss, likely attributed to decreased precipitation and increased temperatures, which accelerate glacial melting. A warming test conducted by Fujita [27] also indicates higher sensitivities for maritime glaciers characterized by a summer accumulation pattern. Conversely, glaciers in the continental climate of the Karakoram and Kunlun regions exhibit smaller negative mass loss rates, with some even gaining mass. This phenomenon can be attributed to cooler temperatures and lower precipitation, resulting in a slower rate of glacial melting. In other words, maritime glaciers are more sensitive to climate change compared to those in continental regions [27,28]. Therefore, the differing responses of glacier types to climate change must be considered in GMB simulation. In this context, ML provides the potential to develop bespoke models, tailored to the unique characteristics of different glacier types, thus more effectively capturing the complex influencing factors. However, research comparing the performance of ML algorithms specifically on maritime and continental glaciers is still lacking, although it is a critical aspect for advancing the precision of GMB simulation.

Therefore, to enhance the accuracy of GMB simulation across various glacier types in the HMA region, this study aims to evaluate the performance of ML models in estimating annual glacier-wide GMB data, specifically focusing on comparing their performance between maritime and continental glaciers. More specifically, continental glaciers selected for this research are situated in the Manas River basin (MRB) of the Tianshan Mountains, while maritime glaciers are represented by those in the Niyang River basin (NRB) in southeast Tibet. Moreover, this study chose ordinary least-square linear regression (OLS), RF, GBDT, and DNN models to assess their performance in GMB simulation, representing three categories of ML: traditional ML, ensemble learning, and deep learning. Thirdly, geodetic mass balance observations [18] are chosen as target GMB datapoints, and ERA5-Land datasets [29] and Randolph Glacier Inventory 6.0 [30] are used to provide meteorological and topographical predictive factors, respectively. Through this research, three pivotal questions are addressed: (1) What are the differences in performance among the four ML models used for simulating GMB data? (2) How does the performance of these models vary across different glacier types? (3) What are the main factors affecting the effectiveness of ML models used for GMB simulation? By addressing these questions, this study seeks to significantly advance GMB simulation methodologies, which can address the shortcomings arising from process-based physical models due to model structure, parameters, and input data, as well as substantial computational resources and time required. This improvement will facilitate a more precise assessment of the impact of climate change on glaciers. Ultimately, this study aspires to provide valuable insights for glacier dynamic detection, water resource planning, disaster prevention, and climate change adaptation in regions that are particularly vulnerable to the effects of climate change.

2. Study Area and Data

2.1. Study Area

Considering glacier type and water vapor source characteristics, glaciers in the Manas River basin (MRB) of the Tianshan Mountains are chosen to represent continental glaciers, primarily influenced by westerlies [26]. Meanwhile, glaciers in the Niyang River basin (NRB) in southeast Tibet are selected as representatives of maritime glaciers, predominantly influenced by the Indian monsoon [26]. The two river basins are depicted in Figure 1. The Manas River, originating on the northern slope of the Tianshan Mountains, creates the largest artificial oasis of China in the northern part of the HMA [31]. In this study, the MRB refers to the upstream basin of the Manas River, which is regulated by the Kensiwate station (Figure 1). This basin covers a drainage area of 5156 km² and is geographically distributed

between 84.5° – 86.5° E and 43° – 44° N. The MRB exhibits a distinct seasonal discharge pattern, with approximately 80% of the flow occurring from June to September. Furthermore, glacial runoff comprises 27% of annual flow, thereby playing a crucial role in its annual runoff regulation [32]. According to the Randolph Glacier Inventory (RGI) version 6 [30], the glacier covers an area of 481.69 km^2 , accounting for 9.32% of the total area in the MRB. Over the last two decades, the average annual rate of glacier mass loss is $0.43 \text{ m w.e. yr}^{-1}$ [18]. Moreover, characterized by its high elevation, often exceeding 3100 m, the MRB receives an annual average precipitation amount of around 550 mm [33]. The annual precipitation in the Manas River basin is predominantly concentrated from April to August, accounting for approximately 70.43% of the average annual precipitation. However, snowfall in the high-mountain areas primarily occurs from November to March of the following year. The precipitation contributes to the presence of perennial snow and glaciers, particularly in lower mountain and hilly areas during the autumn and winter months.

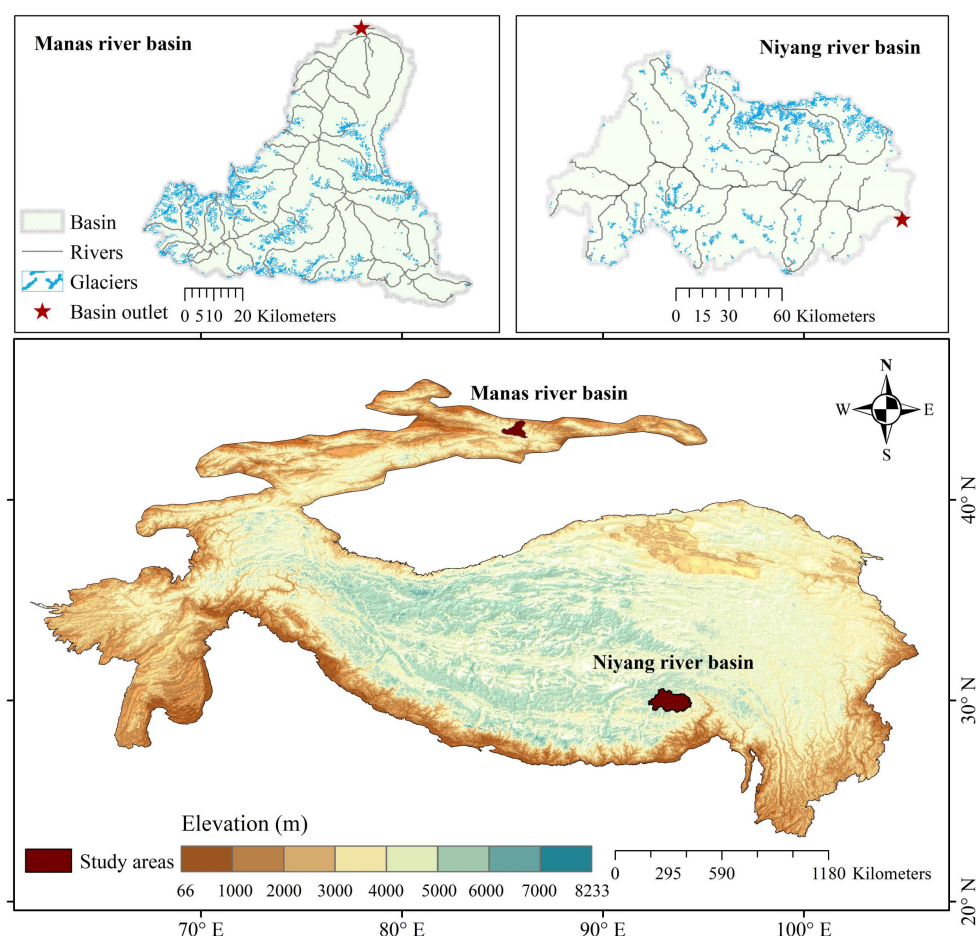


Figure 1. Location of the two headwater catchments.

The Niyang River is the fourth largest tributary in the Yarlung Zangbo River. It meanders for 307.5 km before joining the Brahmaputra at Nyingchi. The NRB is distributed at 92.17° – 94.58° E and 29.47° – 30.52° N, and it is located in the southeastern HMA region (Figure 1). The basin area of the NRB is $16,304 \text{ km}^2$, with elevations ranging from 2924 to 6857 m. It is characterized by a humid temperate climate influenced by the Indian monsoon, with peak temperatures in July and its lowest temperatures in January [34]. The wet season of the NRB lasts from May to October, contributing to the majority of its annual precipitation. The basin has suffered from a steady warming trend, marked by an increase in the annual mean temperature. Recognized as a glacierized region, according to the RGI version 6 [30], the NRB encompasses 953 km^2 of glaciers, accounting for about 5.3%

of its total area. Satellite data have shown a significant reduction in glacial area over the years, with a 30% decrease from 923 km² in 1987 to 650 km² in 2013, equating to an annual shrinkage rate of −0.8%. Importantly, Hugonnet et al. [18] reported an average annual glacier loss rate of 0.81 m w.e. over the past two decades.

2.2. Data

The efficacy of the ML model critically depends on the availability of high-quality training data related to research targets. In this study, GMB samples used for model training were sourced from Hugonnet et al. [18]. This dataset provides global GMB information based on ASTER satellite observations. Covering a 20-year period from 2000 to 2019, it includes annual glacier-wide GMB values for all glaciers worldwide, making it exceptionally comprehensive. Its diverse representation of topographical characteristics, encompassing both maritime and continental glaciers in the HMA, establishes it as an ideal training dataset for the model. For a detailed description, please refer to the paper written by Hugonnet et al. [18].

Topographical factors utilized in the model training are extracted from the RGI version 6 [30]. These factors include a range of glacier characteristics: maximum and minimum glacial elevations, median glacial elevation, glacial area, glacial longitude, glacial latitude, glacial slope, glacial aspect, and glacial length. Due to the lack of weather observations across the HMA, reanalysis datasets are the most viable option [21]. Due to its superior performance and high spatiotemporal resolution ($0.1^\circ \times 0.1^\circ$) [35,36], the ERA5-Land reanalysis dataset [29] is employed to provide meteorological variables (i.e., air temperature and precipitation) of ML models. Employing advanced assimilation techniques, ERA5-Land integrates observational data from various sources, ensuring a higher correlation coefficient (CC) and probability of detection (POD) than IMERG-E and IMERG-L reanalysis precipitation products over the Tibetan Plateau. Its fine spatial resolution and hourly temporal frequency are crucial for elucidating land–atmosphere interactions, rendering it highly valuable for climate research, hydrological modeling, and environmental assessments. This choice is supported by the findings of Wu et al. [35], which suggest that ERA5-Land can capture the long-term spatiotemporal patterns of precipitation. The dataset can be accessed at (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land>) (accessed on 29 December 2023).

