Special Issue Reprint

Hyperspectral Remote Sensing from Spaceborne and Low Altitude Aerial/ Drone-Based Platforms

Differences in Approaches, Data Processing Methods, and Applications

Edited by
Amin Beiranvand Pour, Arindam Guha, Laura Crispini and Snehamoy Chatterjee

MDPI

Hyperspectral Remote Sensing from Spaceborne and Low Altitude Aerial/Drone-Based Platforms—Differences in Approaches, Data Processing Methods, and Applications

Hyperspectral Remote Sensing from Spaceborne and Low Altitude Aerial/Drone-Based Platforms—Differences in Approaches, Data Processing Methods, and Applications

Editors

Amin Beiranvand Pour
Arindam Guha
Laura Crispini
Snehamoy Chatterjee

Editors

Amin Beiranvand Pour
Institute of Oceanography
and Environment (INOS)
University Malaysia
Terengganu (UMT)
Kuala Nerus
Malaysia

Arindam Guha
Mineral Exploration and
Geoenvironment Division
(Geosciences Group)
National Remote Sensing
Centre, Indian Space Research
Organisation Balanagar
Hyderabad
India

Laura Crispini
Department of Earth,
Environment and Life
Sciences (DISTAV)
University of Genova
Genova
Italy

Snehamoy Chatterjee
Department of Geological
and Mining Engineering and
Sciences
Michigan Technological
University
Houghton
USA

This is a reprint of articles from the Special Issue published online in the open access journal *Remote Sensing* (ISSN 2072-4292) (available at: https://www.mdpi.com/journal/remotesensing/special_issues/hyperspectral_rs_platforms).

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. *Journal Name* **Year**, *Volume Number*, Page Range.

Contents

About the Editors

Amin Beiranvand Pour

Amin Beiranvand Pour is an experienced and internationally established researcher in the field of geological remote sensing and mineral exploration. He was listed among the World's Top 2% Scientists by Stanford University for the years 2020, 2021, and 2022. He is a highly cited researcher in the field of geological remote sensing in the world (ranked two on Google Scholar). He has ranked as one among Top Scientists for the year 2023 in earth science in Malaysia. He has a full academic background in applied geology, economic geology, remote sensing, and mineral exploration. He was project leader of numerous mineral exploration research projects using satellite remote sensing technology in arid and semi-arid terrains and Antarctic, Arctic, and tropical environments. His experience ranges widely, from mineral exploration to environmental issues such as geo-hazard, structural mapping, geothermal and geomorphic, and coastal geology investigations. He has published more than 180 research papers in the field of geological remote sensing, mineral exploration, and geohazard modeling.

Arindam Guha

Arindam Guha is currently working at the Mineral Exploration and Geoenvironment Division (Geosciences Group), National Remote Sensing Centre, Indian Space Research Organisation, Balanagar, Hyderabad, in Telengana in India. They are interested in airborne hyperspectral; spaceborne hyperspectral satellite; hyperspectral data processing; geoenvironmental; mineral prospectivity; and target detection.

Laura Crispini

Laura Crispini is a Structural Geologist and Full Professor at the University of Genova (Italy). Her areas of expertise include multiscalar structural analysis, field-based geological mapping, remote sensing applications to geological mapping, tectonic and geodynamic evolution of the oceanic lithosphere and orogenic belts, and the study of solid-earth-cryosphere interactions in Polar Regions in the past to recent times. She has taken part in various scientific expeditions to Antarctica, either as a Scientist, Project Leader, or Scientific Coordinator of an Italian Program. She has also been a part of core logging teams and was Team Leader of the International Ocean Discovery Program (IODP) and International Continental Drilling Program (ICDP) expeditions.

Snehamoy Chatterjee

Snehamoy Chatterjee is an Associate Professor and Witte Family Faculty Fellow of Mining Engineering at Michigan Technological University. His research delves into spatial statistics, operations research, artificial intelligence, and remote sensing, with applications in geosciences. Holding a Ph.D. in Mining Engineering from the Indian Institute of Technology Kharagpur, he worked as a Post-Doctoral Fellow at the University of Alaska Fairbanks and as a research associate at COSMO Stochastic Mine Planning Laboratory, McGill University, Canada. Dr. Chatterjee's research, funded by federal agencies and mining companies, has earned him recognition, including The Editor's Best Reviewer Awards 2014 from Mathematical Geosciences Journal and the 2015 Application of Computers and Operations Research in the Mineral Industry (APCOM) Young Professional Award. He actively contributes to professional organizations like IAMG, SME, GSA, and AGU, serving as a co-convener and technical committee member for international conferences. Dr. Chatterjee is also an editorial board member for the *International Journal of Mining, Reclamation, and*

Environment, and an associate editor for the *Mining, Metallurgy & Exploration* and *Results in Geophysical Sciences* journals.

Editorial for the Special Issue Entitled *Hyperspectral Remote Sensing from Spaceborne and Low-Altitude Aerial/Drone-Based Platforms*—Differences in Approaches, Data Processing Methods, and Applications

Amin Beiranvand Pour [1,*], **Arindam Guha** [2], **Laura Crispini** [3] and **Snehamoy Chatterjee** [4]

[1]   Institute of Oceanography and Environment (INOS), Universiti Malaysia Terengganu (UMT), Kuala Nerus 21030, Malaysia
[2]   Geosciences Group, National Remote Sensing Centre, Indian Space Research Organization, Balanagar, Hyderabad 500010, India; arindam_g@nrsc.gov.in
[3]   Department for Earth, Environment and Life Sciences, University of Genova, Corso Europa 26, 16132 Genova, Italy; laura.crispini@unige.it
[4]   Department of Geological and Mining Engineering and Science, Michigan Technological University, Houghton, MI 49931, USA; schatte1@mtu.edu
*   Correspondence: beiranvand.pour@umt.edu.my; Tel.: +609-6683824; Fax: +609-6692166

1. Introduction

Nowadays, several hyperspectral remote sensing sensors from spaceborne and low-altitude aerial/drone-based platforms with a variety of spectral and spatial resolutions are available for geoscientific applications [1–4]. Advances in hyperspectral remote sensing imagery have fostered the progress of novel image processing techniques, which have yielded auspicious outcomes in a wide range of fields, such as soil geochemistry, water quality assessments, forest species mapping, agricultural stress, mineral alteration mapping, etc. In the last two decades, multiple spaceborne hyperspectral sensors have been launched by different space agencies (e.g., Hyperion in November 2000 by the National Aeronautics and Space Administration (NASA), USA; Hyperspectral Imager Suite (HISUI) in December 2019 by the Japan Aerospace Exploration Agency (JAXA); and Precursore IperSpettrale della Missione Applicativa (PRISMA) in March 2019 by the Italian Space Agency (ASI)) [1,5,6]. These sensors have made significant use of hyperspectral data and also led to innovative approaches to data processing, from noise removal to spectral mapping. Previous studies have highlighted the limitations of hyperspectral spaceborne sensors in identifying a pure target and in identifying spectral targets with subdued spectral signatures, as these hyperspectral sensors have coarse spatial resolution (generally 20 m to 30 m) and poor signal-to-noise ratios (e.g., Hyperion has a poor signal-to-noise-ratio (SNR) in the shortwave electromagnetic domain) [7–10]. However, these spaceborne sensors have yielded encouraging results in environmental monitoring (for example, forest cover classification, the detection of phonological changes in forests, land use/land cover mapping, agriculture land cover characterization, crop stress estimation, mapping lithology and minerals [11–13]). Hyperspectral image processing addresses the main difficulties associated with classification methods, such as the high dimensionality of the associated data and the limited availability of standard processing techniques [14]. To confront these limitations, several machine learning algorithms have recently been established, supplementing the high potential of hyperspectral data processing [14].

Due to the lack of global coverage of spaceborne hyperspectral sensors, routine aircraft-based and drone-based hyperspectral surveys are carried out in different countries using different advanced hyperspectral sensors, such as the advanced visible infrared spectrometer (AVIRIS) and its latest version, AVIRIS-next generation (AVIRIS-NG); HyMap; the digital airborne imaging spectrometer (DAIS), etc. These sensors are capable of collecting

high-spatial- and spectral-resolution data with optimum spectral fidelity [15]. Unmanned aircraft systems (UAS) with low-altitude platforms are suitable for fine-scale remote sensing applications, for the collection of data along asymmetrical design pathways or in close proximity for specific feature observations, and for momentarily altering pathways due to unfavorable environmental conditions. By mounting multispectral and hyperspectral sensors on unmanned aerial vehicle (UAV) platforms, high-resolution, georeferenced data can be acquired for studying spatial and temporal changes in geological and environmental applications [16,17]. The applications of hyperspectral data from spaceborne, aircraft-based platforms to drone-based platforms have not yet been explored and deciphered appropriately. Machine learning and deep learning techniques can be used to understand and utilize the higher-order variation of field grade spectral data collected using these low-altitude airborne sensors to automate spectral feature-based target detection. It is now important to capitalize on the comparatively high potential of spaceborne and airborne hyperspectral remote sensing datasets based on analyzing different applications that have been addressed by hyperspectral data from different platforms to identify the specificity of each of these two platforms.

In this context, this Special Issue, "*Hyperspectral Remote Sensing from Spaceborne and Low-Altitude Aerial/Drone-Based Platforms—Differences in Approaches, Data Processing Methods, and Applications*", presents the latest achievements in the field of hyperspectral remote sensing data processing and its related applications. A total of 18 manuscripts, all of which were evaluated by professional Guest Editors and reviewers, were submitted for publication in this Special Issue. Subsequently, 11 of them were deemed to be of the appropriate level of quality (based on the standard set by the *Remote Sensing* journal) and were revised, accepted, and published in this Special Issue. The contributions each article within this Special Issue makes to the literature are summarized in the following section.

2. Summary of Papers Presented in This Special Issue

Cristóbal et al. [18] used airborne Hyspex imaging spectrometer data to map the Arctic wetlands in the Yukon Flats National Wildlife Refuge, Alaska. A subset of 120 selected spectral bands with 1 m spatial resolution were used for wetland mapping. A six-class legend was documented based on previous U.S. Geological Survey (USGS) and U.S. Fish and Wildlife Service (USFWS) information and maps. Three different classification methods, namely hybrid classification, spectral angle mapper classification, and maximum likelihood classification, were implemented at selected sites. The best classification performance was achieved using the maximum likelihood classifier with a Kappa index of 0.95. The spectral angle mapper (SAM) classifier and the hybrid classifier showed inferior performances, with Kappa indices of 0.62 and 0.51, respectively. Guha et al. [19] investigated the spectral bands of airborne hyperspectral data of Advanced Visible Infrared Imaging Spectrometer-Next Generation (AVIRIS-NG) data to demarcate the surface signatures associated with the base metal mineralization in the Pur-Banera area, Bhilwara district, Rajasthan, India. Ratio images derived from applying the multi-range spectral feature fitting (MRSFF) method were used to identify the Banded Magnetite Quartzite (BMQ), unclassified calcareous silicates, and quartzite as host rocks of Cu, Pb, Zn mineralization in the study region. The surface imprints of mineralization, such as gossans, and sericitization associated with high Pb-Zn-Cu anomalies were detected in a NE-SW trending structure. Ground-based residual magnetic and laboratory studies (X-ray fluorescence analysis and petrographic study) were also implemented to verify the remote sensing study. The results revealed that the spatial alignment of alteration zones along the structural features are high-potential zones for future detailed Pb-Zn-Cu exploration programs in the study area.

Lu et al. [20] estimated water quality from UAV-borne hyperspectral imagery using nine machine learning algorithms in the Beigong Reservoir, Liuzhou, Guangxi Zhuang Autonomous Region, China. The machine learning algorithms were analytically assessed for the inversion of water quality parameters together with chlorophyll-a (Chl-a) and suspended solids (SS). The experimental results showed that the prediction performance

of the Catboost regression (CBR) model was superior compared to the other algorithms. Additionally, the prediction performances of the Multi-layer Perceptron regression (MLPR) and Elastic net (EN) models were highly unacceptable for the inversion of water quality parameters. Finally, a water quality distribution map for the study area, which can be utilized for large-scale and constant inland water quality monitoring, was generated. Letsoin et al. [21] used unmanned aerial vehicle (UAV) RGB imagery and a transfer learning algorithm to detect sago palm trees in Merauke Regency, Papua Province, Indonesia. A transfer learning strategy was implemented using three deep pre-trained networks, namely SqueezeNet, AlexNet, and ResNet-50, to predict the sago palm trees based on their physical morphology (e.g., their leaves, trunks, flowers, or fruits). The results showed that the ResNet-50 model was the primary model for the flowers, leaves, and trunks for sago palm detection. This baseline model is the first of its kind in the field, and in future studies, it is expected to have high accuracy, including a training validation accuracy of up to 90%, with less elapsed time and an improved number of epochs, which also allows for the use of more datasets of sago palms.

Shirazi et al. [22] developed a Neuro-Fuzzy–Analytic Hierarchy Process (NFAHP) technique for fusing Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) alteration mineral image maps and geological datasets such as lithological maps, geochronological maps, structural maps, and geochemical maps to identify high-potential zones for volcanic massive sulfide (VMS) copper mineralization in the Sahlabad mining area of East Iran. In this study, band ratio and Selective Principal Components Analysis (SPCA) techniques were applied to ASTER VNIR and SWIR bands. Ore mineralization, host–rock lithology, alterations, geochronological, geochemistry, and distance from high-intensity lineament factor communities were considered for applying the Back Propagation Neural Network (BPNN) technique. Subsequently, the Fuzzy–Analytic Hierarchy Process (Fuzzy-AHP) method was implemented to fuse and weigh the information layers. Therefore, a potential map of copper mineralization for the study area, which identified some high-potential zones, was generated. Hashim et al. [23] mapped greenhouse gas (GHG) concentration using the unmanned aerial vehicle-based Sniffer4D sensor in the Pasir Gudang and Tanjung Landsat industrial areas in Johor, Peninsular Malaysia. The results revealed that CO_2 has the highest concentration (mean of 625.235 mg/m^3), followed by CH_4 (mean of 249.239 mg/m^3). The mapped UAV GHG concentration also indicated good agreement with the in situ observations, with an RMSE of 7 and 6 mg/m^3 for CO_2 and CH_4 concentration, respectively. An ozone and nitrogen dioxide mixture (O_3 + NO_2) with a mean concentration of 249 μg/m^3 and an RMSE of 9 μg/m^3 had the other significant concentrations described.

Ding et al. [24] fused ASTER and Sentinel-2 remote sensing images using small-scale geochemical data based on a linear regression model that enhanced the resolution of geochemical elemental layers. This study was conducted in the Xianshuigou area of Northwest China. The regression equation was established using the low-frequency images obtained from image decomposition. Subsequently, the detailed spatial information on the high-frequency images was injected for the fusion. Quantitative correlation coefficients, visual observation, and field sampling emphasized the validation of the fusion results. This approach provided fused large-scale regional geochemical data for mineral exploration, which was key because large-scale geochemical data for areas such as Xinjiang and Tibet in Western China are lacking. Mehranzamir et al. [25] implemented a ground-based lightning locating system using a particle swarm optimization (PSO) algorithm for lightning mapping and monitoring in a 400 km^2 study area at the University Technology Malaysia (UTM), Johor, Peninsular Malaysia. The PSO algorithm was used as a mediator to identify the best location for a lightning strike. It was initiated with 30 particles, considering the outcomes of the magnetic direction finding (MDF) and time difference of arrival (TDOA) techniques. The integration of MDF and TDOA techniques in the PSO-based lightning locating system (LLS) resulted in an accurate lightning detection system with a mean location error of 573 m for a specific local region. This development has enhanced the location accuracy of

lightning strikes in the region. Based on the results of this study, it can be inferred that the LLS using PSO is proficient in precisely identifying and charting lightning discharges in the designated coverage region, thereby validating it as an effective lightning detection system.

Logan et al. [26] used UAV-based hyperspectral imaging for river algae pigment estimation in several locations along the Upper Clark Fork River (UCFR) in Western Montana, USA. Image data were collected across a spectral range of 400–1000 nm with a 2.1 nm spectral resolution. Some spectral indices for the detection of algal standing crops were developed using brute-force analysis. The results show that spectral band ratios provide a suitable method for assessing chlorophyll a and phycocyanin standing crops contained within, near, and on the surface of the blooms of the filamentous nuisance algae growing in the UCFR. Hashim et al. [27] mapped and analyzed consecutive changes in land-use–land-cover (LULC) and water yield (WY) between 2000 and 2015 in the Johor River Basin (JRB), Peninsular Malaysia, by explicitly comparing satellite-based and in situ-derived WY and depicting changes in WY in relation to LULC change magnitudes within watersheds. The WY was estimated using the water balance equation, which determines the WY from the equilibrium of precipitation minus evapotranspiration (ET). The precipitation and ET information were derived from the Tropical Rainfall Measuring Mission (TRMM) and moderate-resolution imaging spectroradiometer (MODIS) satellite data, respectively. The LULC maps were extracted from Landsat-Enhanced Thematic Mapper Plus (ETM+) and Landsat Operational Land Imager (OLI). From the year 2000 to 2015, agricultural areas other than oil palm-based areas enlarged to 11.07%, forest areas diminished to 32.15%, oil palm-based areas enlarged to 11.88%, urban areas were amplified to 9.82%, and WY improved to 15.76%. The results of this study indicate a good agreement between the satellite-based derived quantities and in situ measurements, with an average bias of ±20.04 mm and ±43 mm for precipitation and ET, respectively. Abedini et al. [28] developed a novel approach to optimize remote sensing-based evidential variables using constructed mining geochemistry models for machine learning (ML)-based copper mineralization prospectivity mapping (MPM). The geochemical mining methods and satellite remote sensing data (Landsat ETM+) processing were scrutinized to select the optimal evidential variables for constructing a mineralization prospectivity map using a random forest (RF) approach in the Toroud-Chah Shirin (TCS) belt and the Alborz magmatic belt, North Iran. The lithology, structure, alterations, and geochemical evidential layers were integrated using the RF algorithm to achieve the regional-scale prospectivity mineral mapping of porphyry copper deposits. As a result, this study explored buried copper deposits in the TCS belt through innovative approaches by integrating multi-source geoscientific datasets.

3. Concluding Remarks

We would like to thank Ms. Helen Wang (Assistant Editor), as well as all authors and reviewers who donated their time, research, and effort to this Special Issue. We wish to extend our sincere gratitude to the MDPI editorial team for supporting the Guest Editors in efficiently processing each manuscript. The sympathetic and judicious comments delivered by the reviewers improved each of the papers published in this Special Issue, which came to fruition only because their authors were willing to volunteer their time and attention. We hope that the studies published in this Special Issue will provide insights to other remote sensing researchers about the application of hyperspectral remote sensing data for geological mapping and environmental monitoring and modeling.

References

1. Pearlman, J.S.; Barry, P.S.; Segal, C.C.; Shepanski, J.; Beiso, D.; Carman, S.L. Hyperion, a space-based imaging spectrometer. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1160–1173. [CrossRef]
2. Green, R.O.; Pavri, B.E.; Chrien, T.G. On-orbit radiometric and spectral calibration characteristics of EO-1 hyperion derived with an underflight of AVIRIS and in-situ measurements at Salar de Arizaro, Argentina. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1194–1203. [CrossRef]
3. Cocks, T.; Jenssen, R.; Stewart, A.; Wilson, I.; Shields, T. The HyMap airborne hyperspectral sensor: The system, calibration and performance. In Proceedings of the 1998 Proceedings of the 1st EARSeL Workshop on Imaging Spectroscopy, Zurich, Switzerland, 6–8 October 1998; pp. 6–8.
4. Hu, J.; Lanzon, A. An innovative tri-rotor drone and associated distributed aerial drone swarm control. *Robot. Autonomous Syst.* **2018**, *103*, 162–174. [CrossRef]
5. Iwasaki, A.; Ohgi, N.; Tanii, J.; Kawashima, T.; Inada, H. Hyperspectral Imager Suite (HISUI) -Japanese hyper-multi spectral radiometer. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011. [CrossRef]
6. LLoizzo, R.; Guarini, R.; Longo, F.; Scopa, T.; Formaro, R.; Facchinetti, C.; Varacalli, G. Prisma: The Italian hyperspectral mission. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Ruisui, Taiwan, 31 October 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2018; pp. 175–178. [CrossRef]
7. Goodenough, D.G.; Dyk, A.; Niemann, K.O.; Pearlman, J.S.; Chen, H.; Han, T.; Murdoch, M.; West, C. Processing Hyperion and ALI for forest classification. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1321–1331. [CrossRef]
8. Kruse, F.A.; Boardman, J.W.; Huntington, J.F. Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1388–1400. [CrossRef]
9. Pour, A.B.; Hashim, M. Evaluation of Earth Observing-1 (EO1) data for lithological and hydrothermal alteration mapping: A case study from Urumieh-Dokhtar Volcanic Belt, SE Iran. *J. Indian. Soc. Remote Sens.* **2015**, *43*, 583–597. [CrossRef]
10. Pour, A.B.; Hashim, M.; Marghany, M. Exploration of gold mineralization in a tropical region using Earth Observing-1 (EO1) and JERS-1 SAR data: A case study from Bau gold field, Sarawak, Malaysia. *Arab. J. Geosci.* **2014**, *7*, 2393–2406. [CrossRef]
11. Feng, W.; Qi, S.; Heng, Y.; Zhou, Y.; Wu, Y.; Liu, W.; He, L.; Li, X. Canopy vegetation indices from in situ hyperspectral data to assess plant water status of winter wheat under powdery mildew stress. *Front. Plant Sci.* **2017**, *8*, 1219. [CrossRef]
12. Shang, X.; Chisholm, L.A. Classification of Australian native forest species using hyperspectral remote sensing and machine-learning classification algorithms. *IEEE J. Sel. Top. Appl. Earth Obser. Remote Sens.* **2014**, *7*, 2481–2489. [CrossRef]
13. Teke, M.; Deveci, H.S.; Haliloglu, O.; Gürbüz, S.Z.; Sakarya, U. A short survey of hyperspectral remote sensing applications in agriculture. In Proceedings of the 2013 6th International Conference on Recent Advances in Space Technologies (RAST), Istanbul, Turkey, 12–14 June 2013; pp. 171–176.
14. Paoletti, M.E.; Haut, J.M.; Plaza, J.; Plaza, A. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 279–317. [CrossRef]
15. Pour, A.B.; Ranjbar, H.; Sekandari, M.; El-Wahed, M.; Hossain, M.S.; Hashim, M.; Yousefi, M.; Zoheir, B.; Wambo, J.D.T.; Muslim, A.M. Remote sensing for mineral exploration. In *Geospatial Analysis Applied to Mineral Exploration Remote Sensing, GIS, Geochemical, and Geophysical Applications to Mineral Resources*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 17–149. [CrossRef]
16. Park, S.; Choi, Y. Applications of unmanned aerial vehicles in mining from exploration to reclamation: A review. *Minerals* **2020**, *10*, 663. [CrossRef]
17. Yao, H.; Qin, R.; Chen, X. Unmanned aerial vehicle for remote sensing applications—A review. *Remote Sens.* **2019**, *11*, 1443. [CrossRef]
18. Cristóbal, J.; Graham, P.; Prakash, A.; Buchhorn, M.; Gens, R.; Guldager, N.; Bertram, M. Airborne Hyperspectral Data Acquisition and Processing in the Arctic: A Pilot Study Using the Hyspex Imaging Spectrometer for Wetland Mapping. *Remote Sens.* **2021**, *13*, 1178. [CrossRef]
19. Guha, A.; Kumar Ghosh, U.; Sinha, J.; Pour, A.B.; Bhaisal, R.; Chatterjee, S.; Kumar Baranval, N.; Rani, N.; Kumar, K.V.; Rao, P.V.N. Potentials of Airborne Hyperspectral AVIRIS-NG Data in the Exploration of Base Metal Deposit—A Study in the Parts of Bhilwara, Rajasthan. *Remote Sens.* **2021**, *13*, 2101. [CrossRef]
20. Lu, Q.; Si, W.; Wei, L.; Li, Z.; Xia, Z.; Ye, S.; Xia, Y. Retrieval of Water Quality from UAV-Borne Hyperspectral Imagery: A Comparative Study of Machine Learning Algorithms. *Remote Sens.* **2021**, *13*, 3928. [CrossRef]
21. Letsoin, S.M.A.; Purwestri, R.C.; Rahmawan, F.; Herak, D. Recognition of Sago Palm Trees Based on Transfer Learning. *Remote Sens.* **2022**, *14*, 4932. [CrossRef]
22. Shirazi, A.; Hezarkhani, A.; Beiranvand Pour, A.; Shirazy, A.; Hashim, M. Neuro-Fuzzy-AHP (NFAHP) Technique for Copper Exploration Using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and Geological Datasets in the Sahlabad Mining Area, East Iran. *Remote Sens.* **2022**, *14*, 5562. [CrossRef]
23. Hashim, M.; Ng, H.L.; Zakari, D.M.; Sani, D.A.; Chindo, M.M.; Hassan, N.; Azmy, M.M.; Pour, A.B. Mapping of Greenhouse Gas Concentration in Peninsular Malaysia Industrial Areas Using Unmanned Aerial Vehicle-Based Sniffer Sensor. *Remote Sens.* **2023**, *15*, 255. [CrossRef]
24. Ding, H.; Jing, L.; Xi, M.; Bai, S.; Yao, C.; Li, L. Research on Scale Improvement of Geochemical Exploration Based on Remote Sensing Image Fusion. *Remote Sens.* **2023**, *15*, 1993. [CrossRef]

25. Mehranzamir, K.; Pour, A.B.; Abdul-Malek, Z.; Afrouzi, H.N.; Alizadeh, S.M.; Hashim, M. Implementation of Ground-Based Lightning Locating System Using Particle Swarm Optimization Algorithm for Lightning Mapping and Monitoring. *Remote Sens.* **2023**, *15*, 2306. [CrossRef]

26. Logan, R.D.; Torrey, M.A.; Feijó-Lima, R.; Colman, B.P.; Valett, H.M.; Shaw, J.A. UAV-Based Hyperspectral Imaging for River Algae Pigment Estimation. *Remote Sens.* **2023**, *15*, 3148. [CrossRef]

27. Hashim, M.; Baiya, B.; Mahmud, M.R.; Sani, D.A.; Chindo, M.M.; Leong, T.M.; Pour, A.B. Analysis of Water Yield Changes in the Johor River Basin, Peninsular Malaysia Using Remote Sensing Satellite Imagery. *Remote Sens.* **2023**, *15*, 3432. [CrossRef]

28. Abedini, M.; Ziaii, M.; Timkin, T.; Pour, A.B. Machine Learning (ML)-Based Copper Mineralization Prospectivity Mapping (MPM) Using Mining Geochemistry Method and Remote Sensing Satellite Data. *Remote Sens.* **2023**, *15*, 3708. [CrossRef]

Article

Machine Learning (ML)-Based Copper Mineralization Prospectivity Mapping (MPM) Using Mining Geochemistry Method and Remote Sensing Satellite Data

Mahnaz Abedini [1], Mansour Ziaii [1] , Timofey Timkin [2] and Amin Beiranvand Pour [3,*]

[1] Faculty of Mining, Petroleum and Geophysics Engineering, Shahrood University of Technology, Shahrood 3619995161, Iran; abedini.m@shahroodut.ac.ir (M.A.); mziaii@shahroodut.ac.ir (M.Z.)
[2] School of Earth Sciences & Engineering, Tomsk Polytechnic University, 634050 Tomsk, Russia; timkintv@mail.ru
[3] Institute of Oceanography and Environment (INOS), Universiti Malaysia Terengganu (UMT), Kuala Nerus 21030, Terengganu, Malaysia
* Correspondence: beiranvand.pour@umt.edu.my; Tel.: +60-9-6683824; Fax: +60-9-6692166

Abstract: The exploration of buried mineral deposits is required to generate innovative approaches and the integration of multi-source geoscientific datasets. Mining geochemistry methods have been generated based on the theory of multi-formational geochemical dispersion haloes. Satellite remote sensing data is a form of surficial geoscience datasets and can be considered as big data in terms of veracity and volume. The different alteration zones extracted using remote sensing methods have not been yet categorized based on the mineralogical and geochemical types (MGT) of anomalies and cannot discriminate blind mineralization (BM) from zone dispersed mineralization (ZDM). In this research, an innovative approach was developed to optimize remote sensing-based evidential variables using some constructed mining geochemistry models for a machine learning (ML)-based copper prospectivity mapping. Accordingly, several main steps were implemented and analyzed. Initially, the MGT model was executed by studying the distribution of indicator elements of lithogeochemical data extracted from 50 copper deposits from Commonwealth of Independent States (CIS) countries to identify the MGT of geochemical anomalies associated with copper mineralization. Then, the geochemical zonality model was constructed using the database of the porphyry copper deposits of Iran and Kazakhstan to evaluate the geochemical anomalies related to porphyry copper mineralization (e.g., the Saghari deposit located around the Chah-Musa deposit, Toroud-Chah Shirin belt, central north Iran). Subsequently, the results of mining geochemistry models were used to produce the geochemical evidential variable by vertical geochemical zonality (Vz) (Pb × Zn/Cu × Mo) and to optimize the remote sensing-based evidential variables. Finally, a random forest algorithm was applied to integrate the evidential variables for generating a provincial-scale prospectivity mapping of porphyry copper deposits in the Toroud-Chah Shirin belt. The results of this investigation substantiated that the machine learning (ML)-based integration of multi-source geoscientific datasets, such as mining geochemistry techniques and satellite remote sensing data, is an innovative and applicable approach for copper mineralization prospectivity mapping in metallogenic provinces.

Keywords: machine learning; mining geochemistry; remote sensing; random forest; geochemical zonality; copper mineralization; mineral prospectivity mapping

1. Introduction

Numerous well-known mineral deposits are undergoing long-term mining and exploitation along with shallow ore resources have been gradually exhausted around the world; therefore, the exploration of covered and buried mineral deposits is required to generate innovative approaches. Appreciating the comprehensive assessment of deep ore resources effectively, economically, and rapidly has become the frontier field of mineral

exploration research. Mining geochemistry is one of the most essential disciplines characteristically fulfilled for exploration of deep and concealed mineralization [1–23]. The key purpose of mining geochemistry is to propose an approach for differentiating blind mineralization (BM), outcropping, and zone dispersed mineralization (ZDM) [1,4,10,15,17,18,20]. Many mining geochemistry methods have not been practiced for mineral prospectivity mapping (MPM) using geographic information systems (GIS) [22,24]. These techniques could be divided into three classes, namely mineralogical and geochemical type (MGT), geochemical zonality, and metallometry.

In the last decades, numerous approaches have been efficaciously executed to sort geochemical data responsive for MPM, e.g., [22,24–30]. Recognition of anomalies associated with ore mineralization and the integration of multi-source geoscientific data is necessary for MPM. The main emphasis has been placed upon big data analytics designed to recognize and incorporate anomalies linked to multi-mineralogical and geochemical types (MGTs) of mineralization, e.g., [10,15,31–33]. Incidentally, innovative mathematical models for the delimitation and interpretation of the multi-MGT anomalies have been recommended and implemented, e.g., [34–40].

The main applications of remote sensing satellite data in MPM are to provide detailed information for identifying hydrothermal alteration zones, detecting geological structures, and distinguishing lithological units. During the previous decades, the extraction of hydrothermal alterations was considered one of the most essential implementations of remote sensing data for provincial mineral exploration campaigns e.g., [41–48]. Discrimination of different alteration zones is a qualitative method to evaluate shallow ore resources. These different alteration zones have not been categorized (considered) based on the MGT of anomalies for MPM and cannot differentiate blind mineralization (BM) from zone dispersed mineralization (ZDM). The MGT and geochemical zonality models are able to recognize the hydrothermal alteration areas related to ore mineralization and to evaluate deep and buried ore resources. Metallometry provides a quantitative connection among point source soil and stream sediment anomalies to improve anomaly detection. This relation is allied with the concept of productivity (area × concentration) and is not governed by concentrations. Monotonous computation of regional data has not been an easy task, but with the advent of GIS, it has been easily premeditated. Thus, metallometric techniques are developed to map geochemical anomalies associated with deposits [49,50]. There are various methods for GIS-based MPM, and different relevant evidential layers are integrated. Ziaii et al. [24] used two sets of evidential layers to relate the presentation of geochemical anomalies of a pathfinder element (e.g., Cu) and Vz in MPM, and demonstrated that Vz in MPM outperforms Cu in MPM. Thus, using the multiplicative supra-ore (Pb × Zn) and sub-ore elements (Cu × Mo) (instead of the single elements), the limitations of using traditional interpolation methods, such as Kriging, on the data [30] were fundamentally solved.

Machine learning (ML) algorithms, such random forest (RF) have great capability for data integration and MPM. Rodriguez-Galiano et al. [51] compared the performances of MLAs, such as artificial neural networks, regression trees, RF, and support vector machines in MPM. The results indicated that the RF overtook the former MLA algorithms and revealed advanced steadiness and vigor with erratic training parameters. Carranza and Laborte [52] investigated the susceptivity of the RF to diverse sets of training data and the performance of RF compared to data-driven methods of MPM. They concluded that RF precisely considers the spatial correlations among the predictor variables and the training deposit and non-deposit locations, and RF is more beneficial than other methods. Carranza and Laborte [53,54] used RF for MPM with a small number of prospects and data with missing values. They realized that RF overtook WofE where a small number of prospects are known. Moreover, RF can control missing values in evidential data through an RF-based attribution method. It is not a black-box technique, unlike artificial neural networks. Zhang et al. [55] tested the capabilities of RF for MPM. They selected several evidential maps allied with Au mineralization, including singularity indices, principal component scores, and distance to intrusions and faults as inputs to the model. Then, they

used RF to rank the significance and to recognize the susceptivity of the evidential maps generated based on their spatial correlations to the documented gold mineralization. The results demonstrate that RF could be utilized successfully for MPM in areas with some identified mineral occurrences. Wang et al. [56] used recursive feature elimination and RF to select indicator elements and map geochemical anomalies associated with Cu and W mineralization in a regional scale. Zhang et al. [57] interpreted RF modeling using the variable importance and partial dependence plot and also examined the effectiveness of outlier-based training samples in contrast to using known mineralized locations. They concluded that the MPM resulting from an RF built on outlier-based training set was substandard to those produced by an RF constructed on mineralized location-based training sets, which discloses the prejudice concerning documented mineralized sites in training a data-driven MPM algorithm.

The NE-trending Toroud-Chah Shirin (TCS) belt is placed in the Alborz magmatic belt in the north of Iran [58–61] (Figure 1). A few studies were conducted in the TCS belt and most of the acquired information is only related to depths of less than 100 m [59–64]. Exploration for new deep, concealed, and economic ore resources (mineral deposits) in the TCS belt is challenging owing to the insufficiency of comprehensive geological datasets and arid and mountainous geochemical landscapes [15,35,65–67]. In this research, a novel approach was established to optimize remote sensing-based evidential variables using constructed mining geochemistry models for a machine learning (ML)-based copper mineralization prospectivity mapping (MPM). The mining geochemistry methods and satellite remote sensing data processing was examined to select the optimal evidential variables for constructing MPM using the RF technique. Based on the lithogeochemical data extracted from the copper deposits of Commonwealth of Independent States (CIS) countries, the quantitative MGT model was constructed and used to identify the MGT of lithogeochemical anomalies of the Saghari deposit around the Chah-Musa copper deposit in the TCS belt (Figure 1). Subsequently, the alteration zones and geochemical zonality coefficient related to the identified MGT were selected. According to the database of copper deposits in Kazakhstan and Iran, the quantitative zonality model was constructed to assess the lithogeochemical anomalies of the Saghari deposit. Finally, the set of evidential variables, including lithology, structure, alteration and geochemical zonality, and deposit/non-deposit locations were integrated using RF for MPM in the TCS belt.

2. Geological Setting

The TCS magmatic arc (Figure 1) comprises an elevated block confined by the Toroud fault to the south, and the Anjilow fault to the north that have the same trends. In this block, NE–SW-trending volcano-intrusive rocks are composed of middle- to upper Eocene andesites and dacites, and Oligocene diorites. Middle- to late Eocene igneous activity include: (i) local andesitic lava flows associated with marls, tuffaceous marlstones, sandstones, and explosively deposited rhyolitic to rhyodacitic tuffs; (ii) basaltic andesite and lava flows, and pyroclastic rocks of andesite, trachy-andesite; (iii) hypabyssal intrusive rocks and subordinate dacitic–rhyodacitic rocks [63,68]. The faults and fractures controlled ore mineralization in the TCS belt, e.g., [60,61].
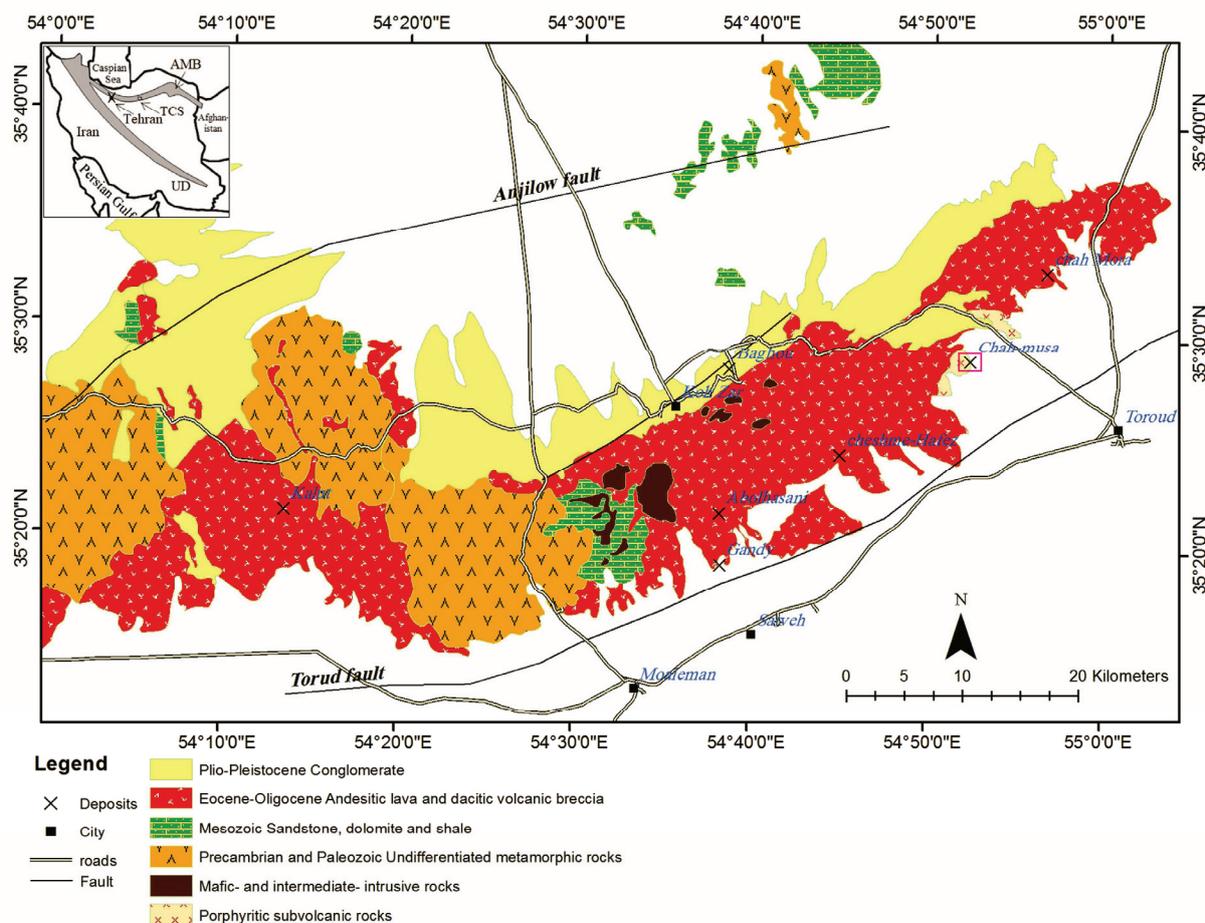
Figure 1. The geographical position of the TCS belt in Iran and regional geological map of the TCS range, showing mineralization zones, geological structures, and the lithological units. (based on 1:250,000 geologic map of Toroud). AMB: Alborz magmatic belt, UD: Urumieh-Dokhtar zone, TCS: Toroud-Chah Shirin range [69].

Figure 2 shows geological map of the Saghari deposit, which is situated in the northeastern part of the geological map of Toroud (1:250,000 scale). The Saghari deposit is placed around the Chah-Musa deposit. The Chah-Musa deposit is an active mine. The Chah-Musa porphyritic hornblende diorite subvolcanic intrusion is positioned in the eastern part of the TSC belt [62,63]. It is emplaced into the Eocene volcanic sequence. The egg-shaped Chah-Musa body hosts a copper deposit and interlopes an Eocene sequence of volcanic breccia, agglomerate, and red tuffaceous sediment. Hand specimens exhibit a porphyritic structure [63]. Disseminated-veinlet copper mineralization in the Chah-Musa deposit occurred in Calc-alkaline subvolcanic porphyry-biotite-hornblende andesite bodies which are associated with phyllic and propylitic alterations. Supergene processes produced extensive alterations in hypogene sulfide minerals, such as pyrite, chalcopyrite, and bornite to secondary chalcocite, covellite, digenite, and malachite supergene minerals. The mineralization system is connected to the evolution of hydrothermal fluid mineralization and mixing with cold and low salinity meteoric water resulting in disseminated-veinlet copper mineralization at deep zones, and vein-type copper, zinc, and lead at shallow zones. This has led to the development of elemental and mineralogical zonation [62]. Despite its importance, geologic information about the Chah-Musa deposit is scarce [63].
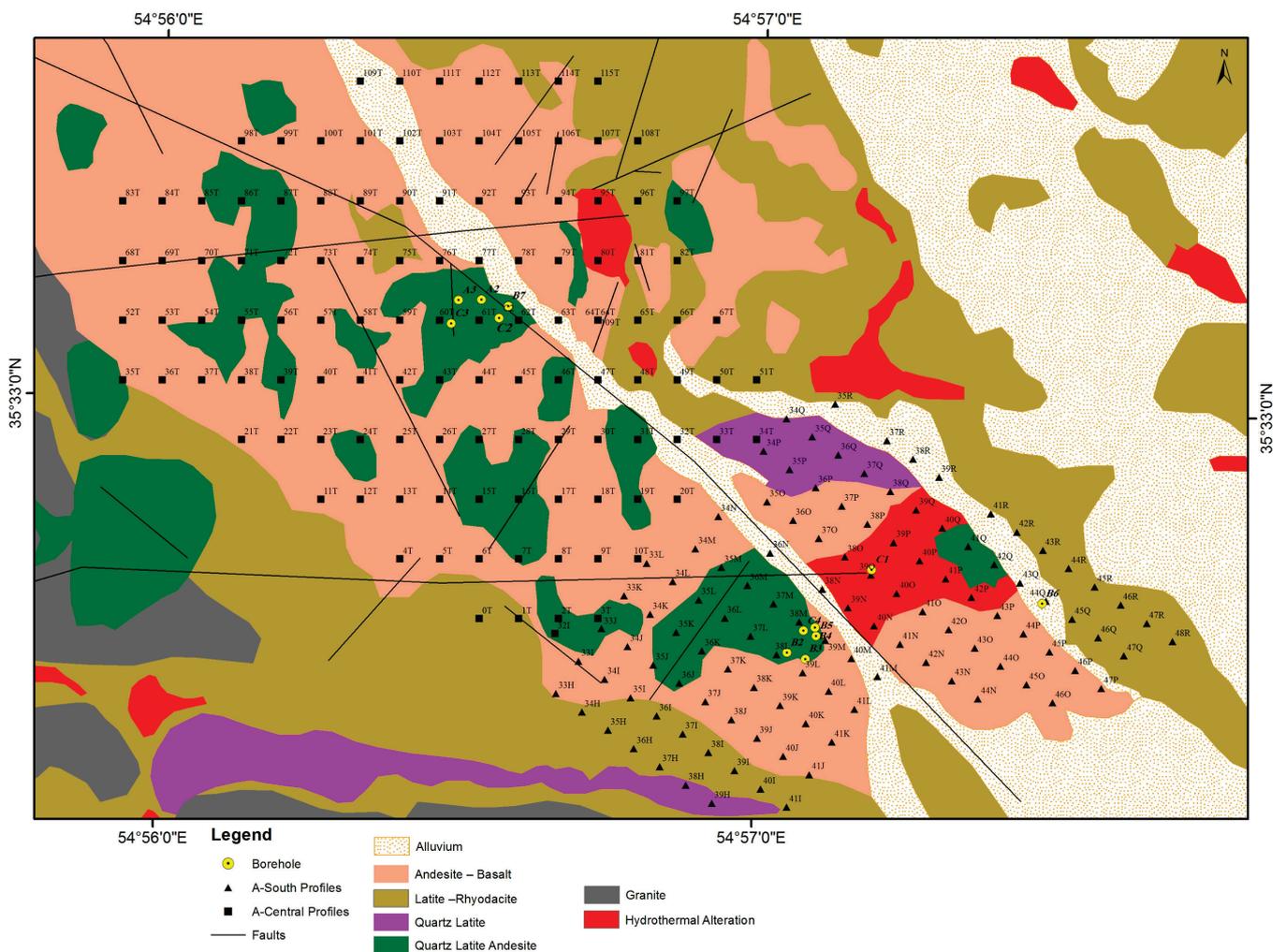
Figure 2. Geological map of the Saghari deposit showing A-Central and A-South profiles (80×100 m^2) and the location of exploratory drilled boreholes.

The study area has an arid climate with deprived vegetation. The occurrence of specific geographical conditions in an arid environment has resulted in the unfeasible execution of geochemical reconnaissance investigations. Many of the identified copper deposits in the study area are considered through advanced zonal patterns of mineralization and hydrothermal alteration zones. The zonal patterns show remarkable differences in element content, reflecting variations in the mineralogical and geochemical compositions of the mineralized and hydrothermally altered zones [24].