According to the RGI [30], the MRB is home to 814 glaciers, while the NRB boasts 1208 glaciers. Due to the low spatial resolution ($0.1^\circ \times 0.1^\circ$) of ERA5-Land meteorological data and the relatively small size of glaciers, to ensure each glacier has representative meteorological variables, this study first established an area threshold for glacier data, selecting only glaciers with an area greater than 0.5 km² as glacier sample data. Then, the ERA5-Land data were bilinearly interpolated to a resolution of 500 m, and, finally, the nearest neighbor method was used to extract the meteorological data corresponding to each grid point glacier. Then, the meteorological variable value for each glacier was calculated using the average of all meteorological data grid points corresponding to the entire glacier. As a result, there were 235 and 363 glacier samples for the MRB and NRB, respectively. Each glacier was subject to 20 years of GMB samples, yielding a total of 4700 samples for the MRB and 7260 samples for the NRB, respectively.

3. Methodology

3.1. Machine Learning Algorithms

ML has witnessed significant evolution, positioning itself at intersects between computer science, statistics, and artificial intelligence. The primary goal of ML is to develop algorithms that enable systems to learn from data, adapt, and make predictions or decisions without explicit programming [37]. The development of ML can be broadly categorized into three phases. (1) Traditional machine learning (1980s): In the early stages, traditional ML algorithms like linear regression, decision trees, and backpropagation artificial neural networks dominated the field. While effective for certain tasks, these algorithms suf-

ferred from limitations in handling complex and unstructured data. (2) Ensemble learning (1990s): This phase was marked by the advent of ensemble learning, exemplified by RF and GBDT models [10]. Ensemble methods enhance the overall performance of ML by combining predictions of multiple models. The RF model, for example, aggregates outputs from multiple decision trees, resulting in enhanced accuracy and robustness. (3) Deep learning (2000s~): Since the early 21st century, deep learning has gained widespread recognition and application. With advancements in computing power and the availability of big data, deep learning has become a focal point in ML and has facilitated numerous industry applications. Particularly after 2012, the surge of deep learning propelled a boom in ML research and a rapid expansion of applications. The advent of deep learning, a subset of ML, inspired by the structure and function of the human brain's neural networks, marked a significant leap. DNNs, a key player in deep learning [10,20], are characterized by multiple layers (deep architectures) that automatically learn hierarchical representations from data. This approach has proven highly effective in tasks like image recognition, natural language processing, and more.

Currently, the ML landscape encompasses a diverse range of algorithms, making it challenging to provide an exhaustive list. The objective of this study is to evaluate the capability of ML in modeling GMB data for various glacier types in the HMA. The selection of ML algorithms in this study is driven by the need to cater to a diverse range of features present in the model input in GMB simulation. Among them, OLS is included as a representative of traditional ML algorithms, while RF and GBDT models are selected to illustrate ensemble learning algorithms. The DNN is chosen to represent deep learning algorithms. The following section will provide a concise overview of the RF, GBDT, and DNN models, elucidating their roles and functionalities in the context of this research.

3.1.1. Random Forest Models

RF classification is a robust ML ensemble method, distinguished for its predictive accuracy and resilience to overfitting. Widely used in geosciences, RF models combine predictions from multiple decision trees, each representing a subset of training data and random features. This approach not only prevents overfitting but also improves its generalization capability. RF models are particularly adept at handling intricate variable relationships, capturing non-linear patterns and managing noisy data, making it a versatile and effective tool in a variety of geoscientific applications [38]. Recently, the RF model has been noted for GMB prediction purposes. By integrating diverse inputs, such as temperature, precipitation, and topographical parameters, RF models can unravel complex patterns of glacier behavior, thereby facilitating assessments of the climate change impacts on glacier dynamics [21].

3.1.2. Gradient Boosting Decision Tree

The GBDT model is a powerful ML algorithm that sequentially constructs decision trees, each designed to rectify the errors of its predecessors. Compared to the RF model, which operates independently, the GBDT model adopts a sequential approach, continually improving model accuracy and predictive performance through iterative refinement. This process of iteratively constructing decision trees allows GBDTs to progressively capture complex relationships between input and output variables. GBDTs excel at managing large-scale datasets and high-dimensional features [39], boosting its various applications in geosciences. For instance, in hydrology, it effectively models complex relationships in rainfall-runoff processes, improving flood prediction accuracy [40]. GBDTs can also be used in GMB modeling, capturing nuanced interactions between climate variables and glacier dynamics. The sequential learning nature of the GBDT model allows it to adapt and refine predictions, making it suitable for diverse geoscientific challenges.

3.1.3. Deep Neural Network

DNNs are advanced ML models inspired by the neural architecture of the human brain. Composed of multiple layers of interconnected nodes, or artificial neurons, they are adept at learning complex patterns and representations from data. This capability makes them particularly effective in extracting hierarchical features. In the field of geosciences, DNNs have shown significant potential across various applications. For instance, in hydrology science, they enhance rainfall–runoff modeling by accurately capturing intricate spatiotemporal patterns [41]. In climate science, they help to improve climate modeling [42], particularly in forecasting temperatures, precipitation levels, and extreme weather events. When applied to glaciology, DNNs can effectively capture the nonlinear response of glaciers to air temperature and precipitation, offering an improved representation of extreme mass balance rates compared to linear statistical and temperature index models [43]. The fundamental principle of DNNs involves learning at different abstraction levels, enabling them to capture intricate patterns and dependencies within geospatial datasets. The advantages of DNNs lie in their automatic learning features, their adaptability to diverse data types, and their robust performance with large-scale datasets. The potential of DNNs in geosciences not only improves predictions but also reveals hidden patterns, which advances our understanding of complex Earth systems.

3.2. Performance Measures

Given the versatility and ease of interpretation, the following two performance measures were used to qualitatively evaluate the performance of ML models: root means squared error (*RMSE*) and coefficients of determination (R^2). The *RMSE* is a commonly used metric to evaluate the accuracy of a regression model. It measures the average deviation between the predicted values of the model and the actual values observed in the data. Mathematically, the *RMSE* is calculated by taking the square root of the average of the squared differences between predicted and observed values. R^2 is a statistical measure that represents the proportion of the variance in the dependent variable that is predictable from the independent variables in a regression model. It ranges from 0 to 1 and is interpreted as the percentage of variation in the dependent variable explained by the independent variables. Detailed expressions of them are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{sim} - y_{obs})^2}{n}} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{obs} - y_{sim})^2}{\sum_{i=1}^n (y_{obs} - \bar{y}_{obs})^2} \quad (2)$$

where n is the number of observations; y_{sim} represents the simulated GMB data; y_{obs} is the target GMB data; and $\bar{y}_{obs}$ denotes the average of the target. The value of 0 indicates that the model does not explain any variability in the dependent variable, whereas the value of 1 indicates that the model explains all the variability. The *RMSE* measures the standard deviation of the differences between the prediction and the target. Lower *RMSE* values are preferable, which signifies a closer agreement between predictions and targets.

3.3. Hyperparameter Selection

Figure 2 depicts the comprehensive workflow employed in this study. The initial step involves preprocessing the curated dataset, which includes the removal of outliers and the normalization of data to enhance the model's adaptability across various data contexts and to improve its generalization performance. Subsequently, the dataset is randomly partitioned, with 70% allocated for model training and the remaining 30% dedicated to evaluating model performance. In the case of the OLS model, the ordinary least-square method is applied for parameter calibration. Given the presence of 60 variables

in this study, each representing a parameter, the parameters of the OLS model are not conveniently displayed.

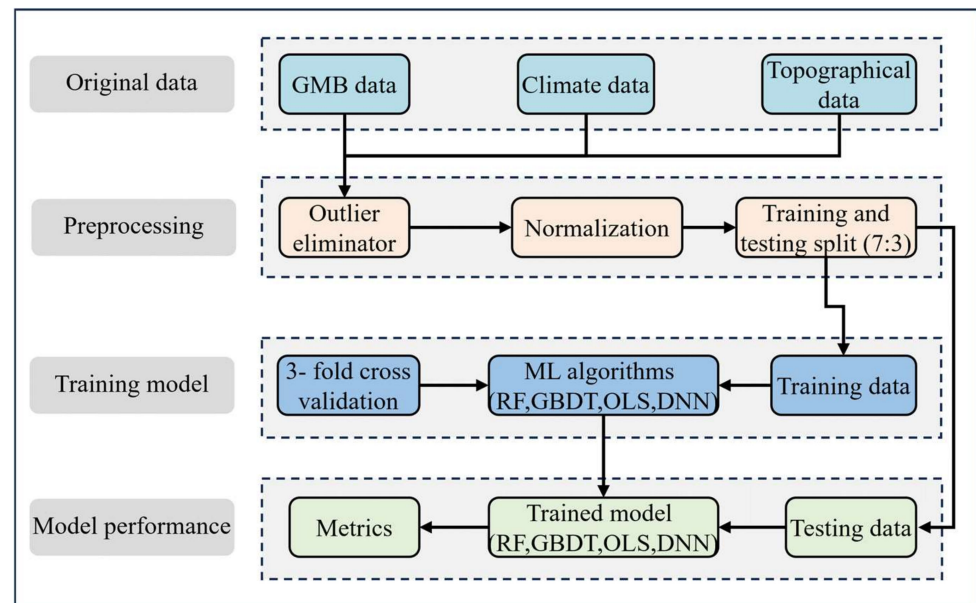


Figure 2. Flowchart of the methodology.