3. Methodology

Figure 3 illustrates an outline of the methodological flowchart applied in this analysis to identify and evaluate the geochemical anomalies of the Saghari deposit around the Chah-Musa deposit and to generate mineral prospectivity mapping in the TCS belt.
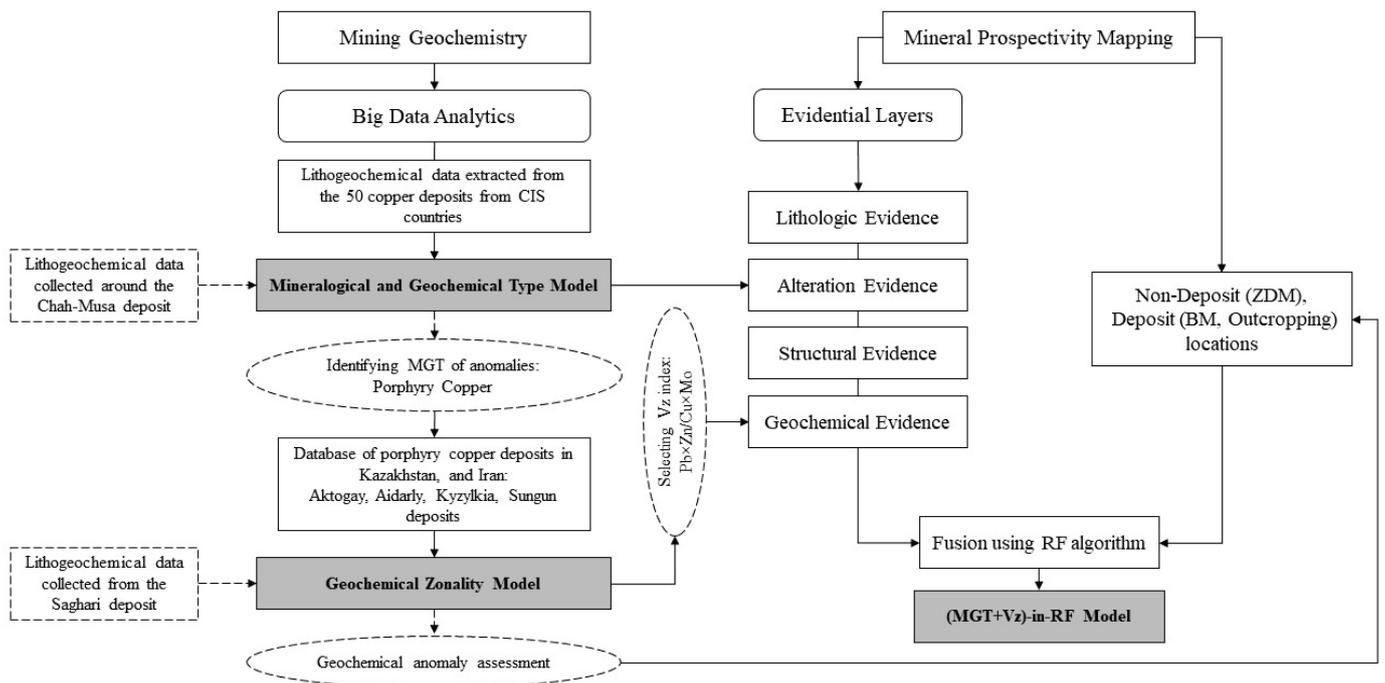
Figure 3. Graphical representation of the workflow used in this study.

3.1. Big Data Analytics

The term "Big Data" refers to large datasets and their characteristics, such as different kinds of data sources. The features of big data are essential to the geosciences. The enormous assortment of earth observation data provides a great opening to execute big data analytics to solve geosciences issues [32]. The implications of implementing big data analytics to mineral exploration are to create the geochemical models utilizing diverse types of big data, to categorize the features of the distribution of trace element contents, and to assess the geochemical anomalies. The fundamental function of big data analytics is prediction, making it an ultimate methodology for evaluating multi-formational geochemical anomalies. The databank used to identify the MGTs of geochemical anomalies of the Saghari deposit comprises the lithogeochemical data extracted from the 50 copper deposits in CIS countries. According to the obtained MGT from the previous model, the database used to assess the anomalies associated with mineralization comprises the porphyry copper deposits in Kazakhstan and Iran.

3.2. Mineralogical and Geochemical Type (MGT)

For the geochemical characteristics of deposits, ore occurrence, and associated lithogeochemical anomalies of a particular metal, it is possible to select indicator elements that determine if an area belongs to a specific mineralogical and geochemical type (MGT). The elements that determine the MGT of deposits make it possible to separate similar geochemical anomalies by depicting their composition on a triple diagram. The relative contents of three groups of elements selected for the characteristics of the corresponding MGT are applied on the sides of an equilateral triangle. The percentage of each three components is calculated from the total amount, and the position of the MGT on the plane of the triangle is marked with a point. The displacement of the point from one of the vertices of the triangle determines whether the ore occurrence belongs to the corresponding MGT. Therefore, the separation of copper ore occurrence is possible by constructing a triple diagram using big data analytics.

3.3. Vertical Geochemical Zonality Method

Kitaev [70] suggested a multidimensional geochemical field study built on the concept that geological space is comprised of geochemical fields signifying the zonality of element associations. This study considers the dispersion of elements to separate multielement anomalies according to Vz values. Grigorian [4] demonstrated that patterns of Vz around exposed deposits are different from patterns of Vz associated with blind deposits. Recognition of the zonality of geochemical haloes associated with blind deposits can be attained by four corresponding analyses [10]: (i) analysis of element associations representing supra-ore and sub-ore haloes of deposits; (ii) analysis of a single component, implying a false anomaly; (iii) analysis of mean values of indicator elements outside significant geochemical anomalies to eliminate background noise; (iv) mapping of multiplicative geochemical anomalies (i.e., Vz coefficients). The geochemical zonality coefficients for the assessment of the anomalies are derived through investigation of the primary geochemical haloes of typical standard deposits. In most cases, residual secondary soil haloes are correlated in composition and structure with the ore bodies and primary haloes which have produced them. The fruitful application is linked to the landscape–geochemical situations in the ore zones. The universal model for discovering blind mineralization and determining the erosional level is according to three criteria: (i) contour zoning [3,4], (ii) natural field of geochemical associations [71], and (iii) metallometry [42,54].

Based on the average contents of ore elements from big data analytics, the Vz values are calculated and converted into the zoning sequence of geochemical haloes. They are interpolated along the averaging line. Therefore, the geochemical zonality model is constructed. This model is used to assess geochemical anomalies. This model adopts a correlation among vertical zonality coefficient and depth of mineralization response. While the Vz was established initially for the study of lithogeochemical data [70], Grigorian [4] has confirmed that it could be executed using stream sediment geochemical data to study the erosional surfaces of multiplicative anomalies signifying different erosional levels of mineral deposits [4]. Therefore, efficacious identification of anomalous erosional surfaces is allied to the landscape situations of mineralized zones.

3.4. Remote Sensing Data

The system of hydrothermal ore formation was fundamentally expressed by the analysis of the composition and physicochemical properties of ore-bearing solutions, sources of ore components and ore-forming fluids, and the conditions and mechanisms of ore deposition [72]. The origin of hydrothermal solutions that take part in the formation of most commercial base metal deposits is one of the most important and complicated problems of the model of ore formation. Remote sensing methods are based on the theory of hydrothermal alterations. Remote sensing data are successfully used for mapping hydrothermally altered rocks by virtue of their spectral signatures [73–75]. Porphyry copper deposits are characteristically associated with hydrothermal alteration mineral zones, such as gossan, argillic, phyllic, and propylitic [73,76]. Iron oxide minerals (e.g., limonite, jarosite, and a hematite called gossan) show spectral absorption features in the visible to near infrared (0.4 to 1.1 μm) region [77]. Hydroxyl-bearing minerals, such as clay groups, sulfate minerals, and carbonates, represent spectral absorption features in the shortwave infrared (2.0 to 2.5 μm) region [77,78]. Previous studies have confirmed the proficiency of multispectral remote sensing sensors (e.g., ETM+ and ASTER) to map the hydrothermal alterations in minerals and zones associated with copper mineralization [79–82].

3.4.1. Landsat ETM+ Data Characteristics, Preprocessing and Processing Techniques

Since July 1999, the Landsat ETM+ multispectral sensor onboard the Landsat 7 satellite has obtained images of the Earth's surface in eight bands. It identifies spectrally-filtered radiation in visible and near infrared (VNIR), short wave infrared (SWIR), longwave infrared (LWIR), and panchromatic bands. It has a 183 km swath width and orbits at an

altitude of 705 km. Landsat 7 collects data in accordance with world-wide reference system (WRS), which has cataloged the world's land mass into 57,784 scenes, each 183 km wide (E–W) by 170 km long (N–S). An ETM+ scene has spatial resolution of 30 m in bands 1–5 and 7 while band 6, thermal infrared (TIR), has a 60 m spatial resolution, and band 8 has spatial resolution of 15 m. In this research, ETM+ image level L1T (Pass 162 and Raw 35) of one scene covering the TCS belt and (date of acquisition 20 July 2000) was used. These data were corrected radiometrically and geometrically.

Principal Component Analysis (PCA)

PCA is a multivariate statistical method which is extensively utilized to diminish the dimension of input data; furthermore, the possibility of useful data loss is minimized. PCA finds a set of linearly uncorrelated components called principal components (PCs) [83,84]. PCs are the projection of input data onto the principal axes or eigenvectors. The output components are arranged based on the variance, in descending order. The PCA builds up a new set of axes orthogonal to each other, and each component is orthogonal to the preceding component [85].

Implementing PCA to map alterations in minerals using multispectral remote sensing data was suggested by Prado and Crosta [86]. The PCA converts the original dataset into a considerably smaller and easier-to-interpret set of uncorrelated variables. The main objective is to eliminate redundancy in multispectral data and extract useful information. PCA is extensively used for mapping hydrothermal alteration minerals associated with ore mineralization in the metallogenic provinces [82,87–89]. This method is worthwhile for multivariate datasets, such as multispectral satellite images, with the tenacity of emphasizing spectral responses associated with specific hydrothermal alteration minerals [90]. The number of output PCs is equal to the input spectral bands. However, a small number of PCs deliver significant information about the dataset. Certain spectral bands are designated which comprise absorption and reflection features of alteration minerals. Thus, a new image (PC) is produced on the axes with the new coordinate system [90,91]. The resulting PC (final image) is usually more intelligible compared to the original images.

A PC image comprises the distinctive involvement of eigenvector loadings (magnitude and sign) for the absorption and reflection bands of alteration minerals. The image tone will be bright if the loading is positive in the reflective band of the mineral, and the image tone will be dark for the enhanced target mineral if it is negative [91]. The PCA technique has been successfully implemented with ETM+ and ASTER multispectral data for highlighting spectral responses allied to particular hydrothermal alteration minerals accompanying porphyry copper deposits [75,79,80,82,86–88,92,93]. In this analysis, for the ascertainment of an image that comprises information correlated to the spectral signatures of particular target minerals, normal data distribution was assumed, and the covariance matrix was used to calculate the PCs. The eigenvector loadings in each PC image were examined. A PC with significant loadings, which shows an analogous trend to the spectral features of the target alteration minerals, is deliberated as a proper component for identifying the target zones. In this study, PCA was implemented to six ETM+ spectral bands for mapping iron oxides/hydroxides and clay minerals.

3.5. Machine Learning Algorithms

Machine learning algorithms have recently received remarkable attention for geochemical anomaly recognition, especially for multi-formational geochemical anomalies and data integration [94]. Machine learning is a field of artificial intelligence that trains an algorithm to wisely find features and patterns hidden in large amounts of data in order to make decisions or predictions based on new data. Machine learning is able to calculate nonlinear and complex patterns even when deprived of a prior hypothesis that the data follow a known multivariate probability distribution [95,96]. Both positive and negative samples are required to train supervised machine learning algorithms for geochemical anomaly identi-

fication. The sites of the documented deposits are considered positive training samples; however, not all selected negative samples are true non-deposit sites [96,97].

Random Forest (RF)

Random forest (RF) is a supervised machine learning method that has been used in geochemical exploration and data integration. This method is an ensemble algorithm that demonstrates an extension of classification and regression trees. It could be utilized to categorize or expect the rate of a target variable built on a quantity of evidential variables [55]. Using the tree structure, the RF algorithm partitions diverse samples with corresponding labels at leaf nodes of ensemble trees [57]. It is consecutively executed from a root node to a terminal node (leaf) to make recurrent predictions [55,98].

The basic classifiers in the RF method are classification and regression trees, which utilize a bagging system to determine that training subsets are casually selected, with each subset forming a decision tree [99]. The bagging system means that nearly one-third of the accessible training samples are not utilized in the building of RF trees; instead, they are utilized to certify the expectation accuracy (also called "out-of-bag" (OOB) samples). The OOB error is a balanced evaluation of the simplification error during RF analysis [96]. The evidential variables utilized for each node in the decision tree are also unsystematically selected. The result of RF modeling is related to the mediocre expectation of all trees convoluted in the model [55]. The RF algorithm starts by dividing parent nodes (i.e., evidential features) into binary portions, where child nodes are purer than the parent node. Examining all of the splits produces optimal splits that amplify the "purity" of the resulting trees. The RF algorithm utilizes the Gini impurity index to compute the information purity of child nodes equated with their parent nodes, with dividing thresholds calculated from the extreme decline in purity rates [98]. This dividing process is reiterated until a halt circumstance is obtained [55].

The RF algorithm as a nonlinear and nonparametric method is suitable for an extensive array of expectation delinquents and delivers adaptable prominence indices [100]. The key RF adaptable prominence indices include the "Gini importance" and the "Accuracy importance". The first designates the change in "Gini gain" dividing criterion presented by each variable through classification. The second equates the expectation accuracy before and after a variable is arbitrarily permuted without changing the rest of the variables, and the difference in prediction accuracy is considered an index of the permuted variable. This measure deliberates the effect of each variable independently and the multivariate interactions with other variables. The elucidation capability of RF modeling is conducted using indices of variable importance. Variable importance indices have been popular for variable selection, but still there is irregularity when diverse sorts of predictor variables are utilized [100]. The variable importance ranking in RF based on diverse training sets delivers insights into evidence maps for MPM [57].

The benefits of the RF algorithm take in the bagging technique, which implicates unsystematic resampling and replacement, and produces diverse training subsets, which could successively be utilized to create decision trees. Therefore, the multiplicity inside the model is increased, and relationships among trees are avoided through the RF process. It permits excellent stability and prediction accuracy because some inputs are not utilized, escaping definite variations. Superlative evidential characteristics are utilized as dividing points to facilitate tree growth throughout the RF process. The unsystematic assortment of evidential characteristics is utilized as fragment of the whole set of input evidential characteristics. As such, it decreases relationships among trees and the simplification error inside RF models [55].

3.6. Properties of Geochemical Data

Secondary and primary geochemical dispersion zones are sampled and analyzed to identify geochemical patterns reflecting the primary geological structures in the Saghari deposit around the Chah-Musa deposit. A total of 233 samples were taken from 2 areas

of the Saghari deposit (A-Central and A-South) at a local scale (see Figure 2). The soil samples were assayed using inductively coupled plasma–mass spectrometry (ICP-MS) and the concentrations of 34 elements were determined. Additionally, 12 exploratory drilled boreholes were excavated in the A-Central and A-South areas, and 146 samples were taken from these boreholes (sampling intervals of 5 m) using a chip-channel sampling technique. The rock samples were submitted to the laboratory for chemical analysis using ICP-MS and the concentrations of 32 elements were measured (see Figure 2). While all soil and rock samples were analyzed for 34 and 32 elements, respectively, only specific elements including Co, Cu, Mo, Pb, and Zn are discussed here.

3.7. Spatial Dataset

Various spatial databases have been obtained from of the Geological Survey of Iran (GSI). Spatial databases include (i) locations of copper deposits/occurrences, (ii) lithologic units from the 1:250,000 scale geological/structural map of Toroud, (iii) fault/fracture lineaments digitized from the 1:250,000 scale geological/structural map of Toroud, (iv) map of hydrothermally altered rocks extracted from ETM+ multispectral satellite data, and (v) stream sediment geochemical data (samples analyzed by the ICP method) related to the TCS belt. The geochemical data represent a total drainage basin area (i.e., sampling density of one sample per 2–3 km^2). Software packages, including ArcGIS (version 10.8), and ENVI (version 5.6) were used for processing the spatial datasets in this research. Subsequently, the datasets were integrated using the RF algorithm in the R Statistical Environment (version R-4.2.2) for regional-scale prospectivity mapping of copper deposits in the TCS belt.

4. Analysis and Results

4.1. The Mineralogical and Geochemical Model

Big data consisting of the results of the analysis of lithogeochemical samples collected from copper deposits, ore occurrences, and associated lithogeochemical anomalies in Russia, Kazakhstan, and Armenia were used to construct the triple diagram with the following coordinates: Mo-Co-(Pb + Zn). On the diagram (see Figure 4), Cu-polymetallic deposits and ore occurrences are clearly distinguished (the vertex of the Pb + Zn triangle) as Cu-Mo and Cu-porphyry (Mo) and Cu-massive Sulfide and Cu-Ni (Co). Since the Clarks of the elements taken for classification are incommensurable, the components are introduced: 100 for Mo, and 10 for Pb and Co. Among Cu-Mo and Cu-porphyry forms (IV), Mo is the determining element, and molybdenite is one of the main minerals. Ores of Cu-polymetallic deposits (III), in addition to pyrite and chalcopyrite, contain galena and sphalerite, which makes it possible to attribute Pb and Zn to the elements whose increased contents determine this MGT. In the ore composition of Cu-massive sulfide (I) and Cu-Ni deposits (V), the role is played by Zn and Co, and by Co, respectively. The displacement of points 7 and 16 (see Figure 4), corresponding, respectively, to the Avangard and Kusmurun deposits (Kazakhstan) towards the vertex "Mo", is due to its increased content in oxidized ores and association with iron hydroxides and the formation of ferrimolybdite [34,35].
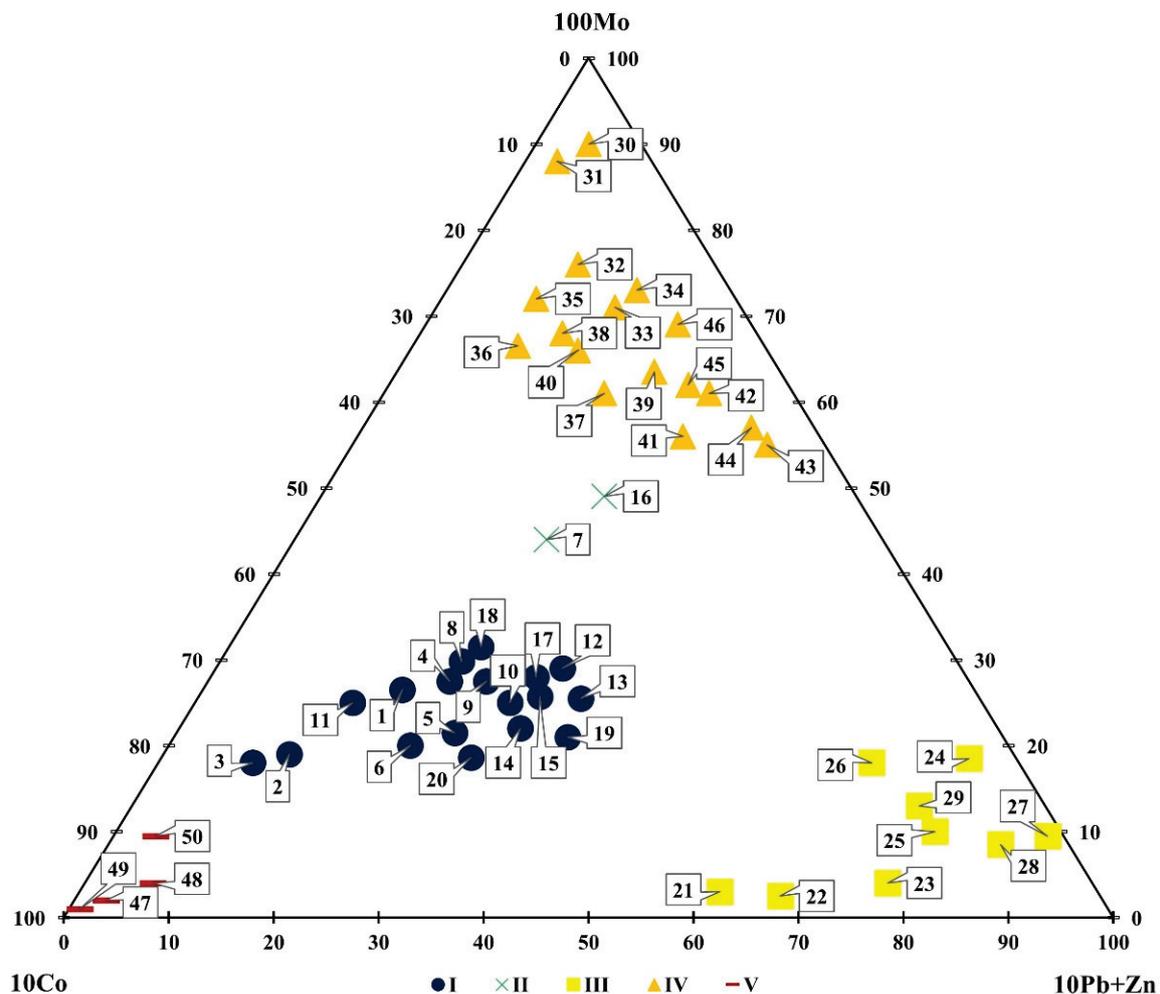
Figure 4. The 3D diagram to classify the mineralogical and geochemical types of Cu deposits, ore occurrence, and associated lithogeochemical anomalies in Russia, Kazakhstan, and Armenia. I: Cu-massive sulfide, II: oxidized Cu-massive sulfide, III: Cu-polymetallic, IV: Cu-Mo-porphyry, V: Cu-Ni deposits [34,35] (see Appendix A).

The 3D MGT model was employed to clarify the relationship between the trace element contenst of lithogeochemical samples and the MGT of geochemical anomaly. Thus, the probability of MGT mineralization of lithogeochemical samples collected from the Saghari deposit was identified. The trace element contents of samples based on the coordinates (Mo-Co-(Pb + Zn)) of the model were plotted to identify their MGT. As shown in Figure 5, the placement of the soil samples collected from the A-Central and A-South areas towards the vertex "Pb + Zn" of the triangle determines the Cu-polymetallic MGT. Also, the placement of the rock samples collected from A-Central and A-South boreholes towards the vertex "Mo" of the triangle determines the Cu-porphyry MGT. The results show a multi-MGT anomaly superposition that is a combination of two MGTs: Cu-polymetallic and Cu-porphyry.
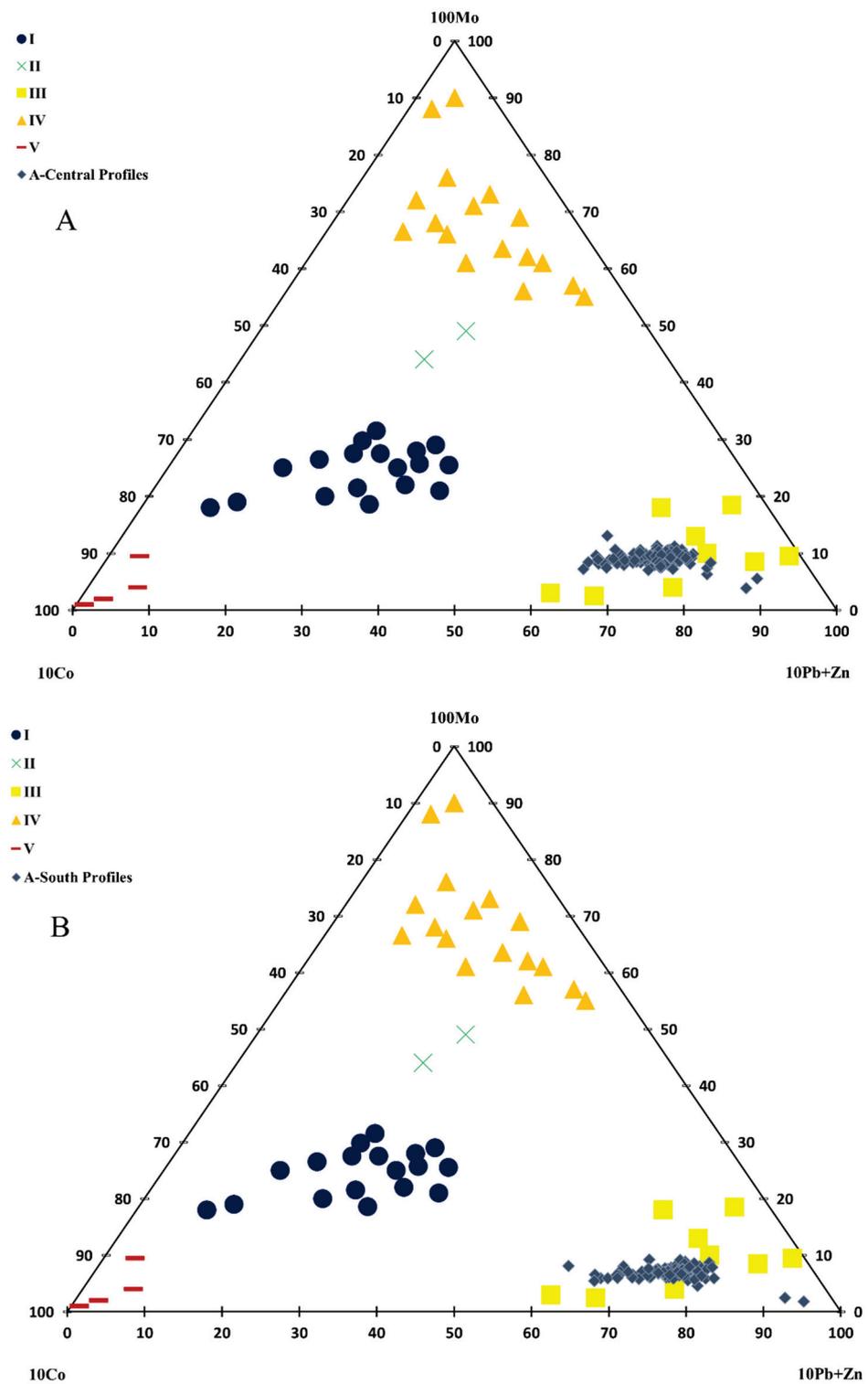
Figure 5. *Cont.*
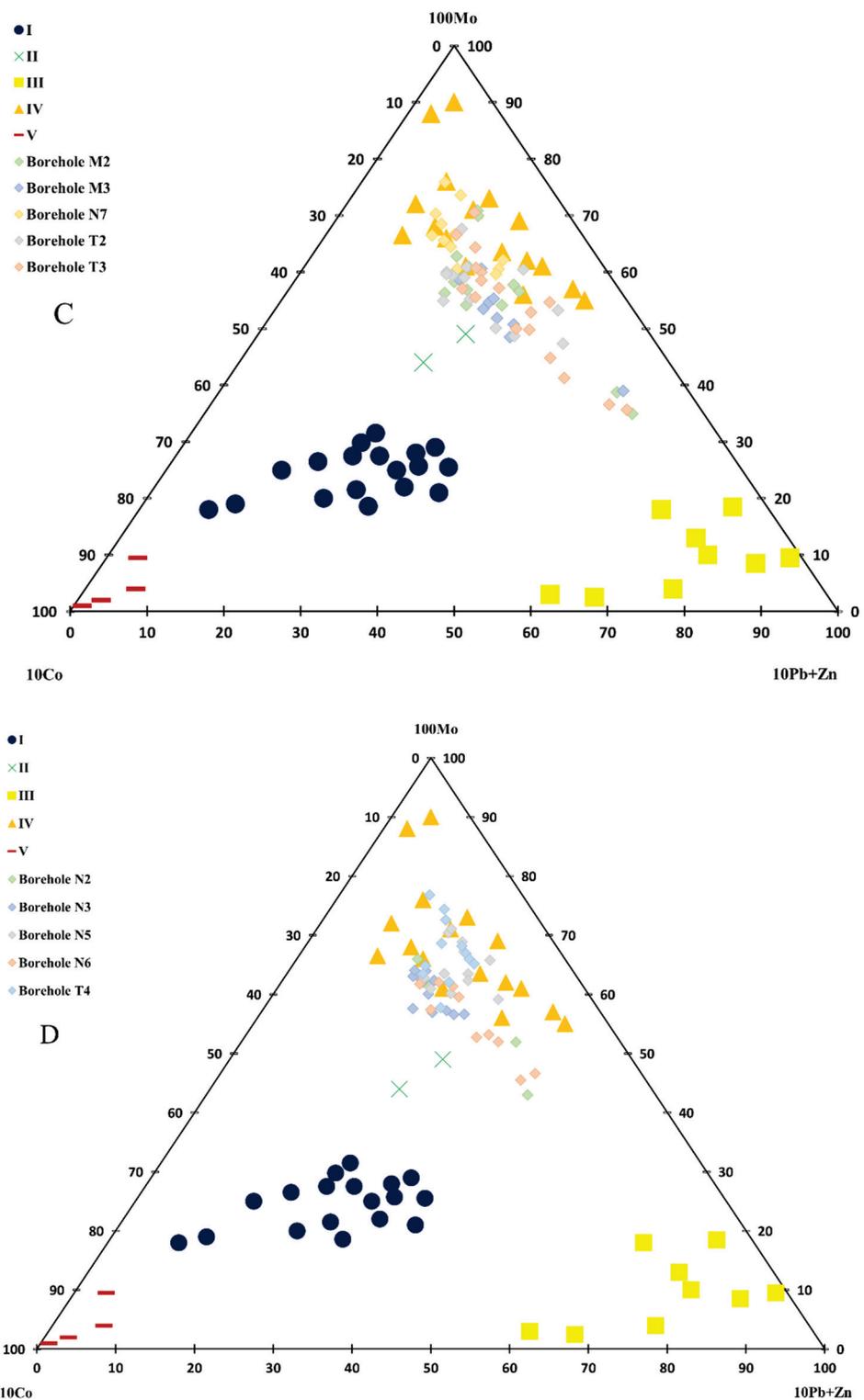
Figure 5. Distribution of lithogeochemical samples of the Saghari deposit in the 3D MGT model: (**A**) A-Central Profiles, (**B**) A-South Profiles, (**C**) A-Central Boreholes, and (**D**) A-South Boreholes.

4.2. Vertical Geochemical Zonality Model

Based on big data containing the average contents of ore elements for Aktogay (Kazakhstan), Aidarly (Kazakhstan), Kyzylkia (Kazakhstan), and Sungun (Iran) deposits, the values of the vertical geochemical zonality coefficient (Vz) were calculated and converted into the zoning sequence of geochemical haloes. They are interpolated along the averaging

straight line. Thus, the graph of the monotonous Vz was constructed in a linear scale along the ordinate axis (depth, m) and a logarithmic scale along the abscissa axis (geochemical zonality coefficient, Vz). The vertical variations in Vz (Pb × Zn/Cu × Mo) in ores and primary haloes associated with four porphyry copper deposits in areas of the same landscape–geochemical conditions in different countries (Kazakhstan, Iran) were shown in Figure 6. Despite the considerable differences in the local geological settings of individual porphyry copper deposits, falling the points on a straight line and decreasing downward the values of Vz indicate the existence of a quantitatively uniform vertical zonality in the primary haloes of the deposits [4,10]. Accordingly, vertical variations in values of Vz permit the discrepancy of mineralization levels and their primary halos (supra-ore, upper-ore, ore, lower-ore, and sub-ore) [4,35,49]. Furthermore, it can be realized from Figure 6 that similar values of Vz imply similar depths of mineralization and primary haloes. Therefore, primary halos of deposits at diverse depths are indicated by particular values of Vz. The importance of Vz is for the identification of erosional surfaces signifying vertical levels of geochemical anomalies. In other words, element data utilized as numerators of Vz signify supra-ore to ore element associations whereas those used as denominators demonstrate ore to sub-ore element associations. Concerning the present level of erosion, high values of Vz imply the presence of subcropping to blind deposits whereas low values of Vz indicate outcropping or already eroded deposits. The geochemical zonality model was used to emphasize that a recognized function (linear or nonlinear) is adequate to model the connections among the parameters of geochemical haloes and responses of the depth of mineralization [4].

The geochemical zonality model makes the erosional level assessment of any geochemical anomaly in a given MGT possible. This model was used to interpret the lithogeochemical samples in the Saghari deposit. Before the analysis of the anomalous patterns, their component anomalies should be reconstructed. These components must consider the following: (a) coexistence of two local maxima for supra-ore and sup-ore (this coexistence implies blind mineralization); (b) existence of a single component implies ZDM; (c) using mean value geochemical indicator elements, outside geochemical anomalies, for eliminating background noise in data interpretation; (d) the multiplicative geochemical anomalies and their spatial associations with particular geological features are essential aspects of mineral distribution for exploration and understanding ore geometry.

To evaluate the potential for the occurrence of blind mineralization associated with the anomalies, the values of Vz were calculated for lithogeochemical data. Based on calculations on the soil samples collected from the A-Central area and the model, blind mineralization (BM) was recognized. This anomaly has been tested by drilling five boreholes. As shown in Table 1 and Figure 7, the A-Central anomaly is promising for blind mineralization. According to the results of Vz for soil samples collected from the A-South area and the model, outcropping mineralization was identified. This anomaly has been tested by drilling seven boreholes. The obtained results of these boreholes and the model also demonstrated outcropping mineralization (Figure 7 and Table 1). Therefore, the geochemical zonality method within the A-Central and A-South areas revealed two geochemical anomalies including BM and outcropping. Furthermore, the field observations of the Saghari deposit (placed around the Chah-Musa deposit) have confirmed the results of the model.
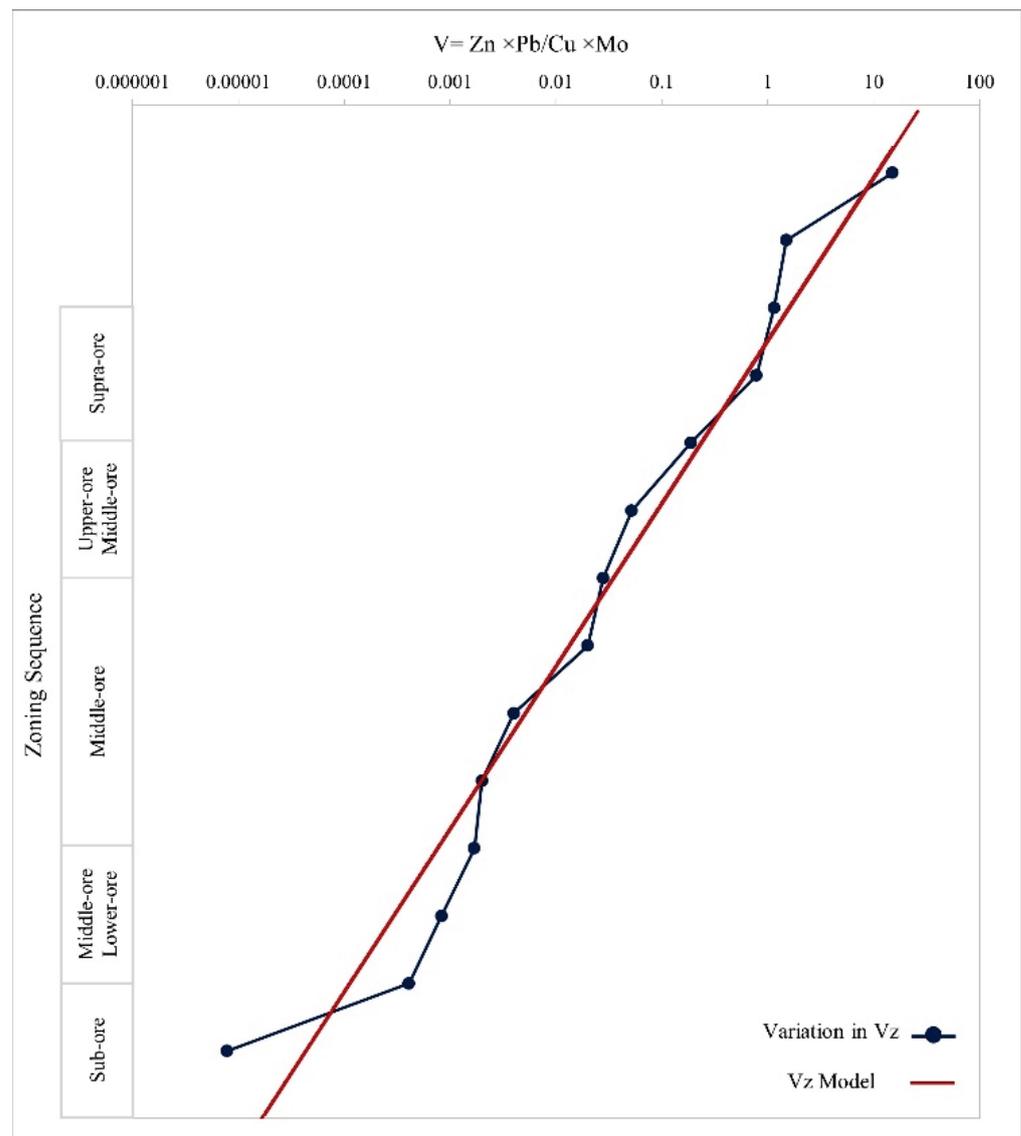
Figure 6. Vertical geochemical zonality (Vz) model for porphyry copper deposits according to typical standard porphyry copper deposits in Kazakhstan and Iran [101].

Table 1. Results of lithogeochemical data of the Saghari deposit and Vz model.

Soil Samples	Vz Index	Assessment	Rock Samples	Vz Index	Assessment
A-Central Profiles	12.32	Blind mineralization	A-Central Boreholes	7.54	BM
35T-51T	13.60	Blind mineralization	A2	7.64	BM
52T-67T	8.84	Blind mineralization	A3	30.23	BM
68T-82T	10.86	Blind mineralization	B7	4.50	BM
Three Profiles	10.73	Blind mineralization	B2	3.17	BM
			C3	10.03	BM
A-South Profiles	2.11	Outcropping mineralization	A-South Boreholes	0.42	Outcropping
33L-41L	2.13	Outcropping mineralization	B2	1.80	Outcropping
34M-41M	1.34	Outcropping mineralization	B3	0.79	Outcropping
34N-44N	2.08	Outcropping mineralization	B4	0.04	Outcropping
35O-46O	2.51	Outcropping mineralization	B5	0.11	Outcropping
34P-47P	1.85	Outcropping mineralization	B6	0.76	Outcropping
Five Profiles	1.90	Outcropping mineralization	C1	1.85	Outcropping
			C4	0.17	Outcropping
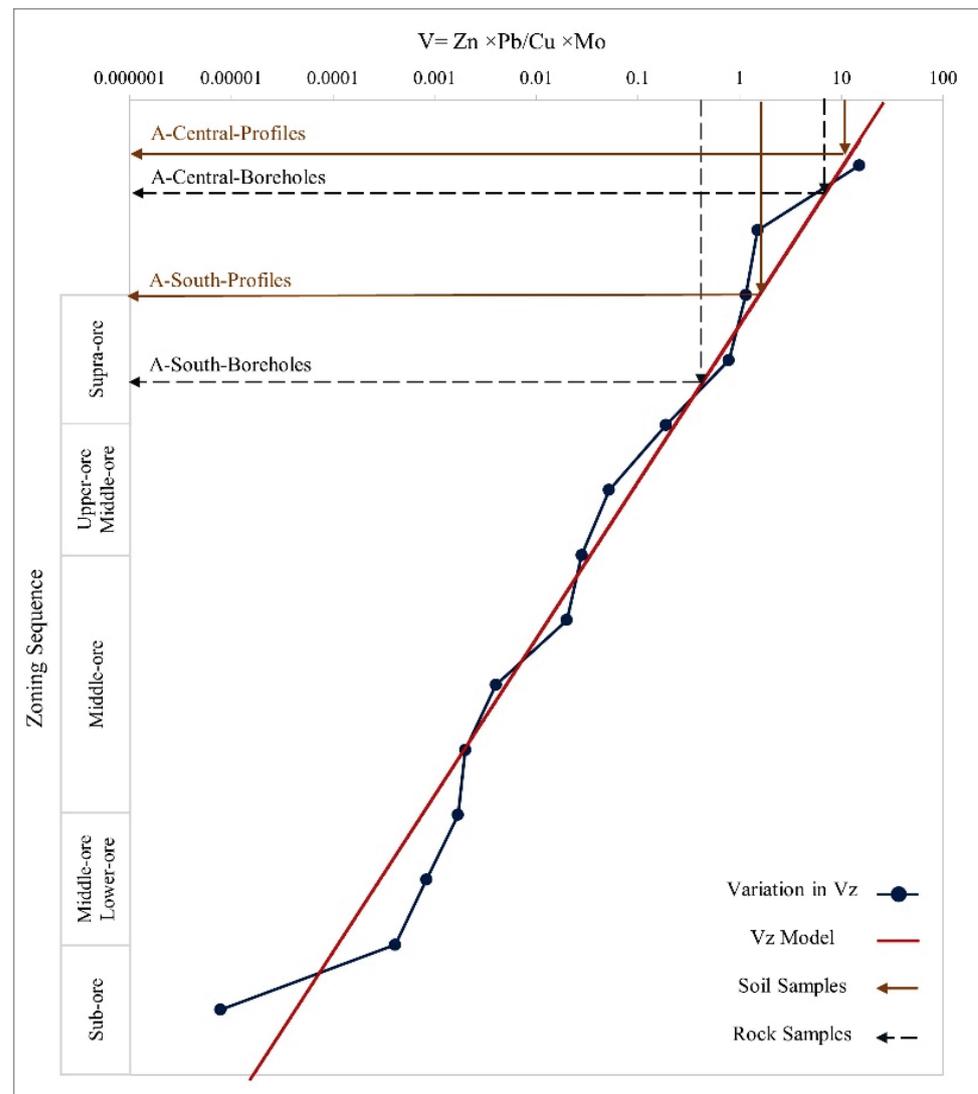
Figure 7. Plot of lithogeochemical samples of the Saghari deposit in the geochemical zonality model.

4.3. Mineral Prospectivity Mapping Model for Porphyry Copper Deposits

There is a main porphyry copper mine namely Chah-Musa deposit located close to the Saghari deposit (see Figure 1). The analyses characterized here intention to realize a provincial-scale answer to the question "Which fragments of the TCS belt have high potential for undiscovered porphyry copper deposits?". The answer will be provided according to the spatial associations of the documented porphyry copper deposits with evidential layers.

4.3.1. Spatial Evidence of Prospectivity for Porphyry Copper Deposit

The set of evidential layers includes the lithological map, the map of fault/fracture lineaments, the alteration map, and the geochemical map. The mining geochemistry-based models were used to optimize the evidential layers used in MPM. The mineralogical and geochemical types (MGTs), and geochemical zonality coefficient (Vz) of multi-elements around deposits and their spatial associations with specific geological, structural, and geochemical features are significant aspects that are valuable to consider in MPM. Based on the MGT model, the Saghari deposit was classified as a porphyry copper type; then, the alteration evidence layer was optimized. Additionally, Vz (Pb × Zn/Cu × Mo) was selected. Based on the geochemical zonality model, the Vz was used to optimize the geochemical evidence layer. Therefore, the spatial datasets utilized for predictive modeling

of prospectivity for porphyry copper deposits in TCS included a lithological map, a map of faults/fractures, the locations of porphyry copper prospects and deposits (see Figure 1), an alteration map, and a map of multi-element geochemical anomalies (Vz).

Lithological and Structural Evidence

Six lithologic units were extracted from the 1:250,000 scale geological map of the TCS range (see Figure 1). This lithological map was used as input to the RF modeling. Analysis of the mineral system associated with the porphyry copper mineralization in TCS indicates that it is worth assessing the spatial association of the target variables (deposit and non-deposit locations) in terms of distances to faults. As such, fault/fracture lineaments were extracted from the 1:250,000 scale geological map of Toroud (see Figure 1); then, distance to fault/fracture was considered to create the fault/fracture evidential map. This evidential map was used as the input for the RF modeling (Figure 8).
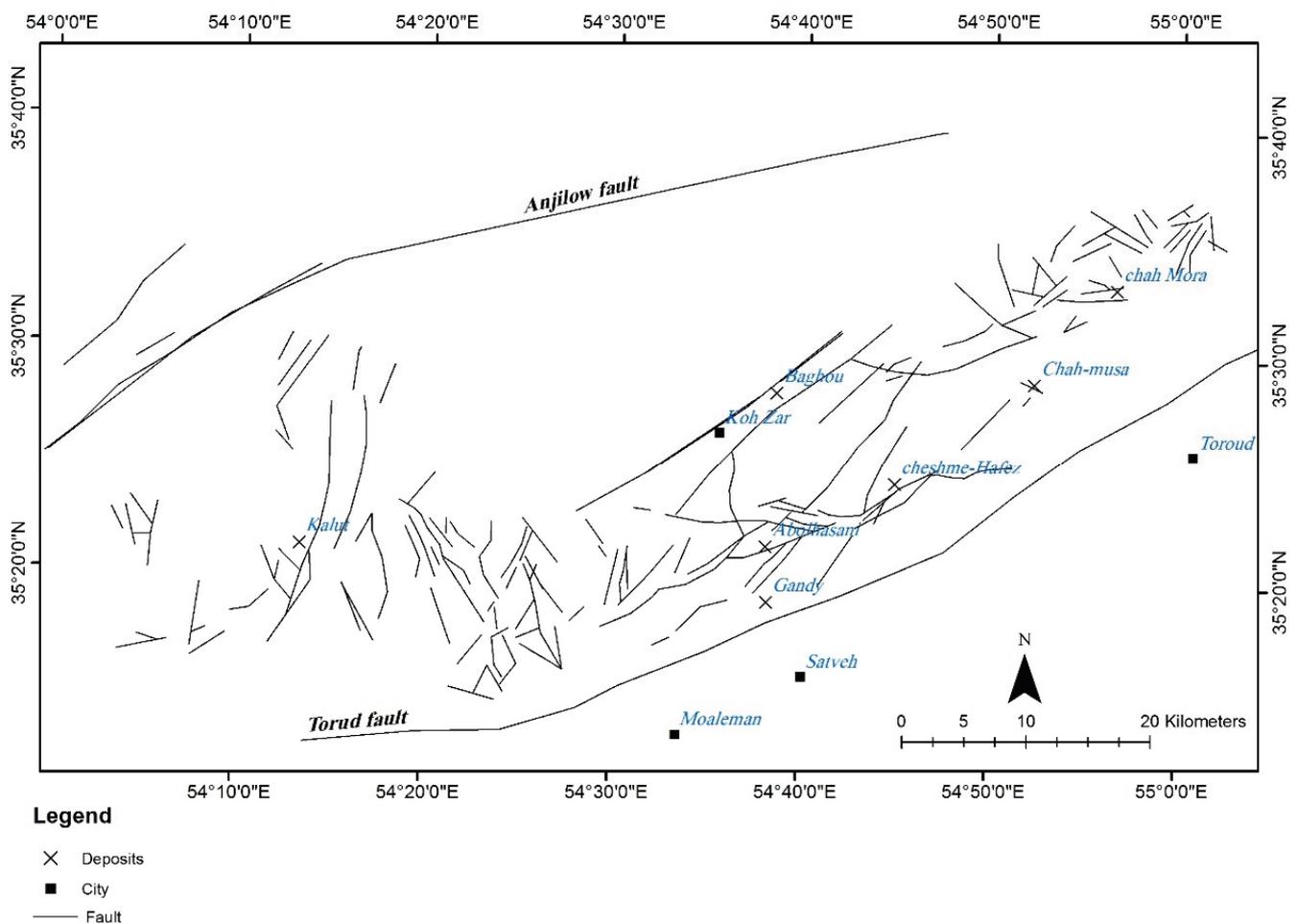
Figure 8. Structural evidential map of the TCS used for RF modeling.