Hyperparameter tuning is crucial in machine learning as it directly influences model performance and generalization. These parameters, set before training, control aspects like model capacity and complexity. GridSearchCV [44] contributes to model optimization by systematically searching through a grid of hyperparameters to find the optimal combination that maximizes model performance. It exhaustively evaluates all possible hyperparameter combinations specified in the grid and uses cross-validation to assess each combination's performance. In this study, the GridSearchCV technique [44] was employed to fine-tune the hyperparameters for the remaining three ML models. Within GridSearchCV, a three-fold cross-validation strategy was chosen to optimize the hyperparameters. As a result, the optimal hyperparameter combinations for the three selected ML models were determined and are presented in Table 1. Finally, the performance of each ML model in simulating GMB data was evaluated using testing data. Additionally, comparative analysis was conducted to assess their performance in simulating GMB data across different types of glaciers.

Table 1. Grid of settings used for the hyperparameter tuning of each model.

Algorithm	Hyperparameter	Value for MRB	Value for NRB
RF	Number of trees	100	120
	Maximum number of features	4	5
	Minimum split samples	5	8
	Maximum depth	None	None
	Minimum samples of leaf nodes	4	6
DNN	Learning rate	0.001	0.001
	Epochs	1000	2000
	Hidden layers	20	20
	Activation function	ReLU	ReLU
GBDT	Learning rate	0.01	0.1
	Number of trees	450	250
	Maximum number of features	3	3
	Minimum split samples	5	4
	Maximum depth	8	8
	Minimum samples of leaf nodes	2	1

4. Results

4.1. Selection of Predictors

According to the simple temperature index-based model [10], temperature and precipitation are primary determinants of glacier accumulation and ablation. Although snowfall significantly affects glacier accumulation, the transition from snowfall to glacier varies regionally, posing challenges to identifying specific snowfall periods that contribute to glacier formation. Therefore, annual, seasonal, and monthly snowfall are considered predictive factors that can be used to clarify the impact of snowfall on glacier accumulation. Temperature and cumulative positive temperature (CPT) play a vital role in glacier ablation. Thus, annual, seasonal, and monthly average temperatures are also considered as predictive factors. CPT primarily influences summer glacier ablation. To fully consider the influence of CPT on glacier ablation, the CPT values in spring, summer, and fall (excluding winter), as well as annual CPT values, are also considered as predictive factors. Additionally, only monthly CPT values from April to October are considered, as temperatures are predominantly below zero from December to the following March. Furthermore, the study distinguishes between specific accumulation (November to April) and ablation (May to October) periods, aiming to capture distinct stages in glacier processes.

Moreover, the dynamics of glacier accumulation and ablation are significantly influenced by topographical features. For example, Zhang et al. [24] underscored vital statistical links between topographical variables and GMB data in the HMA region, particularly highlighting the significant role of median elevation and slope. Meanwhile, the study of Brun et al. [45] reveals that the response of glaciers to climate change is intricately linked to specific environmental conditions, including factors like slope gradients at the glacier's terminus, median altitude, the extent of debris coverage, and avalanche-prone zones. Motivated by these findings, median altitude, maximum glacier length, and slope, obtained from the RGI [30], were used in this study. Moreover, additional topographical factors, including longitude, latitude, minimum and maximum altitudes, aspect, and area of glaciers, were also considered predictive factors for GMB simulation. As outlined in Table 2, an array of 60 predictors were employed with each element corresponding to an individual GMB predictor.

Table 2. Predictor variables for driving ML models.

Topographical Variables	Ablation-Related Variables		Accumulation-Related Variables
Longitude	Annual CPT	Monthly accumulation temperature	Annual snowfall
Latitude	Monthly ablation CPT	Spring temperature	Monthly ablation snowfall
Median altitude	Monthly accumulation CPT	Summer temperature	Monthly accumulation snowfall
Min altitude	Spring CPT	Autumn temperature	Spring snowfall
Max altitude	Summer CPT	Winter temperature	Summer snowfall
Slope	Autumn CPT	1~12 temperature	Autumn snowfall
Aspect	4~10 CPT		Winter snowfall
Max length	Annual temperature		1~12 snowfall
Area	Monthly ablation temperature		

Note: 4~10 CPT denotes monthly CPT values from April to October; similarly, 1~12 temperature denotes monthly mean temperatures from January to December.

Based on the trained GBDT model, Figure 3 illustrates the thirty most impactful predictor variables in GMB simulations for the MRB and NRB, respectively. The graph clearly indicates that the majority of these influential predictors are closely associated with glacial topography. Specifically, in the MRB, nine of the top ten factors influencing GMB data are related to glacial topography, while in the NRB, this is true for eight of the top ten factors. Figure 3 also reveals that the median glacier altitude is the predominant factor affecting GMB data in the MRB, corroborating the findings of Zhang et al. [24]. This significant influence is likely due to the strong connection between median glacier elevation

and equilibrium-line altitude in the Tianshan region, as emphasized by Rabatel [46] and Braithwaite and Raper [47]. Moreover, latitude stands out as the second most influential variable impacting GMB modeling in the MRB. This phenomenon can be ascribed to the basin's north–south orientation, where glaciers located at high latitudes with low altitudes undergo more pronounced melting. In contrast, longitude is identified as the key predictor for GMB modeling in the NRB. This is likely due to the east–west orientation of the NRB, coupled with a decrease in elevation from west to east, resulting in elevated glacial melting at higher longitudes.

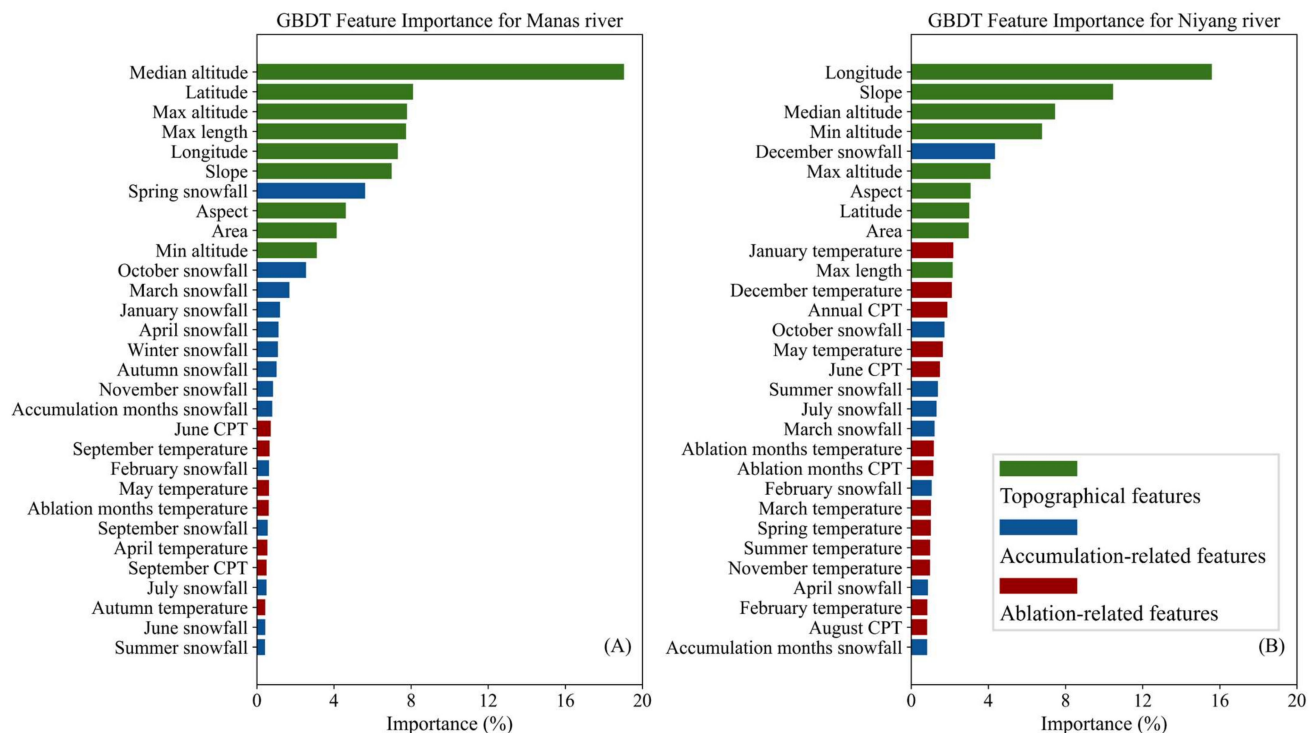


Figure 3. Contribution to the total variance in the top 30 topo-climatic predictors out of 60 predictors using the GBDT model for (A) the MRB and (B) the NRB. Green bars indicate predictors (including topographical features), blue ones include accumulation-related features, and red ones include ablation-related features.