Alteration Evidence

The PCA was applied to ETM+ spectral bands for extracting spectral signatures that reflect the existence of hydrothermally altered rocks associated with porphyry copper deposits in the TCS belt. The eigenvectors and eigenvalues for bands 1, 2, 3, 4, 5, and 7 of ETM+ were calculated. The eigenvector matrix of the selected ETM+ spectral bands is shown in Table 2. Several PC images were produced by running PCA. The specific PC images related to geological features were selected based on eigenvector loadings. For instance, iron oxide minerals, such as limonite, jarosite and hematite, have high reflectance within 0.63 to 0.69 μm (the equivalent to ETM+ band 3) and high absorption within 0.45

to 0.52 µm (the equivalent to ETM+ band 1). Clay minerals and carbonates show high reflectance in the range of 1.55 to 1.75 µm and high absorption in 2.08 to 2.35 µm that correspond with ETM+ bands 5 and 7, respectively [74,75,77,78]. Looking for alteration minerals, including iron oxide/hydroxides and clay minerals and carbonates, indicates that PC4 enhances the presence of iron oxide/hydroxides because it has high eigenvector loadings of different signs in bands 1 (0.401) and 3 (−0.593). Thus, the PC4 image shows iron oxide/hydroxides as dark pixels. PC5 enhances the presence of clay minerals and carbonates because it has strong eigenvector loadings of different signs in bands 5 (−0.50) and 7 (0.64). Therefore, clay minerals and carbonates are mapped as dark pixels in the PC5 image.

Table 2. Eigenvector matrix of six ETM+ spectral bands.

	PC1	PC2	PC3	PC4	PC5	PC6
Band 1	0.31	0.46	0.52	0.40	0.25	0.44
Band 2	0.35	0.37	0.21	0.02	−0.09	−0.83
Band 3	0.49	0.31	−0.19	−0.59	−0.41	0.32
Band 4	0.38	0.11	−0.77	0.39	0.32	−0.01
Band 5	0.47	−0.58	0.15	0.40	−0.50	0.03
Band 7	0.42	−0.45	0.20	−0.42	0.64	−0.04

Because of the arid and mountainous landscapes in the TCS belt, it is unlikely that ore mineralized areas are strongly indicated by the existence of secondary iron oxide/hydroxides [45]. In contrast, outcropping porphyry copper deposits are mostly associated with argillic and phyllic alteration zones [43,45]. Thus, the image of PC5, possibly reflecting the presence of hydrothermal alteration zones associated with ore minerals, can be considered as an evidential layer of porphyry copper mineralization. On the other hand, according to the spectral characteristics of clay minerals and carbonates (i.e., reflection in band 5, absorption in band 7), the image of PC5 must be negated (i.e., multiplied by −1) to manifest hydrothermally altered rocks as bright pixels (Figure 9). This image reflects the spatial distribution of argillic and phyllic alteration zones in the TCS belt. Spectral anomalies of the altered rocks appear as bright pixels in Figure 9.
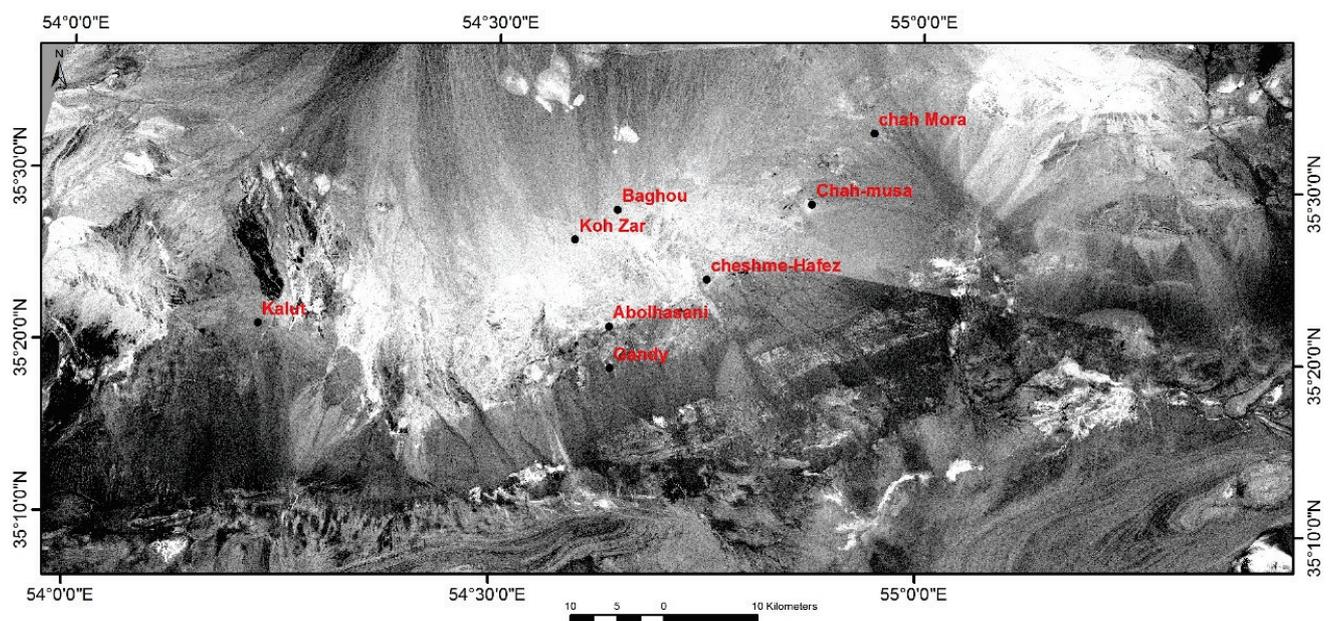
Figure 9. Negated PC5 image showing the presence of argillic and phyllic alteration zones as bright pixels in the TCS belt.

Geochemical Evidence

Multiplicative geochemical data of Pb × Zn and Cu × Mo were initially generated. The interpolated map of Vz was used as spatial evidence of porphyry copper prospectivity (Figure 10). The values in the interpolated Vz map were individually rescaled linearly to the range [0, 1]. High Vz values would suggest supra-ore to ore multielement anomalies accompanying blind to subcropping porphyry copper mineralization (Figure 10). With this hypothesis, the TCS can be segmented into three zones where possible porphyry copper deposits exist at diverse levels. Zone I can be considered as a potential zone for the exploration of subcropping to blind porphyry copper deposits, while the other zone could be a potential zone for the exploration of outcropping porphyry copper deposits.
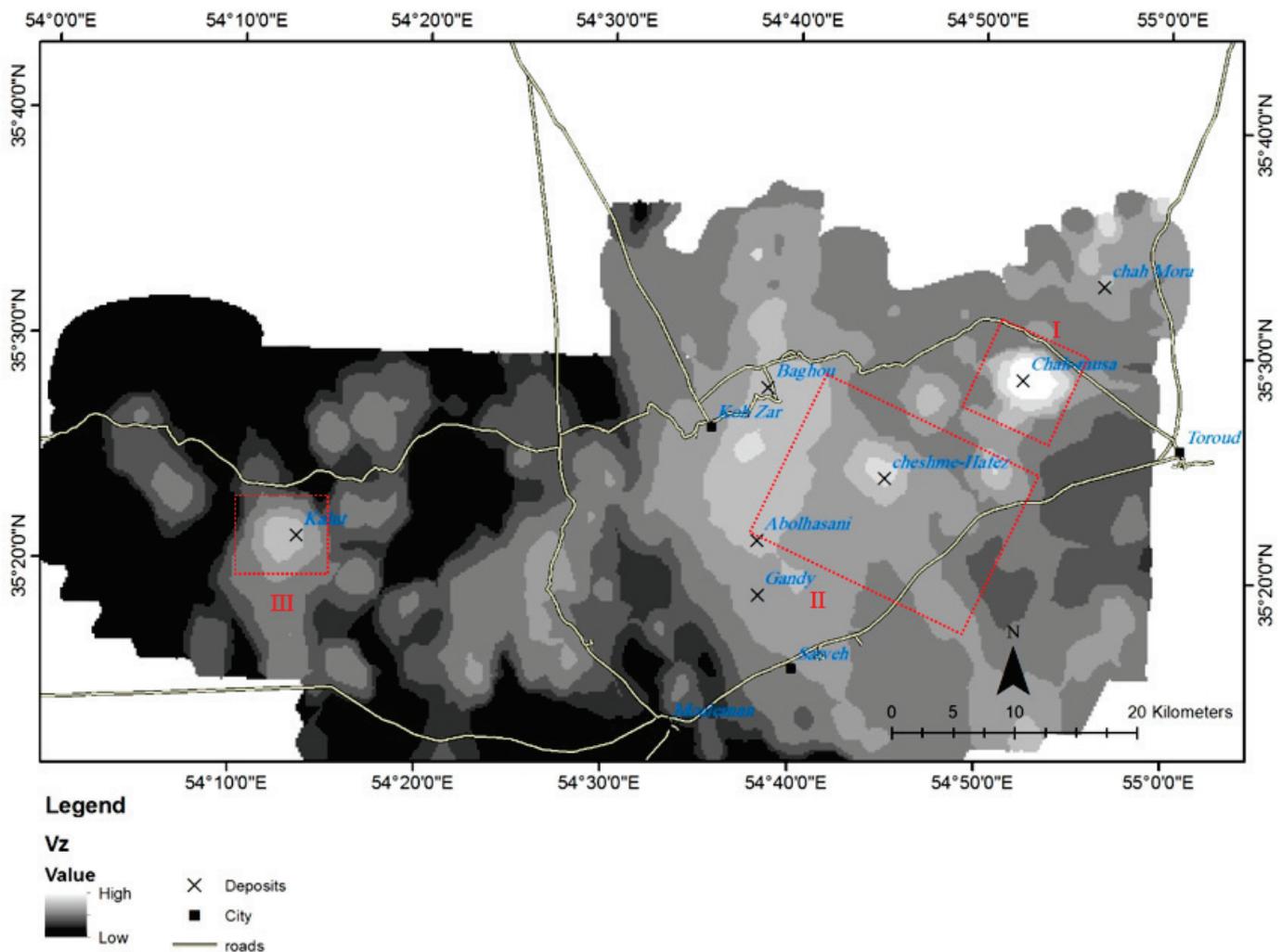
Figure 10. Geochemical evidential map of the TCS used for RF modeling. Vz (Pb × Zn/Cu × Mo) values derived from stream sediment geochemical data. Zone I: potential for the exploration of subcropping to blind porphyry copper deposits, Zone II and III: potential zone for the exploration of outcropping porphyry copper deposits.

4.3.2. Integration of Evidential Layers Using RF Algorithm

The lithology, structure, alteration, and geochemical evidential layers were integrated using the RF algorithm for regional-scale prospectivity mineral mapping of porphyry copper deposits in the TCS belt. The RF package within the R statistical environment was used for the (MGT + Vz)-in-RF modeling to calculate the spatial associations of known porphyry copper deposits in the TCS belt with individual layers of spatial evidence [102,103]. The parameters of the RF model were the number of trees (k) and the number of evidential

layers (m) that were arbitrarily sampled at each split. The m value can be experimentally determined by calculating the square root of the total number of evidential maps. Even though Breiman [98] and Liaw and Wiener [102] designated that an m value as low as 1 can yield accurate results, Gromping [104] described that the m value needs to include at least two evidential variables. Multiple experiments pointed out that the m parameter is constant with the empirical value mentioned above, and the minimum k value of 1000 produces both the lowest prediction errors and the most stable predictions. The appropriate rates of the parameters show that the RF algorithm will realize a fit among the targets (deposits and non-deposits) and evidential layers. The evidential layers input into the model can then be implemented to the model to calculate probabilities for all locations.

According to the results of RF analysis of spatial associations among the documented porphyry copper deposits and the individual layers of spatial evidence, the RF modeling generated a relationship between high-probability areas and areas showing documented porphyry copper deposits. The integration experiment was performed using lithological, structural, alteration, and Vz geochemical maps. The result of the experiment is referred to as the (MGT + Vz)-in-RF model.

Based on (MGT + Vz)-in-RF model, there seem to be four subareas of high potential zones that warrant further exploration (Figure 11). Subareas II, III, and IV could be favorable principally for exploration of outcropping porphyry copper deposits, while Subarea I could be favorable for the exploration of blind to subcropping porphyry copper deposits. Actually, prior to the organization of this this manuscript, blind porphyry copper mineralization was intersected by borehole exploration in Subarea I.
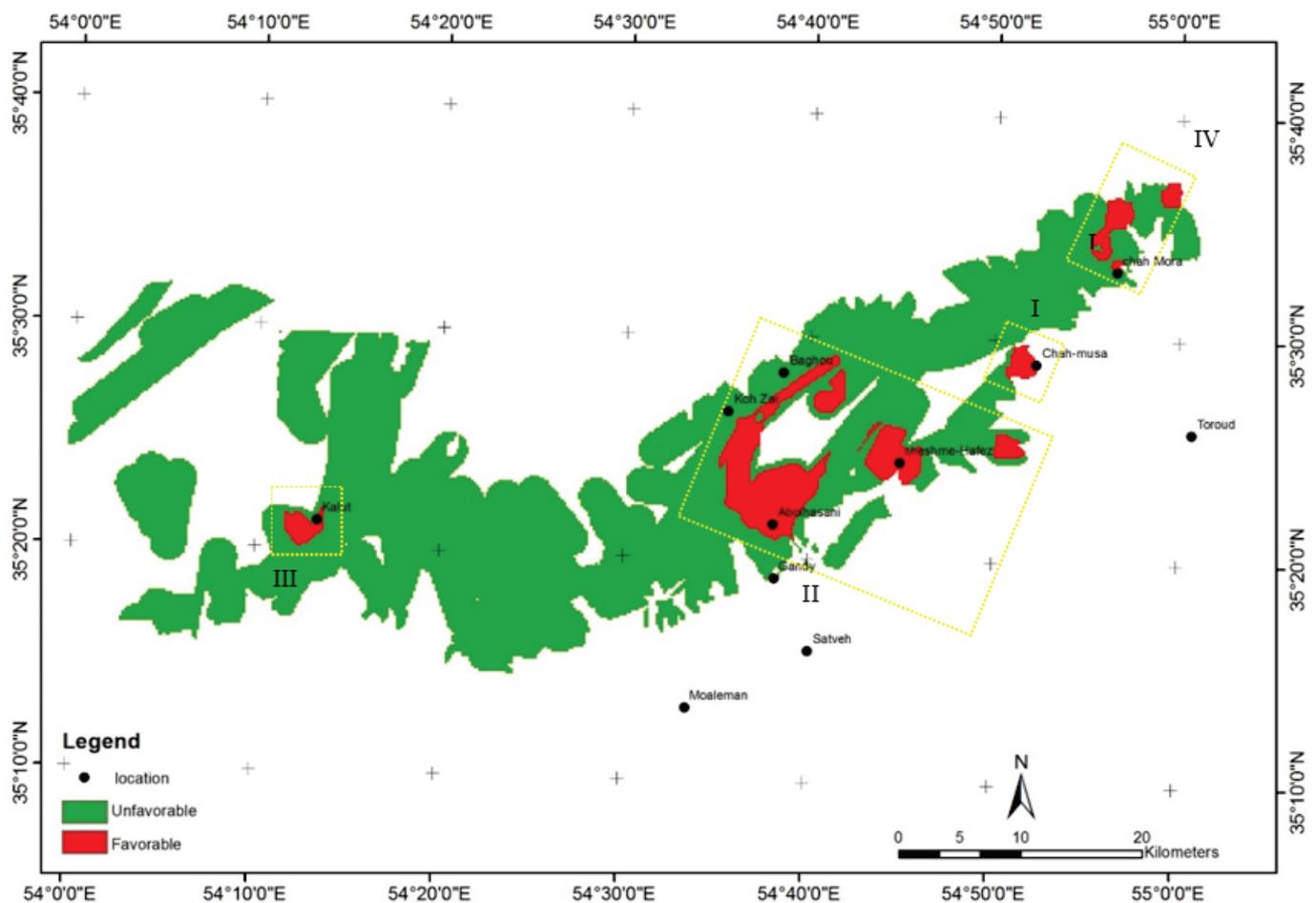

Figure 11. The (MGT + Vz)-in-RF porphyry copper prospectivity map of the TCS belt.

The RF algorithm similarly rates the significance of evidential variables using mean decrease accuracy and mean decrease Gini indices (Figure 12). The first index reflects the decline in accuracy in the whole model using out-of-bag (OOB) data, whereas the second index computes the average gain of purity by using splits of a given variable. The results of both indices indicated that the most important evidential map was Vz. The geochemical evidential layer (i.e., the Vz map), compared to the three geological evidential layers (lithology, Fault/fracture lineaments, and alteration), had stronger spatial associations with the known porphyry copper deposits in the TCS belt (Figure 12).
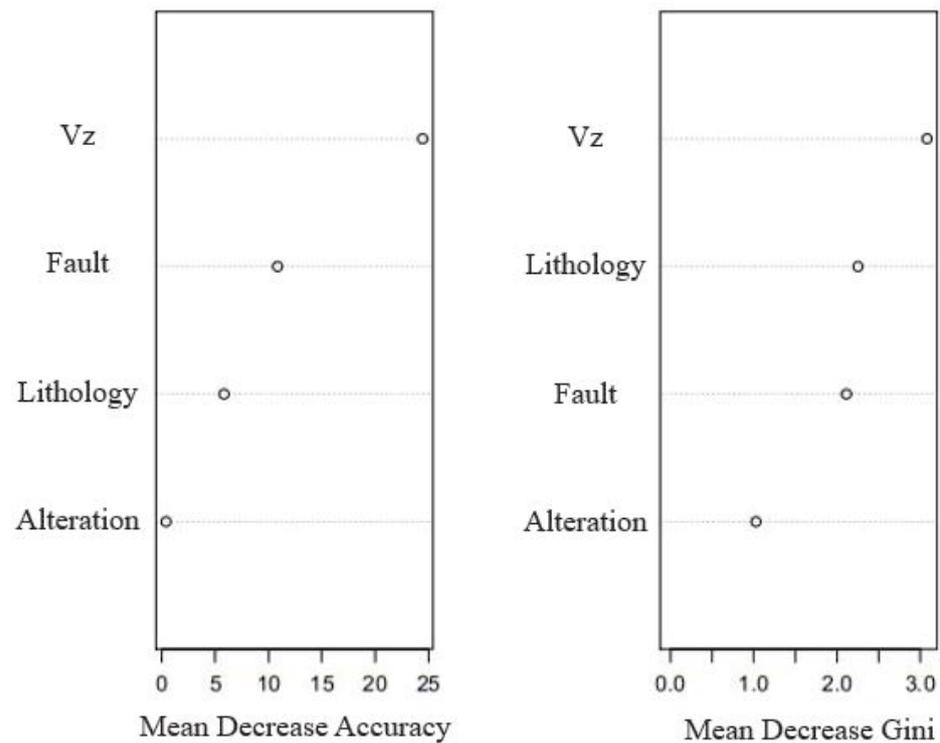

Figure 12. Relative importance of evidential variables used during (MGT + Vz)-in-RF modeling.

Although the geochemical evidential layer represents stream sediments, which are materials transported away from their sources, whereas the geological evidential layers represent in situ materials like the mineral deposits under examination, the Vz map, compared to the other geological evidential maps, provided better spatial evidence of porphyry copper prospectivity. The Vz indices are able to reflect geochemical variations in vertical directions; in fact, it might be unrelated to mineral deposits. It is mainly essential if the geochemical zonality method is executed using stream sediment geochemical data, which characterize earth materials upstream from a sampling site. Faults typically show vital roles in porphyry copper mineralization [4]. In the TCS belt, faults and fractures structurally controlled porphyry copper mineralization [61]. Based on the mean decrease accuracy index, it was the second most important evidential variable.

4.3.3. Validation

Subarea I was investigated in this research. In Subarea I, the Saghari, Fallah, and Tirgah deposits are located around the Chah-Musa deposit. The Fallah and Tirgah deposits were used to validate the presented strategy (Figure 13). The Vz values of the Tirgah and Fallah deposits are 0.11 and 24.6, respectively. The Fallah deposit is promising for blind mineralization.
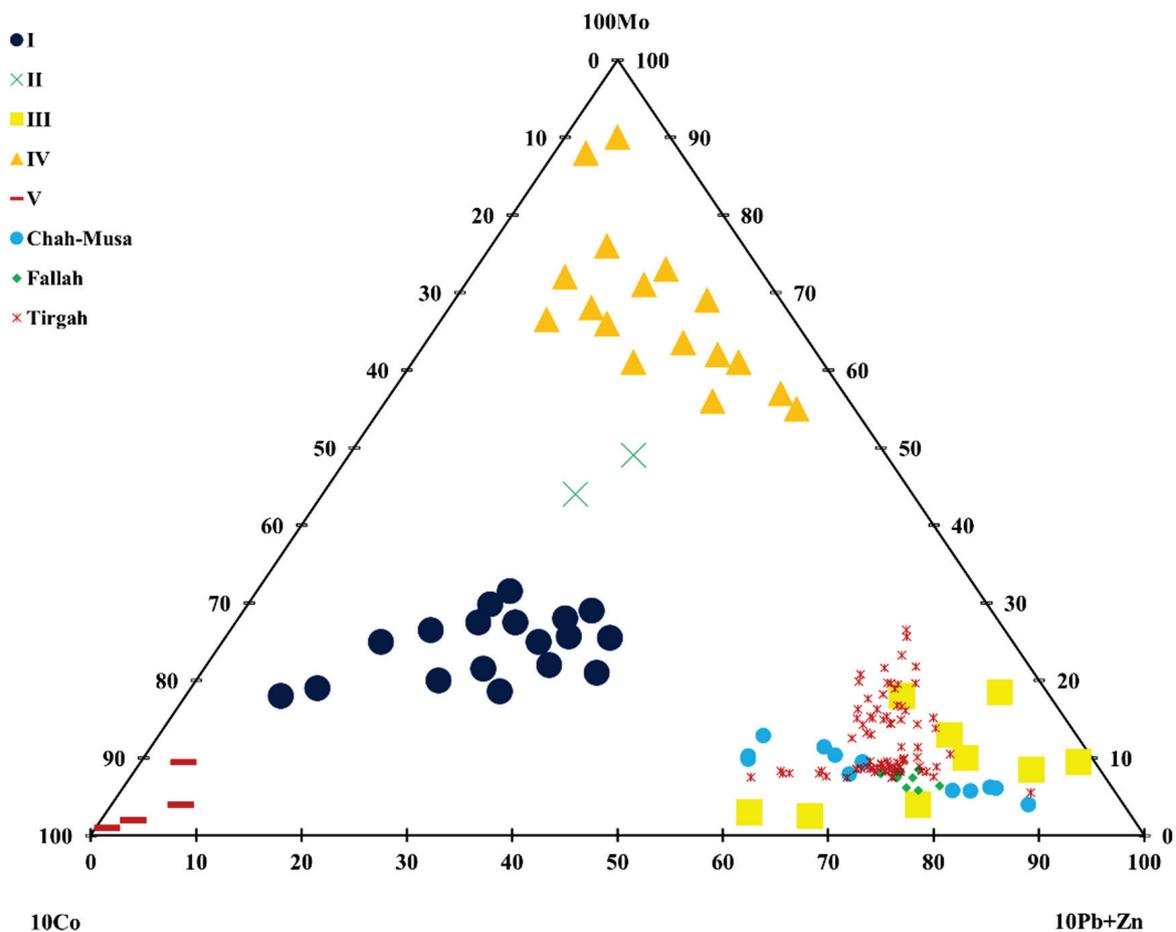
Figure 13. Distribution of lithogeochemical samples of the Fallah and Tirgah deposits in the 3D MGT model.

5. Discussion

It has been long recognized that the level of erosion, not just the ore-forming environment, influences the observed time and space distribution of metallogenic patterns [105–108]. The recent works on the TCS belt are related to depths of less than 100 m. Exploration of new concealed, deep, and economic resources is challenging because of the lack of in-depth geological information and the arid and mountainous geochemical landscapes. The Saghari deposit is located around the Chah-Musa deposit in the eastern part of the TSC belt. Geologic information about the Chah-Musa deposit is scarce despite its importance. As such, the recognition of geochemical patterns, including the spatial association of geochemical patterns, geochemical element associations, and geochemical anomalies related to mineralization is essential. This research demonstrated that the MGT and Vz models can be applied around the Chah-Musa deposit to recognize the mineralogical and geochemical types of the anomalies associated with copper mineralization, and evaluate anomalies (BM, ZDM, and outcropping) and their erosional surfaces.

The MGT model was established to discriminate five types of copper deposits based on trace elements of lithogeochemical samples using Big Data analytics extracted from 50 copper deposits in CIS countries. Supported by the testing set collected from the Saghari deposit (zone A), the performance of the MGT model shows that trace elements in lithogeochemical samples can be used to discriminate copper deposit types which cannot be distinguished from polymetal and porphyry deposits. Furthermore, these results further indicated that lithogeochemical samples of copper deposits carry unique geochemical features inherited from the ore-forming fluids. Hence, big data analytics can find some patterns to construct a well-behaved MGT model. The MGT model showed that the dis-

crimination of copper deposit types is impacted by multi-elements in lithogeochemical samples, while each element has different effects on the discrimination. According to the big data extracted, several elements contributed to the research of lithogeochemical samples, including Mo, Pb, Zn, and Co. Mo is the determining element and molybdenite is one of the main minerals among Cu-Mo and Cu porphyry deposits. In addition to pyrite and chalcopyrite, Cu-polymetallic deposits contain galena and sphalerite, which makes it possible to attribute Pb and Zn to the elements whose increased contents determine this type. Zn and Co, and Co play an important role in Cu-massive sulfide and Cu-Ni deposits, respectively. The information extracted by big data analytics not only provided interpretations for classification results but also contributed to the understanding of their geochemistry.

For the first time, using the constructed MGT model, it has been proved that the MGT type of the Saghari deposit is a multi-mineralogical and geochemical type, namely involving polymetallic and porphyry copper mineralization. Metal associations in porphyry systems not only vary as a function of emplacement depth but also as a function of porphyry intrusion composition. Through time, uplift and erosion events variably destroy and subsidence and burial events variably cover these contrasting levels of the porphyry ore-forming environment. Moreover, cover can protect a porphyry environment from destruction by subsequent erosion. Therefore, the observed metal associations and distribution of porphyry-related systems is a complex reflection of the ore-forming environment, the level of erosion, and the extent of the cover of a porphyry system.

The zonality method resolves the problems of exploration for BM and identification of ZDM. The zonality model was established using the database of copper deposits in Kazakhstan and Iran and the Vz coefficient of porphyry copper ($Pb \times Zn/Cu \times Mo$) to assess the possibilities for the occurrence of blind mineralization associated with the anomalies around the Chah-Musa deposit. The results of mining geochemistry models using lithogeochemical samples determined that the Saghari deposit belongs to the polymetallic type (on the surface, oxide part) and porphyry copper type (at a depth, sulfide part).

The different alteration zones have not been categorized based on the MGT of anomalies for MPM and cannot differentiate BM from ZDM. These mining geochemistry models are able to identify the hydrothermal alteration zones associated with porphyry copper mineralization and to evaluate deep and buried ore resources. Implementation of the Vz model in mineral prospectivity permits further clarification about whether delimited potential areas are acceptable for exploration of outcropping or blind mineral deposits because the perception of the geochemical zonality method detects the discrepancy between sub- and supra-ore anomalies. This added-value information from Vz coefficients is vital in planning exploration projects. Nevertheless, integration of maps of Vz coefficients with other maps utilized as spatial evidence is imperative in filtering out false signals of mineral prospectivity represented in every layer of evidence. Although methods for GIS-based mineral prospectivity mapping are typically well established, it is essential to analyze which techniques of geochemical data processing produce anomaly maps that lead to optimum models of mineral prospectivity mapping. Random forest outperformed the other machine learning algorithms for data integration and MPM. In this research, for the first time, geochemical zonality anomalies instead of anomalies of pathfinder elements (e.g., Cu) were used as one of several evidential maps to construct MPM based on RF for porphyry copper mineralization in the TCS belt. Based on the results of mean decrease accuracy and mean decrease Gini indices, the Vz was the most important evidential map and had stronger spatial associations with the documented porphyry copper deposits in the TCS. The result showed the applicability of the (MGT + Vz)-in-RF for regional-scale prospectivity mapping of porphyry copper deposits. Our attempts can, thus, provide some inspiration for future research to apply mining geochemistry models using big data analytics and remote sensing data to optimize the evidential variables used in MPM.

6. Conclusions

- A novel approach was accomplished in this study to optimize remote sensing-based evidential variables using mining geochemistry models for an RF-based copper mineralization prospectivity mapping (MPM).

- The MGT and Vz models were applied around the Chah-Musa deposit to identify the MGTs of the anomalies related to copper mineralization, and to evaluate anomalies (BM, ZDM, and outcropping) and the erosional surfaces. The mining geochemistry models determined that the Saghari deposit belongs to the polymetallic type (on the surface, oxide part) and porphyry copper type (at a depth, sulfide part). The possibilities for the occurrence of blind mineralization associated with the anomalies around the Chah-Musa deposit were assessed. The hydrothermal alteration zones associated with porphyry copper mineralization (argillic and phyllic alteration) were identified using PCA on ETM+ spectral bands.

- The random forest algorithm was applied to merge the evidence variables to develop a provincial-scale prospectivity map of porphyry copper deposits. Furthermore, a geochemical zonality coefficient ($Pb \times Zn/Cu \times Mo$) was used as one of several evidential maps to construct MPM based on RF for porphyry copper mineralization in the TCS belt. The result showed high applicability of the (MGT + Vz)-in-RF for mineral prospectivity mapping of porphyry copper deposits. This study explored buried copper deposits in the TCS belt through innovative approaches by integrating multi-source geoscientific datasets. The study contributes to the current efforts in searching for innovative and cost-effective methods for mineral exploration in mining sectors, and the approach can be used for MPM in metallogenic provinces around the world.

Appendix A

I: Cu-Massive Sulfide (1. "50 years of October", Southern deposit; 2. "50 years of October", Central deposit; 3. "50 years of October", Northern deposit; 4. Kizil Kibachi, Profile XIV; 5. Limannoye, Profile 60, 6. Vanguard, Profile XVIII; 8. Vanguard, oxidized ore; 9. Novo-Makanskoye, borehole 98; 10. Novo-Makanskoye, secondary scattering halo; 11. Tesiktas, ore zone I; 12. Tesiktas, ore zone III; 13. Tesiktas, ore zone IV; 14. Kusmurun, Profile III; 15. Kusmurun, Profile IV; 17. Kusmurun, oxidized ores; 18. Vlasinchikhnskoye, North, Caucasus; 19. kafan, depth 1040 m; 20. Kafan, depth 945 m);

II: Oxidized Cu-massive sulfide (7. Vanguard; 16. Kusmurun);

III: Cu-polymetallic deposits (21. Priorskoe, Profile VI; 22. Priorskoe, Profile X; 23. Sary-oba, North Mugodzhary; 24. Ayaguzskoe, East Kazakhstan; 25. Mizek, East Kazakhstan; 26. Kenkazgan; 27. "Ukolodtsa"; 28. Efimovskoe; 29. Shaumyan, Northern Armenia);

IV: Cu-Mo-porphyry deposits (30. Chatyrkul, Profile 41; 31. Chatyrkul, Profile 7; 32. Kounrad; 33. Sokurkoy; 34. Kaskyrkazgan; 35. Kenkuduk; 36. Koksay; 37. Aureolny; 38. Borly, 39. Kepcham; 40. Almaly, 41. Zhorga, 42. Stansiya, 43. Dalniy, 44. Kalmakyr, 45. Kajaran, 46. Agarak);

V: Cu-Ni deposits (47. Goryunsky, well, 11; 48. Komkor, Profile III; 49. Komkor, Profile IV; 50. Bakaiskoye, Mugodzhary) [34,35].

References

1. Beus, A.A.; Grigorian, S.V. *Geochemical Exploration Methods for Mineral Deposits*; Applied Publishing Ltd.: Wilmette, IL, USA, 1977; 287p.
2. Levinson, A.A. *Introduction to Exploration Geochemistry*; Applied Publishing: Wilmette, IL, USA, 1980; 924p.
3. Grigorian, S.V. *Secondary Lithochemical Haloes in Prospecting for Hidden Mineralization*; Nedra Publishing House: Moscow, Russia, 1985; p. 176. (In Russian)
4. Grigorian, S.V. *Mining Geochemistry*; Nedra Publishing House: Moscow, Russia, 1992; p. 310. (In Russian)
5. Chen, Y.; Zhao, P. Zonation in primary halos and geochemical prospecting pattern for the Guilaizhuang gold deposit, eastern China. *Nonrenew. Resour.* **1998**, *7*, 37–44.
6. Ziaii, M.; Pouyan, A.A.; Ziaii, M. Geochemical anomaly recognition using fuzzy C-means cluster analysis. *Wseas Trans. Syst.* **2006**, *5*, 2424–2430.
7. Cheng, Q. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. *Ore Geol. Rev.* **2007**, *32*, 314–324. [CrossRef]
8. Chen, Y.; Huang, J.; Liang, Z. Geochemical characteristics and zonation of primary halos of Pulang porphyry copper deposit, Northwestern Yunnan Province, Southwestern China. *J. China Univ. Geosci.* **2008**, *19*, 371–377.
9. Carranza, E.J.M.; Owusu, E.A.; Hale, M. Mapping of prospectivity and estimation of number of undiscovered prospects for lode gold, southwestern Ashanti Belt, Ghana. *Miner. Depos.* **2009**, *44*, 915–938. [CrossRef]
10. Ziaii, M.; Pouyan, A.A.; Ziaei, M. Neuro-fuzzy modelling in mining geochemistry: Identification of geochemical anomalies. *J. Geochem. Explor.* **2009**, *100*, 25–36. [CrossRef]
11. Zuo, R.; Cheng, Q.; Agterberg, F.; Xia, Q. Application of singularity mapping technique to identify local anomalies using stream sediment geochemical data, a case study from Gangdese, Tibet, western China. *J. Geochem. Explor.* **2009**, *101*, 225–235. [CrossRef]
12. Sun, X.; Gong, Q.; Wang, Q.; Yang, L.; Wang, C.; Wang, Z. Application of local singularity model to delineate geochemical anomalies in Xiong'ershan gold and molybdenum ore district, Western Henan province, China. *J. Geochem. Explor.* **2010**, *107*, 21–29. [CrossRef]
13. Hamedani, M.L.; Plimer, I.R.; Xu, C. Orebody modelling for exploration: The Western mineralization, Broken Hill, NSW. *Nat. Resour. Res.* **2012**, *21*, 325–345. [CrossRef]
14. Cheng, Q. Singularity theory and methods for mapping geochemical anomalies caused by buried sources and for predicting undiscovered mineral deposits in covered areas. *J. Geochem. Explor.* **2012**, *122*, 55–70. [CrossRef]
15. Ziaii, M.; Ardejani, F.D.; Ziaei, M.; Soleymani, A.A. Neuro-fuzzy modeling based genetic algorithms for identification of geochemical anomalies in mining geochemistry. *Appl. Geochem.* **2012**, *27*, 663–676. [CrossRef]
16. Zuo, R.; Xia, Q.; Zhang, D. A comparison study of the C–A and S–A models with singularity analysis to identify geochemical anomalies in covered areas. *Appl. Geochem.* **2013**, *33*, 165–172. [CrossRef]
17. Safari, S.; Ziaii, M.; Ghoorchi, M. Integration of singularity and zonality methods for prospectivity map of blind mineralization. *Int. J. Min. Geo-Eng.* **2016**, *50*, 189–194.
18. Safari Sinegani, S.; Ziaii, M.; Ghoorchi, M.; Sadeghi, M. Application of concentration gradient coefficients in mining geochemistry: A comparison of copper mineralization in Iran and Canada. *J. Min. Environ.* **2018**, *9*, 277–292.
19. Ziaii, M.; Safari, S.; Timkin, T.; Voroshilov, V.; Yakich, T. Identification of geochemical anomalies of the porphyry–Cu deposits using concentration gradient modelling: A case study, Jebal-Barez area, Iran. *J. Geochem. Explor.* **2019**, *199*, 16–30. [CrossRef]
20. Safari, S.; Ziaii, M. Singularity of zonality, the indicator for quantitative evaluation of anomalies in mining geochemistry, Case study: Kerver. *J. Min. Eng.* **2019**, *14*, 76–91.
21. Macheyeki, A.S.; Li, X.; Kafumu, D.P.; Yuan, F. *Applied Geochemistry: Advances in Mineral Exploration Techniques*; Elsevier: Amsterdam, The Netherlands, 2020.
22. Timkin, T.; Abedini, M.; Ziaii, M.; Ghasemi, M.R. Geochemical and Hydrothermal Alteration Patterns of the Abrisham-Rud Porphyry Copper District, Semnan Province, Iran. *Minerals* **2022**, *12*, 103. [CrossRef]
23. Sun, W.; Zheng, Y.; Wang, W.; Feng, X.; Zhu, X.; Zhang, Z.; Hou, H.; Ge, L.; Lv, H. Geochemical Characteristics of Primary Halos and Prospecting Significance of the Qulong Porphyry Copper–Molybdenum Deposit in Tibet. *Minerals* **2023**, *13*, 333. [CrossRef]
24. Ziaii, M.; Carranza, E.J.M.; Ziaei, M. Application of geochemical zonality coefficients in mineral prospectivity mapping. *Comput. Geosci.* **2011**, *37*, 1935–1945. [CrossRef]
25. Carranza, E.J.M.; Sadeghi, M. Predictive mapping of prospectivity and quantitative estimation of undiscovered VMS deposits in Skellefte district (Sweden). *Ore Geol. Rev.* **2010**, *38*, 219–241. [CrossRef]

26. Carranza, E.J.M. From predictive mapping of mineral prospectivity to quantitative estimation of number of undiscovered prospects. *Resour. Geol.* **2010**, *61*, 30–51. [CrossRef]

27. Yousefi, M.; Kamkar-Rouhani, A.; Carranza, E.J.M. Geochemical mineralization probability index (GMPI): A new approach to generate enhanced stream sediment geochemical evidential map for increasing probability of success in mineral potential mapping. *J. Geochem. Explor.* **2012**, *115*, 24–35. [CrossRef]

28. Yousefi, M.; Kamkar-Rouhani, A.; Carranza, E.J.M. Application of staged factor analysis and logistic function to create a fuzzy stream sediment geochemical evidence layer for mineral prospectivity mapping. *Geochem. Explor. Environ. Anal.* **2014**, *14*, 45–58. [CrossRef]

29. Chattoraj, S.L.; Prasad, G.; Sharma, R.U.; van der Meer, F.D.; Guha, A.; Pour, A.B. Integration of remote sensing, gravity and geochemical data for exploration of Cu-mineralization in Alwar basin, Rajasthan, India. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *91*, 102162. [CrossRef]

30. Bai, S.; Zhao, J. A New Strategy to Fuse Remote Sensing Data and Geochemical Data with Different Machine Learning Methods. *Remote Sens.* **2023**, *15*, 930. [CrossRef]

31. Zuo, R.; Xiong, Y. Big data analytics of identifying geochemical anomalies supported by machine learning methods. *Nat. Resour. Res.* **2018**, *27*, 5–13. [CrossRef]

32. Xiong, Y.; Zuo, R.; Carranza, E.J.M. Mapping mineral prospectivity through big data analytics and a deep learning algorithm. *Ore Geol. Rev.* **2018**, *102*, 811–817. [CrossRef]

33. Liu, Y.; Carranza, E.J.M.; Xia, Q. Developments in Quantitative Assessment and Modeling of Mineral Resource Potential: An Overview. *Nat. Resour. Res.* **2022**, *31*, 1825–1840. [CrossRef]

34. Matveev, A.A.; Solovov, A.P. *Geochemical Methods of Prospecting for Ore Deposits*; Moscow State University Publishing House: Moscow, Russia, 1985; 232p. (In Russian)

35. Arkhipov, A.Y.; Bugrov, V.A.; Vorobyov, S.A.; Gershman, D.M.; Grigorian, S.V.; Kiyatovsky, E.M.; Matveev, A.A.; Milyaev, S.A.; Nikolaev, V.A.; Perelman, A.I.; et al. *Handbook of Geochemical Prospecting for Minerals*; Nedra: Moscow, Russia, 1990. (In Russian)

36. Ziaii, M. Method of Rational Mineralogical and Geochemical Sampling of Gold Ore Occurrences. Ph.D. Thesis, Russian Academy of Science (IGEM RAN), Moscow, Russia, 1999. (In Russian)

37. Garigorin, S.V.; Liakhovich, T.T.; Getmansky, I.I.; Ziaii, M. Trace elements in minerals as a criterion of geochemical anomaly estimations. *J. Sci. Technol.* **1999**, *1*, 22–26. (In Russian with English Abstract)

38. Grigorian, S.V.; Liakhovich, T.T.; Getmansky, I.I.; Ziaii, M. Geochemical spectrum of minerals as a criterion of gold ores type identification. *J. Sci. Technol.* **1999**, *3*, 5–8. (In Russian with English Abstract)

39. Ziaii, M.; Abedi, A.; Ziaei, M. Geochemical and mineralogical pattern recognition and modeling with a Bayesian approach to hydrothermal gold deposits. *Appl. Geochem.* **2009**, *24*, 1142–1146. [CrossRef]

40. Chen, Z.; Xiong, Y.; Yin, B.; Sun, S.; Zuo, R. Recognizing geochemical patterns related to mineralization using a self-organizing map. *Appl. Geochem.* **2023**, *151*, 105621. [CrossRef]

41. Kazemi, H. Exploring Promising Areas for Copper Deposits in the Trood Area—Using Remote Sensing and GIS. Master's Thesis, Shahrood University of Technology, Shahrood, Iran, 2007.

42. Ziaii, M.; Abedi, A.; Ziaei, M.; Kamkar Rouhani, A.; Zendahdel, A. GIS modelling for Au-Pb-Zn potential mapping in Torud-Chah Shirin area-Iran. *J. Min. Environ.* **2010**, *1*, 17–27.

43. Noori, L.; Pour, A.B.; Askari, G.; Taghipour, N.; Pradhan, B.; Lee, C.W.; Honarmand, M. Comparison of different algorithms to map hydrothermal alteration zones using ASTER remote sensing data for polymetallic vein-type ore exploration: Toroud–Chahshirin Magmatic Belt (TCMB), North Iran. *Remote Sens.* **2019**, *11*, 495. [CrossRef]

44. Zamyad, M.; Afzal, P.; Pourkermani, M.; Nouri, R.; Jafari, M.R. Determination of hydrothermal alteration zones using remote sensing methods in Tirka Area, Toroud, NE Iran. *J. Indian Soc. Remote Sens.* **2019**, *47*, 1817–1830. [CrossRef]

45. Shirmard, H.; Farahbakhsh, E.; Beiranvand Pour, A.; Muslim, A.M.; Müller, R.D.; Chandra, R. Integration of selective dimensionality reduction techniques for mineral exploration using ASTER satellite data. *Remote Sens.* **2020**, *12*, 1261. [CrossRef]

46. Pour, A.B.; Zoheir, B.; Pradhan, B.; Hashim, M. Editorial for the special issue: Multispectral and hyperspectral remote sensing data for mineral exploration and environmental monitoring of mined areas. *Remote Sens.* **2021**, *13*, 519. [CrossRef]

47. Mamouch, Y.; Attou, A.; Miftah, A.; Ouchchen, M.; Dadi, B.; Achkouch, L.; Et-tayea, Y.; Allaoui, A.; Boualoul, M.; Randazzo, G.; et al. Mapping of Hydrothermal Alteration Zones in the Kelâat M'Gouna Region Using Airborne Gamma-Ray Spectrometry and Remote Sensing Data: Mining Implications (Eastern Anti-Atlas, Morocco). *Appl. Sci.* **2022**, *12*, 957. [CrossRef]

48. Yousufi, A.; Ahmadi, H.; Bekbotayeva, A.; Arshamov, Y.; Baisalova, A.; Omarova, G.; Pekkan, E. Integration of Remote Sensing and Field Data in Ophiolite Investigations: A Case Study of Logar Ophiolite Complex, SE Afghanistan. *Minerals* **2023**, *13*, 234. [CrossRef]

49. Solovov, A.P. *Geochemical Prospecting for Mineral Deposits*; Englsih Edition; Mir Publication: Moscow, Russia, 1987; 287p.

50. Moon, C.J. Towards a quantitative model of downstream dilution of point source geochemical anomalies. *J. Geochem. Explor.* **1999**, *65*, 111–132. [CrossRef]

51. Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M.J.O.G.R. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* **2015**, *71*, 804–818. [CrossRef]

52. Carranza, E.J.M.; Laborte, A.G. Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of Random Forests algorithm. *Ore Geol. Rev.* **2015**, *71*, 777–787. [CrossRef]

53. Carranza, E.J.M.; Laborte, A.G. Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Comput. Geosci.* **2015**, *74*, 60–70. [CrossRef]

54. Carranza, E.J.M.; Laborte, A.G. Data-driven predictive modeling of mineral prospectivity using random forests: A case study in Catanduanes Island (Philippines). *Nat. Resour. Res.* **2016**, *25*, 35–50. [CrossRef]

55. Zhang, S.; Xiao, K.; Carranza, E.J.M.; Yang, F. Maximum entropy and random forest modeling of mineral potential: Analysis of gold prospectivity in the Hezuo–Meiwu district, west Qinling Orogen, China. *Nat. Resour. Res.* **2019**, *28*, 645–664. [CrossRef]

56. Wang, C.; Pan, Y.; Chen, J.; Ouyang, Y.; Rao, J.; Jiang, Q. Indicator element selection and geochemical anomaly mapping using recursive feature elimination and random forest methods in the Jingdezhen region of Jiangxi Province, South China. *Appl. Geochem.* **2020**, *122*, 104760. [CrossRef]

57. Zhang, S.; Carranza, E.J.M.; Xiao, K.; Wei, H.; Yang, F.; Chen, Z.; Li, N.; Xiang, J. Mineral prospectivity mapping based on isolation forest and random forest: Implication for the existence of spatial signature of mineralization in outliers. *Nat. Resour. Res.* **2022**, *31*, 1981–1999. [CrossRef]

58. Alavi, M. Tectonostratigraphic synthesis and structural style of the Alborz Mountain system in northern Iran. *J. Geodyn.* **1996**, *21*, 1–33. [CrossRef]

59. Shamanian, G.H.; Hedenquist, J.W.; Hattori, K.H.; Hassanzadeh, J. The Gandy and Abolhassani epithermal prospects in the Alborz magmatic arc, Semnan province, Northern Iran. *Econ. Geol.* **2004**, *99*, 691–712. [CrossRef]

60. TaleFazel, E.; Mehrabi, B.; Siani, M.G. Epithermal systems of the Torud–Chah Shirin district, northern Iran: Ore-fluid evolution and geodynamic setting. *Ore Geol. Rev.* **2019**, *109*, 253–275. [CrossRef]

61. Mehrabi, B.; Ghasemi, S.M.; Tale, F.E. Structural control on epithermal mineralization in the Troud-Chah Shirin belt using point pattern and Fry analyses, north of Iran. *Geotectonics* **2015**, *49*, 320–331. [CrossRef]

62. Imamjomeh, A.; Rastad, E.; Bouzari, F.; Omran, N.R. An introduction to individual disseminated-veinlet and vein mineralization system of Cu (Pb-Zn) in the Chahmoosa-Gholekaftaran mining district, eastern part of Toroud-Chahshirin magmatic arc. *Sci. Q. J. Geosci.* **2008**, *70*, 112–125.