Figure 3A shows that, besides topographical features, the most significant factor affecting GMB data in the MRB is snowfall which belongs to glacier accumulation-related features. This implies that accumulation-related features are more important than ablation-related features in GMB modeling, consistent with the findings of Bolibar [20] for Alpine glaciers classified as continental glaciers. Interestingly, meteorological conditions during transition months, like snowfall in Spring and October, can significantly affect annual GMB data in the MRB. This is primarily because glaciers in the MRB are winter-type glaciers, implying that accumulation mainly occurs in the winter half-year while melting predominates in summer. With lower temperatures during the winter half-year, snowfall is less sensitive to temperature changes. Consequently, the accumulation of snow directly influences glacier mass changes. As snow accumulation increases, glacier ablation decreases; conversely, reduced snow accumulation leads to increased glacier ablation. On the contrary, GMB data in the NRB are more significantly influenced by ablation-related features than accumulation-related features, consistent with recent studies demonstrating a strong link between accelerated glacier mass loss and regional warming in the Himalayan area [48]. Liang et al. [49] also found that GMB modelling was more responsive to changes in air temperature than changes in precipitation for maritime glaciers. This is primarily because summer snowfall could turn into rain with warming, reducing fresh snow accumulation

and causing the glacier surface albedo to decrease, thereby increasing glacier ablation [27]. Conversely, increased snowfall results in reduced glacier ablation.

4.2. Overall Performance of the Four ML Models

Due to the limited temporal coverage of geodetic mass balance observation data provided by Hugonnet et al. [18], spanning only two decades, utilizing this dataset for the ML-based temporal (i.e., two-dimensional outputs) simulation of GMB data poses significant challenges. To overcome this constraint and ensure an adequate sample size, this study adopts a conventional approach [20] that integrates both spatial and temporal dimensions of GMB samples, constructing a spatiotemporal predictive model. Consequently, the dataset comprises 4700 GMB samples from the MRB and 7260 from the NRB, all of which are employed for training and testing ML models.

Figure 4 compares the effectiveness of four ML models in simulating GMB data for the two river basins during the testing period. Results clearly show that the GBDT model outperforms the other three ML models in both of the two basins, achieving R^2 values of 0.72 and 0.67 and RMSE values of 0.21 m w.e. and 0.30 m w.e., respectively. With an R^2 value of 0.70 and an RMSE value of 0.22 m w.e., the RF model emerges as the second best performer in the MRB, closely resembling the performance of the GBDT model. A similar phenomenon is observed in the NRB. This consistency underscores the effectiveness of ensemble learning algorithms. This finding is also supported by Anilkumar et al. [21]. Moreover, the figures also highlight the relatively poor performance of the OLS model, likely due to its limitations in addressing the complex nonlinear relationships inherent in GMB modelling and its predictive factors. While DNNs outperform OLS, they still lag behind RF and GBDT models. This may be attributed to the fact that the DNN requires more parameters to be calibrated, which inherently demands a more comprehensive dataset for effective training.

Furthermore, Figure 4 reveals a consistent tendency that all algorithms in the two basins tend to underestimate high values and overestimate low values. This phenomenon aligns with the findings of Ren et al. [10]. Discrepancies between observed and simulated peaks may be attributed to the following factors: (1) the spatial resolution of ERA5-Land is inadequate for capturing extreme events specific to glaciers; (2) the influence of glacial lakes on glacial melting is not considered but can influence the accuracy of predictions; (3) particular conditions, such as supraglacial debris and glacial movement, are not considered in the models; and (4) inherent challenges of ML models in predicting extreme values due to the limited number of samples. Typically, extreme values are situated in the tail of the data distribution, whereas the training data often fail to adequately represent these extremes. This leads to a scarcity of sufficient examples of extreme values used for accurate prediction in such scenarios.

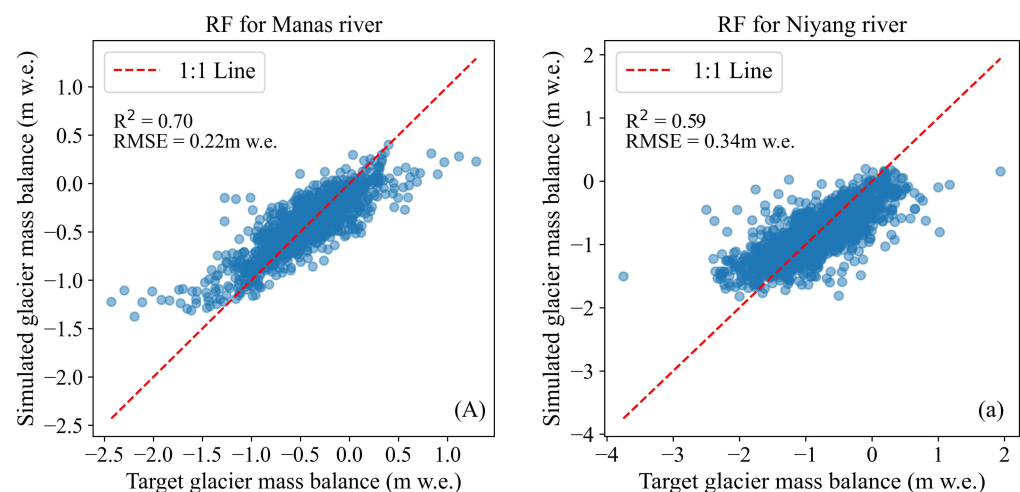


Figure 4. Cont.

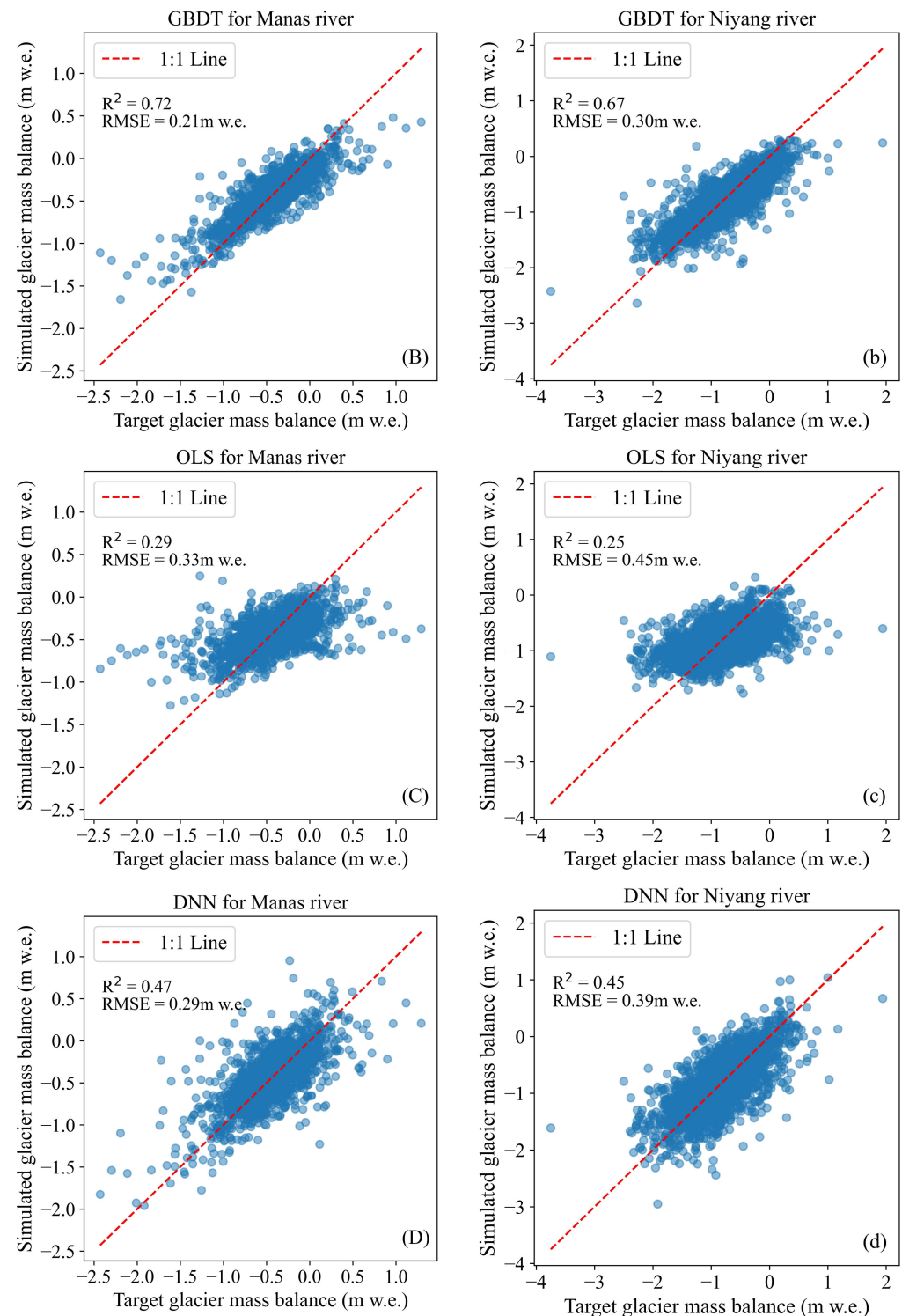


Figure 4. Evaluation of simulated GMB data against target GMB data in the testing period. (A–D) indicate glaciers in the MRB and (a–d) indicate glaciers in the NRB.

A comparison of ML performance between the two basins indicates superior performance in the MRB. This may be attributed to the predominance of continental glaciers in the MRB which are influenced by westerly winds characterized by lower moisture and precipitation. As shown in Figure 3A, the impact of snowfall on the GMB is significantly higher than that of temperature. Conversely, the glaciers in the NRB are maritime glaciers, primarily influenced by the South Asian monsoon carrying plenty of moisture. Glaciers in the NRB are characterized as maritime glaciers and are considered summer-type glaciers,

with the most notable features being accumulation and ablation, occurring predominantly in the summer months. Fujita [27] found that summer snowfall may transition into rain with warming, leading to the reduced accumulation of fresh snow and a decrease in glacier surface albedo, thereby increasing glacier ablation. Consequently, these glaciers are considered to be particularly sensitive to temperature. The second reason why glaciers in the NRB are particularly sensitive to warming is that they mainly accumulate during the summer, resulting in higher temperatures within the ice compared to continental glaciers, which brings them closer to the melting point. Additionally, recent studies have indicated a decreasing trend in precipitation in the NRB [28]. In the context of global warming, the NRB has experienced a more rapid increase in glacier surface temperature compared to the MRB [50]. Because glaciers in the NRB are more sensitive to climate change compared to continental glaciers, using MLs to simulate GMB data amplifies the uncertainty of ERA5-Land temperature data. Additionally, the lack of consideration for an important variable, such as ice temperature or glacier surface temperature, introduces new uncertainties to GMB simulations in this study. Therefore, the performance of MLs in simulating GMB in the NRB is not as advanced as in the MRB.

4.3. Basin Analysis

Numerous research studies on glacier simulation have extensively utilized process-based watershed hydrological models, with degree-day and/or energy balance modules serving as core components in these models, particularly in relation to GMB data. However, the applications of ML models in this domain, especially for simulating GMB data in the basin scale, are limited. The primary reason for this scarcity is the lack of sample data for GMB modelling. Therefore, the performance of ML models in the basin scale is assessed to enable a meaningful comparison with results derived from physical models.

Usually, when evaluating the simulation accuracy of the process-based physical model for GMB data in a basin, the corresponding evaluation index is calculated using the basin average of the GMB data. Figure 5 presents the basin-averaged GMB simulation results driven by spatiotemporal glacier samples over the whole study period following the use of this method. When benchmarked against remote sensing observation data from Hugonnet et al. [18], both RF and GBDT models were found to achieve amazing results in the MRB, with an R^2 value of 0.99 and an $RMSE$ value of 0.01 m w.e. This underscores the robust ability of RF and GBDT models in accurately simulating basin-scale GMB data. In contrast, OLS and DNN models were found to exhibit inferior performance, with R^2 values of 0.99 and 0.94 and $RMSE$ values of 0.01 m w.e. and 0.03 m w.e., respectively. Ren et al. [10] simulated basin GMB data in the MRB using the physically based SPHY model, attaining a Nash–Sutcliffe efficiency (NSE) coefficient of -3.26 and a correlation coefficient (CC) of 0.64. Even so, the results were considered satisfactory as they could generally capture the trend of basin-scale glacier changes. In fact, previous studies have shown that a limited number of physical-based GMB simulations can reproduce the signal of glacial changes, even at the basin scale. As a result, all four of the ML models hold significant potential in basin glacier simulation.

In the NRB, the GBDT model was also found to exhibit excellent performance, with a R^2 value of 0.99 and a $RMSE$ value of 0.02 m w.e. In contrast, both OLS and DNN models presented a slightly lower R^2 value of 0.88 and higher $RMSE$ values of 0.06 m w.e. and 0.04 m w.e., respectively. Surprisingly, the RF model showed the worst performance, with a R^2 value of 0.86 and a $RMSE$ value of 0.06 m w.e. Using the SPHY model, He et al.'s [51] simulation of GMB data in the NRB from 2006 to 2010 achieved a R^2 value of 0.37. This result falls significantly short of the achievements of all four ML models. This result also shows that all the four ML models, especially GBDTs, hold significant potential in average glacial basin simulation.

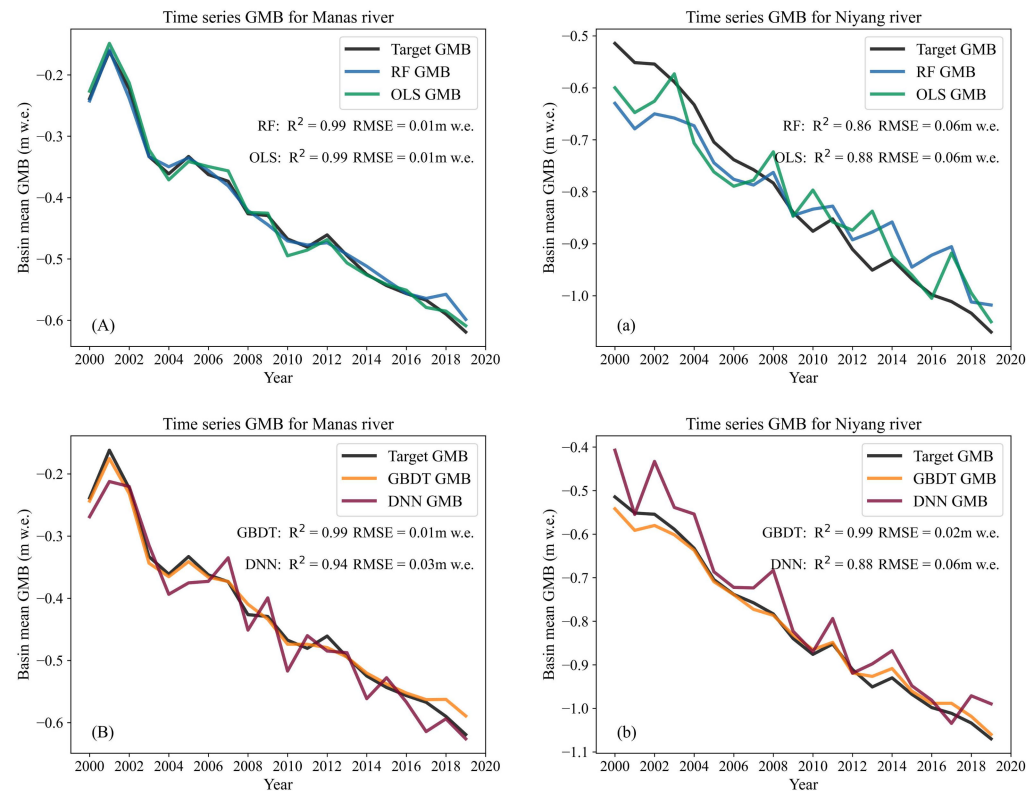


Figure 5. A comparison of basin-averaged GMB simulation from the four ML models against geodetic mass balance observations from Hugonnet et al. [18]. (A,B) indicate the MRB and (a,b) indicate the NRB.

5. Discussion

5.1. The Influence of Outliers

Outliers, indicative of a small and unrepresentative sample size, can adversely affect the performance of ML models. Firstly, they compromise the robustness of the model by guiding it towards unreasonable learning directions, making it excessively sensitive to training data. Consequently, the model may attempt to fit outliers instead of capturing the true underlying patterns in the data, rendering it overly responsive to even minor variations in the input data. Moreover, the existence of outliers leads to diminished model generalization. Overfitting to outliers during training reduces the model's ability to generalize new data, resulting in poor test set performance due to excessive reliance on outliers during training.

To mitigate the impact of outliers on ML models, various outlier removal methods have been proposed, with the most commonly used approaches being standard deviation and boxplot methods. In this study, the standard deviation method was employed to remove outliers, using the NRB as an example to illustrate the necessity of outlier removal. The GMB sample values larger than $+3\sigma$ and less than -3σ were removed. The probability density map before and after outlier removal is shown in Figure 6.

Figure 6A illustrates the results before removing the outliers, while Figure 6B displays the results after outlier removal. Upon careful examination of the figures, it is apparent that the NRB includes samples with glacier mass loss exceeding -3σ , resulting in a deviation from a normal distribution of GMB samples. An examination based on the RGI revealed that these samples are primarily relatively small but significantly affected by global warming, leading to a notably accelerated melting rate. Though there are many similar small glaciers, the limitations imposed by the spatial resolution of meteorological data necessitate the selection of glaciers with larger areas. Consequently, this particular glacier emerges as an outlier within the samples.

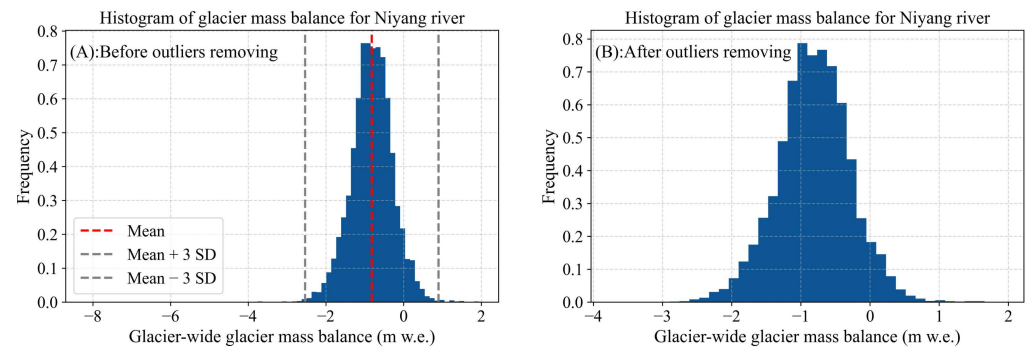


Figure 6. Comparison of the sample histograms: (A) before outliers were removed and (B) after outliers were removed for this study in the NRB.