63. Seifivand, A.; Sheibi, M. Ballooning emplacement and alteration of the Chah-Musa subvolcanic intrusion (NE Iran) inferred from magnetic susceptibility and fabric. *Geol. Mag.* **2020**, *157*, 621–639. [CrossRef]

64. Khalaj, M.; Hassan-Nezhad, A.; Alizadeh, H.; Babaei, A.H.; Ghorbani, G. Investigation of mineralization fluid evolution of hydrothermal vein copper deposits: Based on studies of fluid inclusions at Chah Mousa area (north of central Iran). *Adv. Appl. Geol.* **2021**, *11*, 116–135.

65. Bugrov, V. Geochemical sampling techniques in the Eastern Desert of Egypt. *J. Geochem. Explor.* **1974**, *3*, 67–75. [CrossRef]

66. Perelman, A.I. *Landscape Geochemistry*; Higher School: Moscow, Russia, 1975; 341p. (In Russian)

67. Li, Q.; Xi, X.; Ren, T.; Zhao, Y.; Che, G. Geochemical exploration in the semiarid steppe terrain of eastern Inner Mongolia—A case history. *J. Geochem. Explor.* **1989**, *33*, 27–46.

68. Houshmandzadeh, A.R.; Alavi, M.N.; Haghipour, A.A. *Evolution of Geological Phenomenon in Toroud Area (Precambrian to Recent)*; Report H5; Geological Survey & Mineral Exploration of Iran: Tehran, Iran, 1978.

69. Geological Survey of Iran (GSI). *Geological Map of Toroud*; 1:250,000 Series; Geological Survey of Iran (GSI): Tehran, Iran, 1978.

70. Kitaev, N.A. Multidimensional analysis of geochemical fields. *Math. Geol.* **1991**, *23*, 15–32. [CrossRef]

71. Baranov, E.V. *Endogenetic Haloes Associated with Massive Sulphide Deposits*; Nedra Publishing House: Moscow, Russia, 1987; p. 294. (In Russian)

72. Safonov, Y.G. Hydrothermal gold deposits: Distribution, geological-genetic types, and productivity of ore-forming systems. *Geol. Ore Depos.* **1997**, *30*, 20–32.

73. Mars, J.C.; Rowan, L.C. Regional mapping of phyllic- and argillic-altered rocks in the Zagros magmatic arc, Iran, using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data and logical operator algorithms. *Geosphere* **2006**, *2*, 161–186. [CrossRef]

74. Di Tommaso, I.; Rubinstein, N. Hydrothermal alteration mapping using ASTER data in the Infiernillo porphyry deposit, Argentina. *Ore Geol. Rev.* **2007**, *32*, 275–290. [CrossRef]

75. Pour, B.A.; Hashim, M. The application of ASTER remote sensing data to porphyry copper and epithermal gold deposits. *Ore Geol. Rev.* **2012**, *44*, 1–9. [CrossRef]

76. Lowell, J.D.; Guilbert, J.M. Lateral and vertical alteration-mineralization zoning in porphyry ore deposits. *Econ. Geol. Bull. Soc. Econ. Geol.* **1970**, *65*, 373–408. [CrossRef]

77. Hunt, G.R.; Ashley, P. Spectra of altered rocks in the visible and near infrared. *Econ. Geol.* **1979**, *74*, 1613–1629. [CrossRef]

78. Crowley, J.K.; Vergo, N. Near-infrared reflectance spectra of mixtures of kaolin group minerals: Use in clay mineral studies. *Clays Clay Miner.* **1988**, *36*, 310–316. [CrossRef]

79. Pour, B.A.; Hashim, M. Identification of hydrothermal alteration minerals for exploring of porphyry copper deposit using ASTER data, SE Iran. *J. Asian Earth Sci.* **2011**, *42*, 1309–1323. [CrossRef]

80. Pour, B.A.; Hashim, M. Identifying areas of high economic-potential copper mineralization using ASTER data in Urumieh-Dokhtar Volcanic Belt, Iran. *Adv. Space Res.* **2012**, *49*, 753–769. [CrossRef]

81. Hu, B.; Xu, Y.; Wan, B.; Wu, X.; Yi, G. Hydrothermally altered mineral mapping using synthetic application of Sentinel-2A MSI, ASTER and Hyperion data in the Duolong area, Tibetan Plateau, China. *Ore Geol. Rev.* **2018**, *101*, 384–397. [CrossRef]

82. Pour, A.B.; Hashim, M.; Hong, J.K.; Park, Y. Lithological and alteration mineral mapping in poorly exposed lithologies using Landsat-8 and ASTER satellite data: North-eastern Graham Land, Antarctic Peninsula. *Ore Geol. Rev.* **2019**, *108*, 112–133. [CrossRef]

83. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [CrossRef]

84. Jolliffe, I.T. *Principal Component Analysis for Special Types of Data*, 2nd ed.; Springer: New York, NY, USA, 2002.

85. Yang, J.; Cheng, Q. A comparative study of independent component analysis with principal component analysis in geological objects identification, Part I: Simulations. *J. Geochem. Explor.* **2015**, *149*, 127–135. [CrossRef]

86. Prado, I.M.; Crosta, A.P. Evaluating Geoscan AMSS Mk-II for gold exploration in the Fazenda Maria Preta District, Rio Itapicuru Greenstone Belt, Bahia State, Brazil. Boletim IG-USP. *Série Científica* **1997**, *28*, 63–83. [CrossRef]

87. Pour, A.B.; Park, T.S.; Park, Y.; Hong, J.K.; Zoheir, B.; Pradhan, B.; Ayoobi, I.; Hashim, M. Application of Multi-Sensor Satellite Data for Exploration of Zn-Pb Sulfide Mineralization in the Franklinian Basin, North Greenland. *Remote Sens.* **2018**, *10*, 1186. [CrossRef]

88. Pour, A.B.; Park, T.-Y.S.; Park, Y.; Hong, J.K.; Muslim, A.M.; Läufer, A.; Crispini, L.; Pradhan, B.; Zoheir, B.; Rahmani, O.; et al. Landsat-8, advanced spaceborne thermal emission and reflection radiometer, and WorldView-3 multispectral satellite imagery for prospecting copper-gold mineralization in the northeastern Inglefield Mobile Belt (IMB), northwest Greenland. *Remote Sens.* **2019**, *11*, 2430. [CrossRef]

89. Pour, A.B.; Park, Y.; Crispini, L.; Läufer, A.; Kuk Hong, J.; Park, T.-Y.S.; Zoheir, B.; Pradhan, B.; Muslim, A.M.; Hossain, M.S.; et al. Mapping Listvenite Occurrences in the Damage Zones of Northern Victoria Land, Antarctica Using ASTER Satellite Remote Sensing Data. *Remote Sens.* **2019**, *11*, 1408. [CrossRef]

90. Crosta, A.P.; De Souza Filho, C.R.; Azevedo, F.; Brodie, C. Targeting key alteration minerals in epithermal deposits in Patagonia, Argentina, using ASTER imagery and principal component analysis. *Int. J. Remote Sens.* **2003**, *24*, 4233–4240. [CrossRef]

91. Loughlin, W.P. Principal component analysis for alteration mapping. *Photogramm. Eng. Remote Sens.* **1991**, *57*, 1163–1169.

92. Honarmand, M.; Ranjbar, H.; Shahabpour, J. Application of principal component analysis and spectral angle mapper in the mapping of hydrothermal alteration in the Jebal–Barez Area, Southeastern Iran. *Resour. Geol.* **2012**, *62*, 119–139. [CrossRef]

93. Khaleghi, M.; Ranjbar, H.; Shahabpour, J.; Honarmand, M. Spectral angle mapping, spectral information divergence, and principal component analysis of the ASTER SWIR data for exploration of porphyry copper mineralization in the Sarduiyeh area, Kerman province, Iran. *Appl. Geomat.* **2014**, *6*, 49–58. [CrossRef]

94. Zuo, R. Machine learning of mineralization-related geochemical anomalies: A review of potential methods. *Nat. Resour. Res.* **2017**, *26*, 457–464. [CrossRef]

95. Zuo, R.; Xiong, Y.; Wang, J.; Carranza, E.J.M. Deep learning and its application in geochemical mapping. *Earth-Sci. Rev.* **2019**, *192*, 1–14. [CrossRef]

96. Zuo, R.; Wang, J.; Xiong, Y.; Wang, Z. The processing methods of geochemical exploration data: Past, present, and future. *Appl. Geochem.* **2021**, *132*, 105072. [CrossRef]

97. Zuo, R.; Wang, Z. Effects of random negative training samples on mineral prospectivity mapping. *Nat. Resour. Res.* **2020**, *29*, 3443–3455. [CrossRef]

98. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

99. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]

100. Strobl, C.; Boulesteix, A.L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [CrossRef]

101. Ziaii, M. Lithogeochemical Exploration Methods for Porphyry Copper Deposit in Sungun, NW Iran. Unpublished Master's Thesis, Moscow State University (MSU), Moscow, Russia, 1996; p. 98. (In Russian).

102. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.

103. R Development Core Team. The R Project for Statistical Computing. 2008. Available online: http://www.R-project.org (accessed on 28 October 2018).

104. Grömping, U. Variable importance assessment in regression: Linear regression versus random forest. *Am. Stat.* **2009**, *63*, 308–319. [CrossRef]

105. Barton, M.D. Granitic magmatism and metallogeny of southwestern North America. *Earth Environ. Sci. Trans. R. Soc. Edinb.* **1996**, *87*, 261–280.

106. Groves, D.I.; Vielreicher, R.M.; Goldfarb, R.J.; Condie, K.C. Controls on the heterogeneous distribution of mineral deposits through time. *Geol. Soc. Lond. Spec. Publ.* **2005**, *248*, 71–101. [CrossRef]

107. Kesler, S.E.; Wilkinson, B.H. The role of exhumation in the temporal distribution of ore deposits. *Econ. Geol.* **2006**, *101*, 919–922. [CrossRef]
108. Zürcher, L.; Hammarstrom, J.M.; Mars, J.C.; Ludington, S.D.; Zientek, M.L. *Relationship between Porphyry Systems, Crustal Preservation Levels, and Amount of Exploration in Magmatic Belts of the Central Tethys Region*; Society of Economic Geologists: Littleton, CO, USA, 2016.

MDPI

Article

Analysis of Water Yield Changes in the Johor River Basin, Peninsular Malaysia Using Remote Sensing Satellite Imagery

Mazlan Hashim [1,2,*], **Babangida Baiya** [2,3], **Mohd Rizaludin Mahmud** [1,2], **Dalhatu Aliyu Sani** [2], **Musa Muhammad Chindo** [2,4], **Tan Mou Leong** [5] and **Amin Beiranvand Pour** [1,6]

1   Geoscience & Digital Earth Centre (INSTEG), Research Institute for Sustainable Environment (RISE), Universiti Teknologi Malaysia (UTM), Johor Bahru 81310, Malaysia; beiranvand.pour@umt.edu.my (A.B.P.)
2   Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia (UTM), Johor Bahru 81310, Malaysia; chindo@graduate.utm.my (M.M.C.)
3   Department of Surveying and Geoinformatics, School of Environmental Science Technology, Federal Polytechnic, Mubi 360231, Nigeria
4   Department of Surveying and Geoinformatics, School of Environmental Studies, Federal Polytechnic, Nasarawa 962106, Nigeria
5   Geography Section, School of Humanities, Universiti Sains Malaysia, Gelugor 11800, Malaysia
6   Institute of Oceanography and Environment (INOS), University Malaysia Terengganu (UMT), Kuala Nerus 21300, Malaysia
*   Correspondence: mazlanhashim@utm.my; Tel.: +60-197173861

Abstract: Changes in land-use–land-cover (LULC) affect the water balance of a region by influencing the water yield (WY) along with variations in rainfall and evapotranspiration (ET). Remote sensing satellite imagery offers a comprehensive spatiotemporal distribution of LULC to analyse changes in WY over a large area. Hence, this study mapped and analyse successive changes in LULC and WY between 2000 and 2015 in the Johor River Basin (JRB) by specifically comparing satellite-based and in-situ-derived WY and characterising changes in WY in relation to LULC change magnitudes within watersheds. The WY was calculated using the water balance equation, which determines the WY from the equilibrium of precipitation minus ET. The precipitation and ET information were derived from the Tropical Rainfall Measuring Mission (TRMM) and moderate-resolution imaging spectroradiometer (MODIS) satellite data, respectively. The LULC maps were extracted from Landsat-Enhanced Thematic Mapper Plus (ETM+) and Landsat Operational Land Imager (OLI). The results demonstrate a good agreement between satellite-based derived quantities and in situ measurements, with an average bias of ± 20.04 mm and ± 43 mm for precipitation and ET, respectively. LULC changes between 2000 and 2015 indicated an increase in agriculture land other than oil palm to 11.07%, reduction in forest to 32.15%, increase in oil palm to 11.88%, and increase in urban land to 9.82%, resulting in an increase of 15.76% WY. The finding can serve as a critical initiative for satellite-based WY and LULC changes to achieve targets 6.1 and 6.2 of the United Nations Sustainable Development Goal (UNSDG) 6.

Keywords: remote sensing imagery; land use land cover; water yield; United Nations Sustainable Development Goal (UNSDG) 6

1. Introduction

Satellite-based spatio-temporal land-use land-cover (LULC) change analysis can provide comprehensive and widespread water yield (WY) distribution information [1]. However, changes in land use affect the water balance of a region by influencing WY along with variations in rainfall and evapotranspiration (ET) [2]. Land use patterns can directly change ecosystem types, landscape patterns, and ecological processes [3]. These have a significant influence on the WY on the surface, a crucial parameter for planning and managing raw water supply [4,5]. This affects ecosystem services such as biodiversity, WY characteristics, and rainfall [6].

The conventional WY measurement is based on a plot representation within a catchment/watershed. The plot could be observations of a single or a few points to measure rainfall and ET for a given period, producing a WY value based on the representation of the point or the entire polygon. However, with advances in remote sensing (RS) technologies, WY information could be obtained spatially and temporally at the pixel level through specific processing. The advantages of using RS to extract WY information include (1) comprehensiveness for every pixel; (2) reduced costs—minimum field observations are made for calibration and validation only; (3) feasibility even for remote and ungagged watersheds; and (4) derivation of WY from satellite using selected hydrologically established methods, i.e., the water balance equation, which provides a simple and robust approach. However, the approach is replicable at any location, provided there is available corresponding local satellite data and the needed processing parameters [7].

Previous studies established that satellite-based precipitation data offer high accuracy worldwide [8,9]. The higher-level data products of the Tropical Rainfall Measuring Mission (TRMM) and Global Precipitation Measurement (GPM) data products of the satellite-based precipitation data provide daily rainfall estimation within a three-hour interval at a high spatial resolution of 0.25°, which can be localized to reach a higher resolution [10,11]. Several other studies have raised concerns about the effects of changes in land cover due to LULC changes in the WY [12,13]. An assessment of the relationship between land use and hydrological cycle components was also carried out [14–17]. Runoff modelling for drought and flood monitoring was reported [18]. These studies, as mentioned above, only widely reported the effects of LULC changes on WY, without taking cognizance of the influence of individual LULC classes on WY. This could lead to an inadequate understanding of WY dynamics.

Between 1984 and 2015, about 90,000 km^2 of permanent surface water was lost from the Earth's surface, primarily due to human extraction and increased evapotranspiration caused by climate change [19]. Thus, accounting for water is critical to developing policies and procedures that ensure the maintenance of a regional or basin-scale water balance over time [19,20]. Although Malaysia is considered a low-water-stress country based on the World Resource Institute Ranking [21], Johor, despite its good river system [22] centred around the Johor River, which provides 60% of its water supply [23], is known for water scarcity [24]. According to the National Water Resource study, demand for water in Johor has consistently increased, and this trend will continue [25]. This may not be unconnected to the increasing population, urbanisation, industrialisation and large-scale agricultural activities [26], increasing demands for water [27].

The increasing human activities in Johor are also causing large-scale LUCCs [23,28,29], which, in turn, are affecting the water balance with intermittent and regular incidences of droughts and flooding [27,29–33]. At present, the city-state of Singapore draws 40% of its water from the Johor River [34]. Hence, the Johor River Basin (JRB) plays a pivotal role in the water resource management of both Peninsular Malaysia and the island state of Singapore. Consequently, this study achieved precise satellite-based reporting of changes in WY from six LULC classes (agriculture, bare land, forest, oil palm, urban, and water bodies) with specific objectives: (a) to compare satellite-based WY with the in-situ-derived WY, and (b) to characterise WY changes in relation to LULC within watersheds according to their corresponding magnitude of LULC. The results of this study will play a crucial role in achieving the United Nations Sustainable Development Goal 6 targets 6.1 and 6.2 through satellite-based analysis of changes in WY due to land use.

2. Materials and Methods

2.1. Study Area

The study was carried out in the Johor River Basin (JRB), located in Johor, the southernmost state of Peninsular Malaysia (PM). The state recorded a total population of about 1,638,219 people at the 2020 census (Department of Statistics Malaysia). The basin covers an area of approximately 2636 km^2, with a main river length of 123 km [4]. The elevation

ranges between 3 and 977 m above mean sea level. The site is an agricultural basin that lies between latitudes 1°30′ and 2°10′N and longitudes 103°20′ and 104°10′E (Figure 1). The primary soil type within the basin is ultisol (Rengam–Jerangau series). This soil is characterised by yellowish–brown sandy clay with moderate permeability and is well-drained, making it suitable for oil palm and rubber plantations [35]. The main land uses of the basin are oil palm and forest [36]. The river flows from the southwest of Johor from south Gunung Belumut at 1010 m elevation. The main tributaries are the Linggiu, Sayong, Penggeli, Jengeli, and Belitong rivers. These tributaries serve as sources of fresh water for the populations of Johor and Singapore. Since the middle of the 1960s, the Public Utility Board (PUB) and the Johor Water Company, Johor, Malaysia, (SAJ) have each drawn approximately 0.25×10^6 km^3/day from the JRB [37].
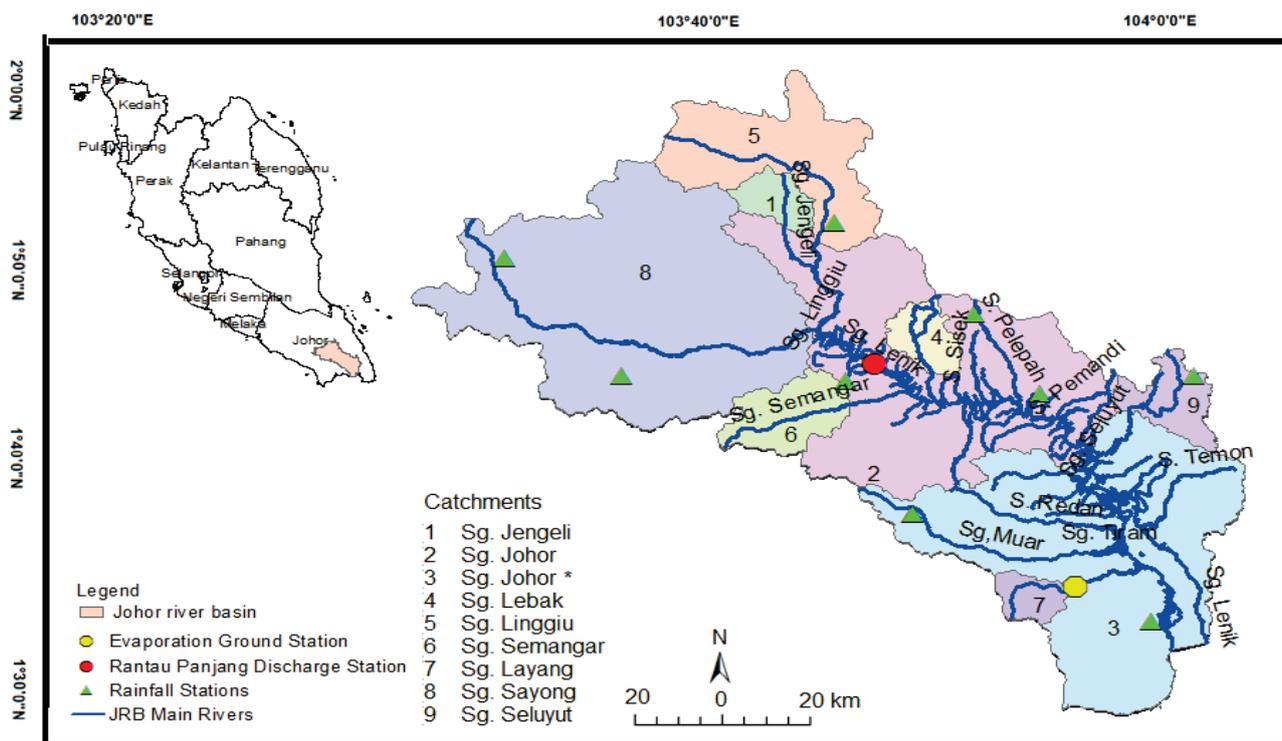
Figure 1. The Johor River Basin of Peninsular Malaysia * Sg. Johor.

The basin receives an average annual rainfall of 2500 mm, and the rainfall pattern is influenced by two seasons, namely the northeast monsoon (November–February) and the southwest monsoon (May–August). In between these two monsoons, JRB experiences intermonsoon periods that usually happen in March and April and September and October. During the northeast monsoon, most of the eastern coast of Peninsular Malaysia, including JRB, receives heavy rainfall compared to the southwest monsoon and other seasons [38]. The east coast regions receive higher rainfall (>350 mm per month), mainly during December and January.

2.2. Data Used

Five types of data were used in this study. Multitemporal satellite images of Landsat-7-enhanced thematic mapper plus (ETM+) and Landsat operational land imager (OLI) with 30 m spatial resolution, acquired from the United States Geological Survey (USGS) (http://glovis.usgs.gov, accessed on 3 December 2018), were used to generate LULC maps. Table 1 summarises the attributes of the four Landsat images used in this study, which were selected based on the longevity of data archiving. An advanced space-borne thermal emission and reflection radiometer digital elevation model (ASTER GDEM), with a spatial resolution of 30 m, was used to derive the vertical slope. A topographic map on a scale

of 1:25,000 obtained from the Department of Surveying and Mapping Malaysia (JUPEM) was used to collect ground control points (GCP) for image geometric corrections and the validation of vertical information. TMPA 3B42 version 7 daily data in hierarchical data format (HDF) with a high spatial and temporal resolution from 2000 to 2015 were used to obtain precipitation information for WY extraction. These TMPA data also provide information on rainfall relative error and gauge relative weighting. Because rainfall is the most vital input for modelling WY [39,40], corresponding rainfall observations from rainfall gauges within the JRB were also collected and used for calibration and validation of the TMPA-derived precipitation. A moderate-resolution imaging spectroradiometer (MODIS16A2) satellite data from between January 2000 and December 2015 in HDF, from TERRA and AQUA satellite platforms with 1 km spatial and monthly temporal resolutions, was used to obtain ET. The MODIS is used to provide high-quality ET data [41], containing information on potential latent energy and an ET quality check.

Table 1. The attributes of Landsat ETM+ and OLI images were used in the study.

Sensor	Scene ID	Path/Row	Date of Acquisition	* Monsoon
Landsat 7 ETM+	LEO71250582000104EDC00	125/058	29 September 2000	Post-SW
	LEO71250592005356EDC00	125/059	25 September 2005	Post-SW
	LEO71260582010292EDC00	126/058	8 August 2009	SW
Landsat 8 OLI	LE081260592015181EDC00	126/059	2 November 2015	NE

* Notes: SW = southwest, NE = northeast.

2.3. Methodology

A series of activities were involved in the processing and analysis of all the data acquired for the analysis of changes in JRB. These include LULC mapping, satellite-based water balance extraction, calibration, and validation of TMPA rainfall and MODIS ET data, the extraction of WY changes from individual LULC classes, and analysis of the WY trend. The complete methodological flow chart of this study is illustrated in Figure 2.

2.3.1. LULC Mapping

The LULC mapping was accomplished by Landsat multitemporal mapping to identify and classify different types of land use, such as forests, oil palm, agriculture other than oil palm, urban areas, bare land, and water bodies. This process consists mainly of satellite data preprocessing and image classification.