It should be noted that the samples after the outliers' removal still have relatively large or small values, as shown in Figure 4. This is because there is no clear boundary between outliers and normal values, and outliers can also be real values caused by rare phenomena. Therefore, the ML models with samples treated with previously removed outliers will also encounter phenomena based on the overestimation of low values and the underestimation of high values, as shown in Figure 4.

To further illustrate the necessity of excluding outliers, this study provides ML simulation results for the NRB without outlier removal, as depicted in Figure 7. A comparison between the glacier simulation results for the NRB in Figures 4 and 7 revealed varying degrees of decline in the modeling capabilities of the RF, GBDT, OLS, and DNN algorithms regarding GMB data. Specifically, the R^2 exhibited reductions of 5.08%, 8.95%, 4.00%, and 8.89%, respectively. Notably, the GBDT and DNN models showed the most significant reductions, while OLS was the least affected. This finding suggests that not excluding outliers has a more pronounced impact on nonlinear algorithms. In contrast, linear models are generally simpler and exhibit lower levels of data fitting, making them somewhat more resilient to the influence of outliers. Nonlinear algorithms, on the other hand, tend to be more flexible and complex, with a greater number of parameters to adjust, making them more susceptible to the influence of extreme values in the data. Therefore, it is suggested that outliers must be eliminated in nonlinear ML simulations.

5.2. The Potential of Temporal Predictive Models

The time series prediction of GMB data is of great significance for assessing the temporal variations in watershed glaciers. However, constructing a stable ML-based model with two-dimensional outputs using only 20 years of remote sensing GMB samples proved to be challenging. The results of the model constructed by combining the spatiotemporal samples of glaciers in a ratio of 7:3 for training and testing samples are presented in Figure 5. Due to the random partitioning of training and testing samples, direct simulation results for the GMB time series during the testing period were difficult to obtain. To illustrate the modeling capability of ML models for the GMB time series, this study, based on the GBDT model, explores the simulation results for different training and testing sample ratios. The time series simulation results and evaluation metrics are presented in Figure 8 and Table 3.

From Figure 8, it can be observed that in the MRB, when the training sample proportion is 30%, the overall time series simulation results are significantly better than those when the training sample proportion is 20%. Subsequently, as the proportion of training samples increases, the overall simulation results show little variation. In contrast, for the NRB, the overall simulation results only begin to stabilize when the sample proportion is 50%. This suggests that the glacier changes in the MRB are relatively small, with a more evenly distributed sample, and 30% of the samples already have overall representativeness. On the other hand, glaciers in the NRB undergo drastic changes, with a more scattered sample distribution, and a 50% sample proportion is needed for some representativeness.

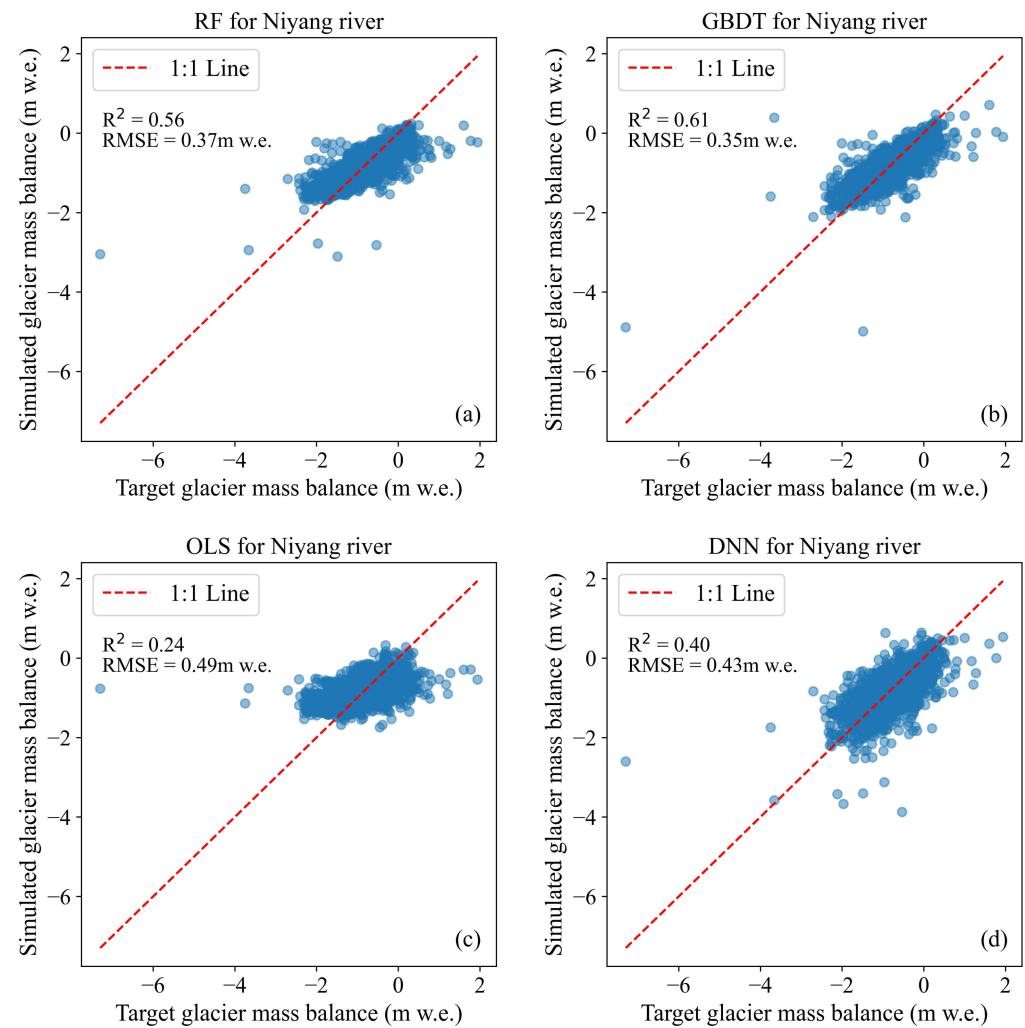


Figure 7. An evaluation of simulated GMB data against target GMB data in the testing period for glaciers in the NRB. (a–d) Scatter plot of the RF, GBDT, OLS and DNN model results, respectively.

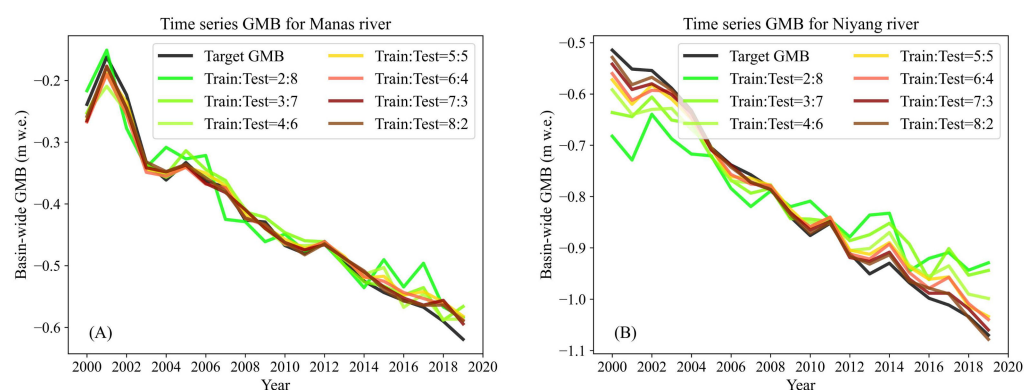


Figure 8. The simulation of basin-scale GMB data under different proportions of train samples and test samples for (A) the MRB and (B) the NRB based on the GBDT model.

As evident in Table 3, for the Manas River training data with a sample proportion of 30%, the R^2 and RMSE values were 0.97 and 0.02 m w.e., respectively. This result is significantly better than the watershed GMB data simulated by Ren et al. [10] in the MRB based on the SPHY model. Even for the NRB, which performs slightly worse, the R^2 and RMSE values were 0.96 and 0.03 m w.e., respectively, when the training sample proportion was 50%. This result is noticeably better than the results recorded by He et al. [51], who

simulated NRB GMB data based on the SPHY model for the 2006–2010 period. This suggests that the size of the training sample proportion has a certain impact on the simulation of the GMB time series, but as the proportion increases to a certain extent, this impact becomes smaller and can even be negligible. This also indicates that ML algorithms in the simulation of the GMB time series in watershed areas are significantly superior to physical models.

Table 3. Evaluation metrics of basin-scale GMB simulation under different proportions of train and test samples based on the GBDT model.

Train Sample/Test Sample	MRB		NRB	
	R^2	RMSE (m w.e.)	R^2	RMSE (m w.e.)
2:8	0.93	0.03	0.72	0.09
3:7	0.97	0.02	0.85	0.07
4:6	0.98	0.02	0.92	0.05
5:5	0.98	0.02	0.96	0.03
6:4	0.98	0.02	0.97	0.03
7:3	0.99	0.01	0.99	0.02
8:2	0.99	0.01	0.99	0.01

5.3. Limitations of the Current Study

The current meteorological variables used in ML for GMB simulation are primarily derived from traditional simple temperature index-based models, relying solely on temperature and precipitation. However, there is untapped potential to improve these models' predictive capabilities by incorporating variables from more sophisticated models, such as energy balance models. These models include longwave and shortwave radiation, turbulent fluxes, and albedo levels, providing a more comprehensive explanation of the underlying processes. Research conducted by Xiao et al. [52] has shown a decline in precipitation levels, resulting in reduced glacier albedo levels in the eastern Himalayas. Additionally, studies on glaciers in southeastern Tibet suggest that while black carbon contributes less than 5% to the melting volume of fresh snow, its concentration increases during the melting process, potentially raising its contribution to up to 15% [53]. Therefore, albedo or radiation should be considered in GMB modeling. Recent research by Anilkumar et al. [21] also confirms the importance of net solar radiation and albedo as critical factors influencing glacier ablation across five ML models.

Furthermore, glaciers with higher ice temperatures are more sensitive to climate change compared to those with lower ice temperatures. Maritime glaciers exhibit an amplified response to warming climates, primarily due to their higher ice temperatures nearing the melting point. Consequently, relying solely on air temperature and precipitation data may insufficiently capture the mass balance dynamics of maritime glaciers. Another aspect not considered in this study is the impact of debris cover on the GMB model. Research by Zhang et al. [24] has indicated that debris cover significantly contributes to the spatial heterogeneity of GMB data. Therefore, it is essential to incorporate considerations of radiation or albedo, ice temperature, and debris cover when modeling these glaciers. Expanding the ML model to include these new inputs may improve the precision of GMB simulation. However, an increased consideration of predictors affecting GMB data must account for different glacier types, as the above results reveal notable differences in ML model performance for GMB simulation in the MRB and NRB.

Another primary concern is the uncertainty inherent in meteorological predictor variables, which might impair the models' ability to generalize. In this research, topographical factors have emerged as more influential than meteorological predictors in glacier simulations across both studied basins. Glaciers, as manifestations of climatic conditions, are fundamentally affected by temperature and precipitation during their accumulation and ablation periods. However, terrain elements also exert considerable control over these processes. The topographical variables employed in this study were sourced from high-resolution satellite remote sensing data, surpassing the glacier area threshold of 0.5 km²

set for this study. In contrast, the ERA5-Land meteorological dataset offers a spatial resolution of $0.1^\circ \times 0.1^\circ$, which is significantly coarser than the glacier areas analyzed here. Consequently, these findings deviate from those of Bolibar et al. [20], who observed the predominant influence of meteorological data in GMB simulations. Additionally, Wu et al. [35] noted a positive bias in precipitation measurements in ERA5-Land, although it accurately captures intra-annual variations. Similarly, Zhao and He [54] found that ERA5-Land reliably tracks temperature trends but often misrepresents the magnitude of these values. Therefore, future studies should incorporate the spatial downscaling of meteorological data to improve its accuracy and applicability.

6. Conclusions

The accelerating retreat of global glaciers, driven by climate change, presents significant risks. The accurate simulation of GMB data is vital for evaluating these impacts. While traditional models encounter challenges and uncertainties, ML holds promise as a robust alternative. To enhance the precision of GMB simulation across diverse glacier types in the HMA region, this study assessed four ML models for annual glacier-wide GMB simulation. It specifically compared maritime glaciers in the NRB with continental glaciers in the MRB. Meteorological predictors obtained from ERA5-Land monthly datasets, along with topographical predictors from the RGI 6.0, were utilized to drive the following four ML models: the RF model, the GBDT model, the DNN model, and the OLS model.

The key findings of this study are as follows. ML models generally show superior performance in modeling GMB data for continental glaciers, primarily attributed to varying influencing factors for different glacier types. The GBDT model consistently outperforms the other three models, supported by higher R^2 and lower $RMSE$ values in the two glacier types, and demonstrates significant potential in temporal GMB simulation compared to temperature index models. The OLS model, constrained by its linear property, demonstrates limited effectiveness in capturing complex relationships inherent in GMB simulation. The influential factors for GMB simulation vary between continental and maritime glaciers. For continental glaciers, topographical and accumulation-related factors are the most significant factors affecting GMB simulation, whereas for maritime glaciers, topographical and ablation-related predictors are more influential.

The study reveals the robust potential of ML in modeling GMB data across diverse glacier types, presenting a valuable approach for investigating the impacts of climate change on glacier dynamics in the HMA region. However, to improve the accuracy of GMB modeling, several considerations are imperative. For example, incorporating additional predictors like debris cover, glacier surface temperatures, and surface albedo could provide a more comprehensive explanation of glacier behavior. Moreover, improving the representativeness of meteorological data, potentially through spatial downscaling, can reduce the uncertainty associated with meteorological predictive factors. These measures potentially contribute to the performance improvement of glacier simulation.

Author Contributions: All authors contributed to the work. Conceptualization, X.L. and D.Z.; methodology, Y.W., Z.Z. and W.R.; software, J.S., Z.Z. and W.R.; validation, J.S., R.Z. and W.R.; formal analysis, W.R.; investigation, Z.Z. and W.R.; resources, W.R.; data curation, W.R.; writing—original draft preparation, W.R.; writing—review and editing, X.L., D.Z. and R.Z.; visualization, W.R.; supervision, X.L. and D.Z.; project administration, X.L. and D.Z.; funding acquisition, W.R. and J.S. 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 (grant numbers 42101406 and 42101397).

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors would like to thank the editors and the reviewers for their crucial comments and suggestions which improved the quality of this paper.

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

References

1. Biemans, H.; Siderius, C.; Lutz, A.F.; Nepal, S.; Ahmad, B.; Hassan, T. Importance of snow and glacier meltwater for agriculture on the Indo-Gangetic Plain. *Nat. Sustain.* **2019**, *2*, 594–601. [\[CrossRef\]](#)
2. Immerzeel, W.W.; Lutz, A.F.; Andrade, M.; Bahl, A.; Biemans, H.; Bolch, T.; Hyde, S.; Brumby, S.; Davies, B.J.; Elmore, A.C.; et al. Importance and vulnerability of the world's water towers. *Nature* **2020**, *577*, 364–369. [\[CrossRef\]](#)
3. Miles, E.; McCarthy, M.; Dehecq, A.; Kneib, M.; Fugger, S.; Pellicciotti, F. Health and sustainability of glaciers in High Mountain Asia. *Nat. Commun.* **2021**, *12*, 2868. [\[CrossRef\]](#)
4. Marzeion, B.; Hock, R.; Anderson, B.; Bliss, A.; Champollion, N.; Fujita, K.; Huss, M.; Immerzeel, W.W.; Kraaijenbrink, P.; Malles, J.H.; et al. Partitioning the uncertainty of ensemble projections of global glacier mass change. *Earth's Future* **2020**, *8*, e2019EF001470. [\[CrossRef\]](#)
5. 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. [\[CrossRef\]](#)
6. Huss, M.; Hock, R. Global-scale hydrological response to future glacier mass loss. *Nat. Clim. Chang.* **2018**, *8*, 135–140. [\[CrossRef\]](#)
7. Li, X.; Cheng, G.; Fu, B.; Xia, J.; Zhang, L.; Yang, D.; Zheng, C.; Liu, S.; Li, X.; Song, C.; et al. Linking critical zone with watershed science: The example of the Heihe River basin. *Earth's Future* **2022**, *10*, e2022EF002966. [\[CrossRef\]](#)
8. Li, X.; Feng, M.; Ran, Y.; Su, Y.; Liu, F.; Huang, C.; Shen, H.; Xiao, Q.; Su, J.; Yuan, S.; et al. Big Data in Earth system science and progress towards a digital twin. *Nat. Rev. Earth Environ.* **2023**, *4*, 319–332. [\[CrossRef\]](#)
9. Zeng, R.; Ren, W. The spatiotemporal trajectory of US agricultural irrigation withdrawal during 1981–2015. *Environ. Res. Lett.* **2022**, *17*, 104027. [\[CrossRef\]](#)
10. Ren, W.; Li, X.; Zheng, D.; Zeng, R.; Su, J.; Mu, T.; Wang, Y. Enhancing Flood Simulation in Data-Limited Glacial River Basins through Hybrid Modeling and Multi-Source Remote Sensing Data. *Remote Sens.* **2023**, *15*, 4527. [\[CrossRef\]](#)
11. Liu, H.; Wang, L.; Zhou, J.; Shrestha, M.; Chai, C.; Li, X.; Ahmad, B. Energy-balance modeling of heterogeneous glacio-hydrological regimes at upper Indus. *J. Hydrol. Reg. Stud.* **2023**, *49*, 101515. [\[CrossRef\]](#)
12. Ismail, M.F.; Bogacki, W.; Disse, M.; Schäfer, M.; Kirschbaur, L. Estimating degree-day factors of snow based on energy flux components. *Cryosphere* **2023**, *17*, 211–231. [\[CrossRef\]](#)
13. Ren, W.; Yang, T.; Shi, P.; Xu, C.Y.; Zhang, K.; Zhou, X.; Shao, Q.; Ciais, P. A probabilistic method for streamflow projection and associated uncertainty analysis in a data sparse alpine region. *Glob. Planet. Chang.* **2018**, *165*, 100–113. [\[CrossRef\]](#)
14. Lutz, A.F.; Immerzeel, W.W.; Shrestha, A.B.; Bierkens, M.F.P. Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. *Nat. Clim. Chang.* **2014**, *4*, 587–592. [\[CrossRef\]](#)
15. Khanal, S.; Lutz, A.F.; Kraaijenbrink, P.D.A.; Hurk, B.v.D.; Yao, T.; Immerzeel, W.W. Variable 21st century climate change response for rivers in High Mountain Asia at seasonal to decadal time scales. *Water Resour. Res.* **2021**, *57*, e2020WR029266. [\[CrossRef\]](#)
16. Gao, R.; Torres-Rua, A.F.; Aboutaleb, M.; White, W.A.; Anderson, M.; Kustas, W.P.; Agam, N.; Alsina, M.M.; Alfieri, J.; Hipps, L.; et al. LAI estimation across California vineyards using sUAS multi-seasonal multi-spectral, thermal, and elevation information and machine learning. *Irrig. Sci.* **2022**, *40*, 731–759. [\[CrossRef\]](#)
17. Ren, W.W.; Yang, T.; Huang, C.S.; Xu, C.Y.; Shao, Q.X. Improving monthly streamflow prediction in alpine regions: Integrating HBV model with Bayesian neural network. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 3381–3396. [\[CrossRef\]](#)
18. 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*, 726–731. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Hoinkes, H.C. Glacier variation and weather. *J. Glaciol.* **1968**, *7*, 3–18. [\[CrossRef\]](#)
20. Bolibar, J.; Rabatel, A.; Gouttevin, I.; Galiez, C.; Condom, T.; Sauquet, E. Deep learning applied to glacier evolution modelling. *Cryosphere* **2020**, *14*, 565–584. [\[CrossRef\]](#)
21. Anilkumar, R.; Bharti, R.; Chutia, D.; Aggarwal, S.P. Modelling point mass balance for the glaciers of the Central European Alps using machine learning techniques. *Cryosphere* **2023**, *17*, 2811–2828. [\[CrossRef\]](#)
22. Brun, F.; Berthier, E.; Wagnon, P.; Kääb, A.; Treichler, D. A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to 2016. *Nat. Geosci.* **2017**, *10*, 668–673. [\[CrossRef\]](#)
23. Zhao, F.; Long, D.; Li, X.; Huang, Q.; Han, P. Rapid glacier mass loss in the Southeastern Tibetan Plateau since the year 2000 from satellite observations. *Remote Sens. Environ.* **2022**, *270*, 112853. [\[CrossRef\]](#)
24. Zhang, Z.; Gu, Z.; Hu, K.; Xu, Y.; Zhao, J. Spatial variability between glacier mass balance and environmental factors in the High Mountain Asia. *J. Arid. Land* **2022**, *14*, 441–454. [\[CrossRef\]](#)
25. Kraaijenbrink, P.D.; Bierkens, M.F.; Lutz, A.F.; Immerzeel, W.W. Impact of a global temperature rise of 1.5 degrees Celsius on Asia's glaciers. *Nature* **2017**, *549*, 257–260. [\[CrossRef\]](#)
26. Yao, T.; Thompson, L.; Yang, W.; Yu, W.; Gao, Y.; Guo, X.; Yang, X.; Duan, K.; Zhao, H.; Xu, B.; et al. Different glacier status with atmospheric circulations in Tibetan Plateau and surroundings. *Nat. Clim. Chang.* **2012**, *2*, 663–667. [\[CrossRef\]](#)
27. Fujita, K. Influence of precipitation seasonality on glacier mass balance and its sensitivity to climate change. *Ann. Glaciol.* **2008**, *48*, 88–92. [\[CrossRef\]](#)
28. Yao, T.D.; Yu, W.S.; Wu, G.J.; Xu, B.; Yang, W.; Zhao, H.; Wang, W.; Li, S.; Wang, N.; Li, Z.; et al. Glacier anomalies and relevant disaster risks on the Tibetan Plateau and surroundings. *Chin. Sci. Bull.* **2019**, *64*, 2770–2782. (In Chinese)

29. Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. [CrossRef]
30. RGI Consortium. GLIMS: Global Land Ice Measurements from Space, A Dataset of Global Glacier Outlines Version 6.0 Technical Report, Colorado, USA. 2017. Available online: <https://www.glims.org/RGI/randolph60.html> (accessed on 10 January 2024).
31. Zhang, Z.; Li, Z.; He, X. Progress in the research on glacial change and water resources in Manas river basin. *Int. Soil Water Conserv. Res.* **2014**, *25*, 332–337. (In Chinese)
32. Wang, X.; Yang, T.; Xu, C.Y.; Yong, B.; Shi, P. Understanding the discharge regime of a glacierized alpine catchment in the Tianshan Mountains using an improved HBV-D hydrological model. *Glob. Planet. Chang.* **2019**, *172*, 211–222. [CrossRef]
33. Ji, X.; Chen, Y. Characterizing spatial patterns of precipitation based on corrected TRMM 3B43 data over the mid Tianshan Mountains of China. *J. Mt. Sci.* **2012**, *9*, 628–645. [CrossRef]
34. Zhang, M.; Ren, Q.; Wei, X.; Wang, J.; Yang, X.; Jiang, Z. Climate change, glacier melting and streamflow in the Niyang River Basin, Southeast Tibet, China. *Ecohydrology* **2011**, *4*, 288–298. [CrossRef]
35. Wu, X.; Su, J.; Ren, W.; Lü, H.; Yuan, F. Statistical comparison and hydrological utility evaluation of ERA5-Land and IMERG precipitation products on the Tibetan Plateau. *J. Hydrol.* **2023**, *620*, 129384. [CrossRef]
36. Yilmaz, M. Accuracy assessment of temperature trends from ERA5 and ERA5-Land. *Sci. Total Environ.* **2023**, *856*, 159182. [CrossRef] [PubMed]
37. Zhou, L.; Pan, S.; Wang, J.; Vasilakos, A.V. Machine learning on big data: Opportunities and challenges. *Neurocomputing* **2017**, *237*, 350–361. [CrossRef]
38. Schoppa, L.; Disse, M.; Bachmair, S. Evaluating the performance of random forest for large-scale flood discharge simulation. *J. Hydrol.* **2020**, *590*, 125531. [CrossRef]
39. Rahman, A.; Hosono, T.; Quilty, J.M.; Das, J.; Basak, A. Multiscale groundwater level forecasting: Coupling new machine learning approaches with wavelet transforms. *Adv. Water Resour.* **2020**, *141*, 103595. [CrossRef]
40. Seydi, S.T.; Kanani-Sadat, Y.; Hasanlou, M.; Sahraei, R.; Chanussot, J.; Amani, M. Comparison of machine learning algorithms for flood susceptibility mapping. *Remote Sens.* **2022**, *15*, 192. [CrossRef]
41. He, X.; Luo, J.; Zuo, G.; Xie, J. Daily runoff forecasting using a hybrid model based on variational mode decomposition and deep neural networks. *Water Resour. Manag.* **2019**, *33*, 1571–1590. [CrossRef]
42. Kadow, C.; Hall, D.M.; Ulbrich, U. Artificial intelligence reconstructs missing climate information. *Nat. Geosci.* **2020**, *13*, 408–413. [CrossRef]
43. Bolibar, J.; Rabatel, A.; Gouttevin, I.; Zekollari, H.; Galiez, C. Nonlinear sensitivity of glacier mass balance to future climate change unveiled by deep learning. *Nat. Commun.* **2022**, *13*, 409. [CrossRef]
44. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
45. Brun, F.; Wagnon, P.; Berthier, E.; Jomelli, V.; Maharjan, S.B.; Shrestha, F.; Kraaijenbrink, P.D.A. Heterogeneous influence of glacier morphology on the mass balance variability in High Mountain Asia. *J. Geophys. Res. Earth Surf.* **2019**, *124*, 1331–1345. [CrossRef]
46. Rabatel, A.; Dedieu, J.P.; Vincent, C. Spatio-temporal changes in glacier-wide mass balance quantified by optical remote sensing on 30 glaciers in the French Alps for the period 1983–2014. *J. Glaciol.* **2016**, *62*, 1153–1166. [CrossRef]
47. Braithwaite, R.J.; Raper, S.C.B. Estimating equilibrium-line altitude (ELA) from glacier inventory data. *Ann. Glaciol.* **2009**, *50*, 127–132. [CrossRef]
48. Maurer, J.M.; Schaefer, J.M.; Rupper, S.; Corley, A.J.S.A. Acceleration of ice loss across the Himalayas over the past 40 years. *Sci. Adv.* **2019**, *5*, eaav7266. [CrossRef]
49. Liang, L.; Cuo, L.; Liu, Q. Mass balance variation and associative climate drivers for the Dongkemadi Glacier in the central Tibetan Plateau. *J. Geophys. Res. Atmos.* **2019**, *124*, 10814–10825. [CrossRef]
50. Rani, S.; Mal, S. Trends in land surface temperature and its drivers over the High Mountain Asia. *Egypt. J. Remote Sens. Space Sci.* **2022**, *25*, 717–729. [CrossRef]
51. He, Q.; Kuang, X.; Chen, J.; Feng, Y.; Zheng, C. Reconstructing runoff components and glacier mass balance with climate change: Niyang River basin, southeastern Tibetan Plateau. *Front. Earth Sci.* **2023**, *11*, 1165390. [CrossRef]
52. Xiao, Y.; Ke, C.Q.; Shen, X.; Cai, Y.; Li, H. What drives the decrease of glacier surface albedo in High Mountain Asia in the past two decades? *Sci. Total Environ.* **2023**, *863*, 160945. [CrossRef] [PubMed]
53. Zhang, Y.; Kang, S.; Cong, Z.; Schmale, J.; Sprenger, M.; Li, C.; Yang, W.; Gao, T.; Sillanpää, M.; Li, X.; et al. Light-absorbing impurities enhance glacier albedo reduction in the southeastern Tibetan Plateau. *J. Geophys. Res. Atmos.* **2017**, *122*, 6915–6933. [CrossRef]
54. Zhao, P.; He, Z. A first evaluation of ERA5-Land reanalysis temperature product over the Chinese Qilian Mountains. *Front. Earth Sci.* **2022**, *10*, 907730. [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.