Pre-Processing of Satellite Images

Usually, deficiencies and errors are found in the raw data obtained via remote sensing satellite-sensor platforms. Therefore, the raw data acquired for the LULC mapping were enhanced by subjecting them to various preprocessing steps to certify their originality. The preprocessing consisted of three major steps: radiometric correction to eliminate systematic errors introduced by sensors and atmospheric conditions, geometric correction to correct for variations in satellite altitude and orientation during data acquisition, and image enhancement to improve the visual quality and highlight important features. All satellite data processing was carried out using the ENVI Digital Image Processing System and ArcGIS system software, as explained in detail below. This differs from image to image depending on the category of information being extracted, the composition of the image scene, and the initial condition of the image.

Initially, the images were radiometrically corrected by applying the rescaling factors and parameters obtained in the metadata file that came with the images. The correction was carried out in two-step processes: (1) the image digital numbers (DNs) were converted to top atmospheric reflectance, and (2) the top atmospheric correction of Landsat OLI and Landsat ETM+ spectral bands was achieved using the FLASH programme of ENVI v.5 software. The atmospheric correction using FLAASH was adopted to lessen the atmospheric uncertainties of nadir-viewing images via inclusivity to correct the effects of adjacency. This is vital to minimise scattering effects [42].
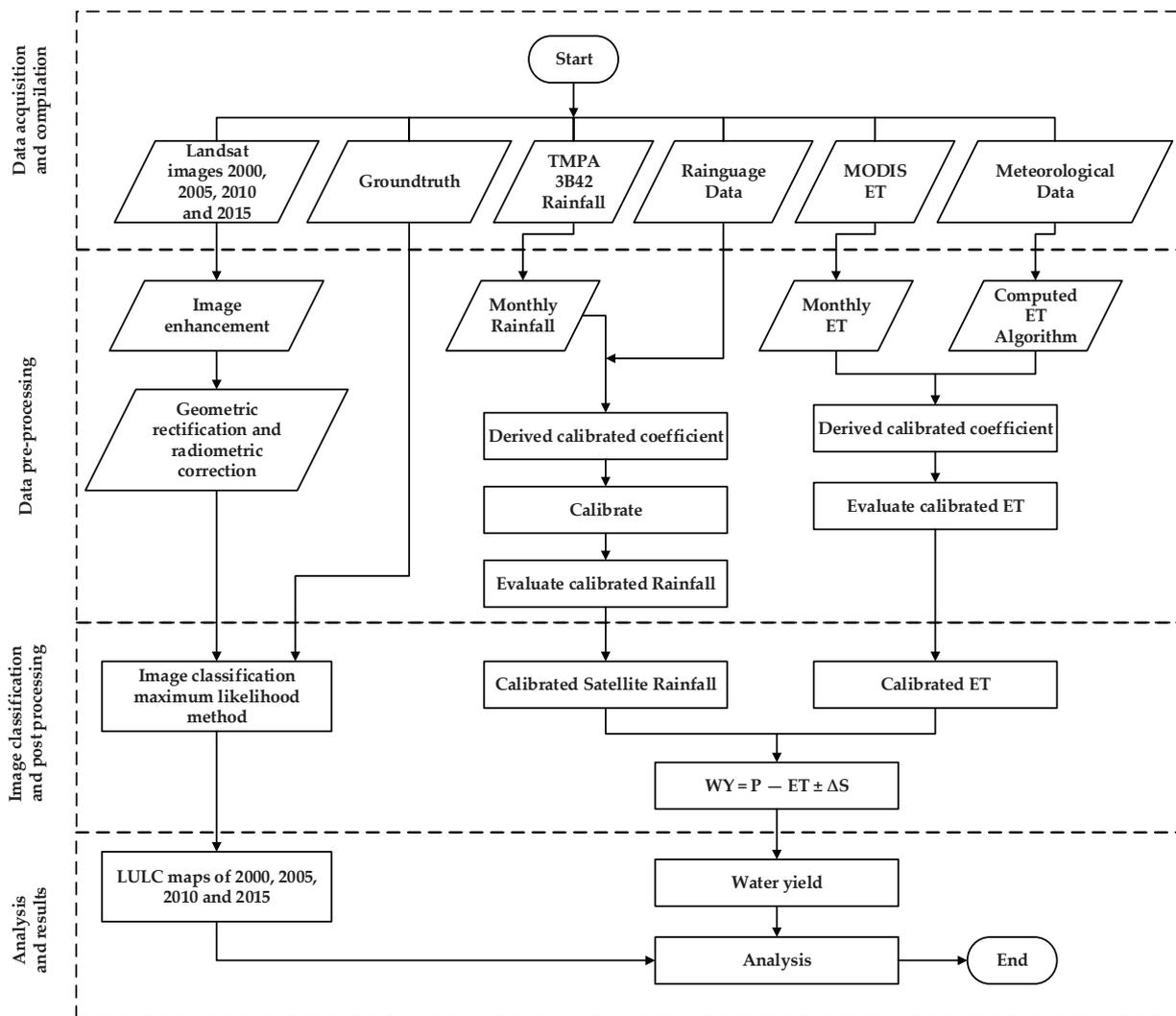
Figure 2. Schematic flow chart for water yield assessment.

The image-to-map geometric correction technique uses 30 GCP extracted from the topographic maps to correct the images' [43]. Similarly, corrections were carried out in two-step procedures: (1) a second-degree polynomial was employed to transform the images to map geometry, and (2) the pixel intensity values were created into the transformed geometry by applying the nearest neighbour resampling scheme. The root mean square error (RMSE) of the transformation was ensured to be ±0.5 pixels, and the chosen nearest neighbor resampling scheme in each case was used to avoid loss of details.

Satellite Images Classification

The maximum likelihood classifier, a popular supervised classification method, was used in the LULC mapping. The classification accuracies were evaluated using cross-validation statistics through confusion matrices using a set of 600 stratified random points. An overall accuracy that certifies the minimum threshold of 85% is needed for the effective and steadfast analysis and modelling of the LULC changes [44]. The entire classification assessment comprises the confusion matrices, producers, users, and overall accuracies with the Kappa indexes of the classification images.

2.3.2. Satellite-Based Water Balance Extraction

Satellite-based extraction of water balance operation is focused on the derivation of WY from the rainfall and ET data from TMPA and MODIS satellites, respectively. The WY

is normally determined according to the concept of the water balance equation method, which extracts the WY from the equilibrium of precipitation minus ET (see Equation (1));

$$WY_R = P_i - ET_{i,} \pm \Delta S \qquad (1)$$

where WY_R is water yield (mm), P_i is precipitation (mm) for the ith month, ET_i is evapotranspiration for the ith month [43], and ΔS is the change in soil moisture, which is insignificant in this study environment [18].

This process was realised by applying digital image processing to obtain the spatial aspect where the WY for each pixel is computed. The WY equation is a standard operation method applied by the Department of Irrigation and Drainage Malaysia (DID). The parameters used in the process of computing the WY are calibrated rainfall and calibrated ET from MODIS. In the computation of WY, there are three components inside the water balance, and all three components cover the initial stages of the water cycle up to the WY. The two main elements that were determined were ET and rainfall.

The rainfall intercepts the LULC, which infiltrates the soil, producing soil moisture; it is in this coincidental phenomenon that ET occurs. Since ET occurs when vegetation exists because of stomatal activities, soil moisture was not addressed. The difference between rainfall and ET was calculated. The results of the two elements can be positive or negative. A positive value shows that precipitation is higher than the ET. This shows the possibility that the WY can be harvested. If the outcome is negative, it means that ET is higher than rainfall.

Before entering rain and ET from TMPA and MODIS satellites into Equation (1), all these data sets were calibrated with the corresponding ground-based observations. Once calibrated, an independent assessment of the calibrated datasets was also performed as part of the validation process. Satellite-based WY assessment was carried out by (1) validation against the observed river flow, (2) comparison with similar studies in nearby watersheds using a water balance equation, and (3) modelling WY using the soil water assessment tool (SWAT).

Modeling WY Using Soil Water Assessment Tool (SWAT)

The SWAT is a semi-dispersed, comprehensive, process-based catchment, and time-uninterrupted model applied to model the potential influence of LULC changes and tue supervision of water quality and quantity [45–49]. SWAT was established by The United States Agricultural Development (USAID) developed the model using its research services in the 1998. The model, combined with a geographical information system (GIS), permits digital LULC, topography, and soil data input. In addition, the model permits the modelling of evaporation and water yield losses of drainage channels. The model repeatedly enhanced SWAT to the very recent version of 2012 [50].

Arnold et al. [49] highlight the model's flexibility in integrating upland and channel procedures and the modelling of land management. SWAT has several progressive and global applications [50]. The literature has connected to SWAT in several ways [17]. It is an appropriate model for carrying out influence studies, as the effect of LULC, climate change, or both on hydrology has been assessed by applying the SWAT model [51,52]. Additionally, the model has proved its ability to model water changes in regions with inadequate data readiness [53,54]. SWAT is gradually being applied on a larger scale [55,56].

In this study, the model was run in ArcSWAT 2012 interface, which works as an extension in ArcGIS 10.3 software. The SWAT model is freely available on the internet and downloadable from the official SWAT website https://swat.edu/software/arcswat/ (accessed on 1 November 2018). Generally, SWAT modelling is made up of six main steps, as shown in Figure 3. The first step needed during the model run set-up is watershed delineation using a digital elevation model, and the topographic characteristics of the watershed are estimated. The next step is the analysis of hydrological response units (HRU), in which layers of LULC and soil maps were added, and slope ranges were defined. Subsequently, climate station data were added. Next, the input parameters were edited,

checked, and validated through data-quality checking, data-sorting or data-processing. After running the model successfully, sensitivity, calibration, and validation were carried out. Similar studies were also carried out [57]. Finally, the extra sub-basin outlet was added based on the location of the hydro-gauging station and Table 2. The primary purpose of adding an outlet point at the Rantau Panjang Monitoring Station is to compare simulated and observed discharge.
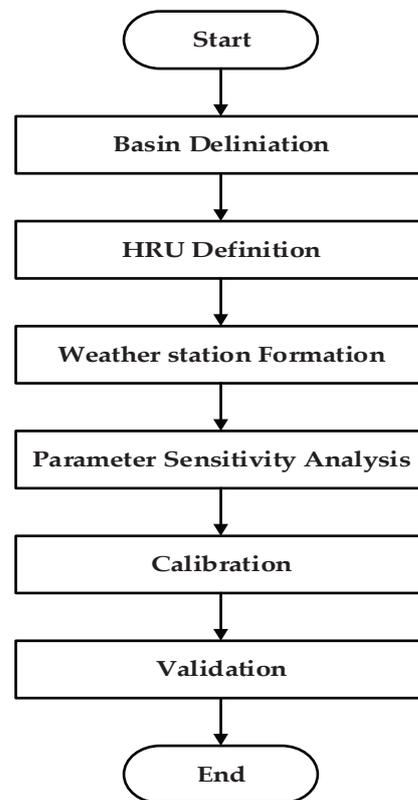

Figure 3. SWAT simulation processes.

Table 2. Climate parameters in SWAT model database.

Parameters	Details
WLATITUDE	Latitude of the climate station
WLONGITUDE	Longitude of the climate station
WELEV	Height of the climate station in meters above mean sea level
RAIN_YRS	Number of years of maximum monthly half-hour rainfall data used to define values for average per month of the year
TMPMX	Average daily maximum air temperature for each month in degree Celsius (°C)
TMPMX	Average daily maximum air temperature for each month in degree Celsius (°C)
TMPSTDMN	The standard deviation for daily minimum air temperature for each month in degree Celsius (°C)
PCPMM	The average total of monthly precipitation in milometers (mm)
PCPSTD	The standard deviation for daily precipitation for each month expressed as in mm of water per day.
PCPSKW	The skew coefficient for daily rainfall for each month
PR_W1	Possibility of a wet day following a dry day for each month
PR_W2	Possibility of a wet day following a wet day for each month
PCPD	mean numbers of days of rainfall for each month
RAINHHMX	Most extreme half-hour precipitation for each month
SOLARAV	Average daily solar radiation for each month.
DEWPT	Average daily few points temperature per month of the year in degree Celsius (°C)
WNDAV	Average daily wind speed for each month (m/s)

Source: [58].

LULC and soil map are inputs to determine the land/soil categories required to establish hydrological response units (HRU). The SWAT categorised the closes climate station to the centroid of sub-catchment and applied it to all the HRUs inside the sub-basin.

Modification of the SWAT model involves the climate and soil parameters in the database. For the climate, SWAT requires daily variables to populate the weather matrix with averages for each month of the year over the total period covered by the station. The parameters are generated from a daily record of the observed data (ideally more than 30 years). For this study, the climate data were from 1980 to 2009, while discharge data were from 1970 to 2018. Table 2 demonstrates the climate parameters that are essential to modify the weather generator (WGN) database in the SWAT model.

Soil information varies from place to place, so modification of the soil database in the SWAT is significant to ensure the better modelling of water within the layers of soil. The soil parameters required to modify the SWAT model's soil database are presented in Table 3. According to the water balance Equation (5), SWAT carries out the simulation.

$$(SW_t = SW_o + \sum_{i=1}^{t} \left(R_{day} - Q_{surf} - E_a - W_{seep} - Q_{gw} \right) \tag{2}$$

where SW_t = final soil water content (mm), SW_o = initial soil water content on day i (mm), t = time (days), R_{day} = amount of precipitation on day i (mm), Q_{surf} = amount of surface runoff on day i (mm), E = amount of evapotranspiration on day i (mmH$_2$O), W_{seep} = amount of water entering the vadose zone from the soil profile on day i (mm), and Q_{gw} = amount of return flow on day i (mm).

Table 3. Parameters for sensitivity analysis in SWAT model.

Parameter	Details
TEXTURE	Texture of soil layer
HYDGRP	Soil hydrologic group
SOL_ZMX	Maximum rooting depth for soil profile (mm)
ANION_EXCL	A fraction of porosity (void space) from which union are excluded
SOL_CRK	Crack volume potential of soil
SOL_PH1	A soil PH of the first layer of soil.
SOL_Z1	Depth from the soil surface to the bottom of the first layer of soil (mm).
SOL_BD1	Moist bulk density of the first layer of soil (g/cm^3)
SOL_AWC1	Accessible water capacity of the soil layer number one (mm)
SOL_K1	Wet hydraulic conductivity of the first layer of soil (mm/h).
SOL_CBN1	The organic carbon content of layer one of the soil (%)
CLAY1	The clay content of layer one of the soil (%).
SILT1	Silt content of layer one of the soil (%).
SAND1	Sand content of layer one of the soil (%)
ROCK1	Rock fragment content of layer one of the soil (%)
SOL_ALB1	Moist soil albedo of layer one of the soil.
NLAYERS	Number of layers in the soil
SOL_EC1	Soil electrical conductivity of the first layer of soil(ds/m).
SOL_CAL1	Calcium carbonate content of layer one of the soil (%)
USLE_k1	USLE equation of soil erodibility (K) factor of the first layer of soil.

Source: [58].

The parameters listed in Table 3 are considered the most sensitive parameters that affect the hydrological responses of the river basin. These parameters are applied to control the amount of erosion from the channel and its catchment as they affect the rate of runoff, sediment, and soil nutrient loss to the maximum extent.

The SWAT-Cup 2012 program developed by [59] was applied for a sensitivity analysis of the calibration and validation of the SWAT model. The LULC, DEM, soil and climate data (precipitation, temperature minimum and maximum, solar radiation, relative humidity, and wind speed) were applied as SWAT inputs. The observed monthly discharge (1970–2018) at

Rantau Panjang point was applied for calibration and validation. The model was calibrated for 1985–1999 (15 years) after an initial 5-year model warmup period (1970–1984) was used to obtain a better parameterisation of the simulation based on local conditions [49].

The global sensitivity analysis method was used to test 10 parameters with 500 runs (each run has various combinations) performed in parallel with calibration. The fresh parameters obtained during calibration of the model were used for validation of the model. The Sequential Uncertainty Fitting algorithm (SUFI-2), a semi-automatic inverse modelling procedure in the SWAT-CUP, was selected because of its handle and ability to analyse many parameters using the smallest number of model runs [60,61]. A detailed description and processing procedure for the SUFI-2 algorithm within the SWAT-CUP can be found in [62].

2.3.3. Calibration and Validation
Calibration of TMPA Rainfall with Reference to Observed Rain Gauge Data

Monthly rainfall data for nine (9) stations in nine catchments in the JRB from 2000 to 2010 were collected from the Malaysian Meteorological Department (MMD) and DID. Because the reliability of WY measurement depends on the quality of applied rainfall data [63], the performance of the observed rain gauge and TMPA rainfall in the study area was evaluated first using conventional statistical indices [64], using long-term daily rainfall records available at nine catchments in the study area.

The TMPA rainfall observation was also subjected to calibration. This was achieved by a direct comparison of the TMPA observations with the corresponding rainfall obtained from the rain gauge observations using a linear regression analysis approach.

Thus, the calibration function of the annual average TMPA observation is formed and obtained via Equation (3) below:

$$RF_{calibrated} = C_{RF} + m_{RF} * (RF_{TRMM})$$ (3)

where $RF_{calibrated}$ = calibrated TMPA rainfall, m_{RF} = slope, c_{RF} = constant, and RF_{TMPA} = observed TMPA. The rainfall averages for individual months over 11 years ($11 \times 12 = 132$ months), m_{RF} and c_{RF} are 0.93 and 0.04, respectively, with r = 0.90 and R^2 = 0.73, at s 99% level of significance. These demonstrated that the average monthly data enhance the relationship between the TMPA and the rain gauge data; thus, the average monthly data reduce the differences between the two measurements. The slope and constant are obtained through Equation (3). Various months for each station are suitable for calibrating the TMPA for rainfall at the station level.

Validation of the calibrated TMPA data was also performed. A total of 70% of the rain gauge data randomly selected from nine (9) stations was applied for the TMPA calibration, and the remaining 30% was applied for validation. The performance of the calibrated TMPA was evaluated using RMSE (Equation (4)), where R_{sat} is the calibrated TMPA, R_g is the observed ground rainfall, and n is the number of pixels.

$$RMSE = \sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}\left(R_{sat} - R_g\right)^2\right]}$$ (4)

Calibration and Validation of MODIS ET Data

The ET obtained from the MODIS satellite data products is the MODIS high-level data product formatted by tiles, and the coordinates of the study area were applied to download the ET. This study used ET retrieved from the digital number (DN) of the MODIS 16A data product. The ET values were multiplied by a constant to convert DN into millimetres per month (Equations (5) and (6)).

$$E_T = MODIS\ 16A_{HDF\ xa}$$ (5)

$$MODIS\,16A_{HDF} = \frac{S\,x\,A\,x\,C_p\left(\frac{e_{sat}-e}{r_a}\right)}{s+y\left(l+\frac{r_s}{r_a}\right)} \tag{6}$$

where ET is the total ET estimates at monthly intervals (mm/month), MODIS16A$_{HDF}$ is unitless ET in HDF format, a is a constant which is set to 0.1, s = d(e_{sat})/dts = denotes the slope of the curve relating to saturated water-vapour pressure (e_{sat}) to temperature, A is the available energy divided between sensible heat, latent heat, and soil flux on the land surface, ρ is the air density, C_ρ is the specific heat capacity of air, r_a is the aerodynamic resistance, and r_s is the surface resistance. Surface resistance was parameterized by applying the satellite leaf area index and vegetation fraction cover. γ is the psychrometric constant. The detailed procedure for estimating MODIS 16A (ET) is presented in Figure 4.
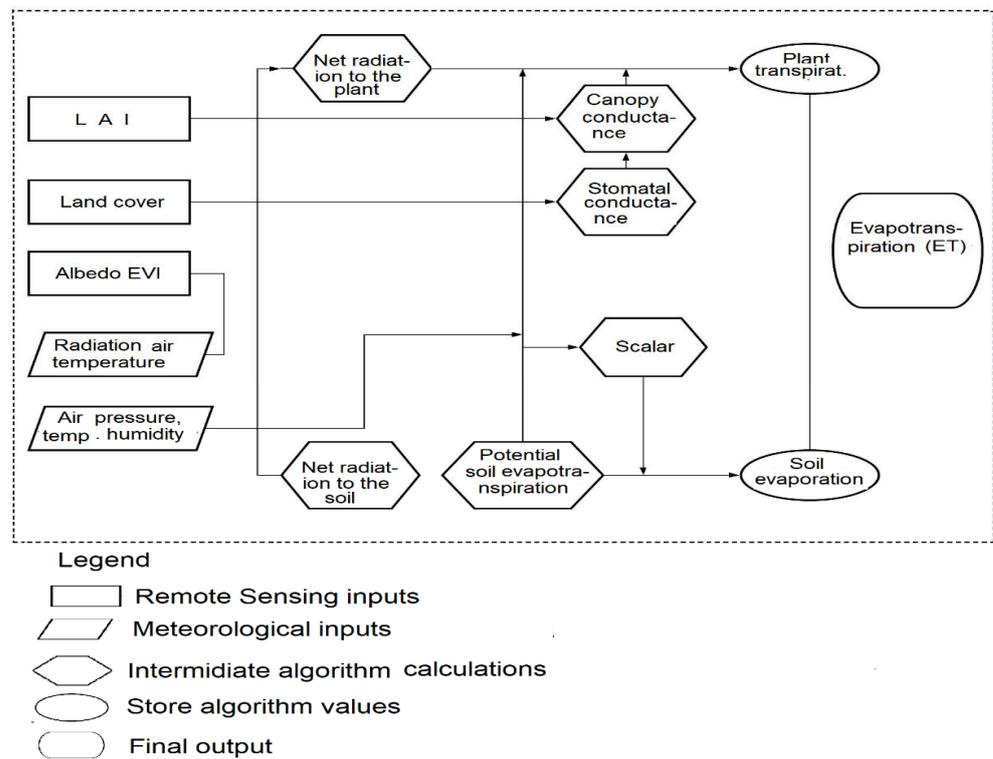
Figure 4. Scheme used for the estimation of evapotranspiration using the MODIS 16 algorithm. LAI denotes the leaf area index, and EVI symbolizes the enhanced vegetation index [18].

Satellite-based ET provides data on a global scale. Although there are numerous validations of those data sets for various climates, including in America and Asia, specific validation, especially for the tropical region, was carried out by [18]. These studies suggested that it is necessary to calibrate the data. In addition, errors related to seasonal variation have been shown to require calibration by using a linear regression function (Equation (7)) based on monsoon characteristics.

$$ET_{calib} = a * ET_{MODIS} + c, \tag{7}$$

where ET_{calib} is the calibrated satellite ET and a and c are the calibration coefficients with values of 0.36 and 54.7, respectively.

The calibration of the ET on MODIS 16A2 was performed by computing the ET from the Kluang meteorological station [18], which is the nearest weather station, using the

Penman–Monteith (PM) method [65], involving surface and water-vapor aerodynamics based on Equation (8):

$$ET = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T+273}u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{8}$$

where ET = reference ET rate (mm month^{-1}), Δ = slope of the vapor pressure curve (kpa°C^{-1}), γ = psychrometric constant (kpa °C^{-1}) R_n = net radiation at the crop surface (MJ m^{-2} d^{-1}), G = soil heat flux density (MJ m^{-2} month^{-1}) (scale G is daily assumed to be zero), e_s = saturation vapor pressure (kpa), e_a = actual vapor pressure (kpa), $e_s - e_a$ = saturation vapor pressure deficit (kpa).

The accuracy of the calibrated ET MODIS 16A was checked and validation tests were also conducted. The MODIS data were calibrated using rain gauge data collected in the periods between 2000 and 2006 and validated using data collected in the periods between 2007 and 2010. The performance of the calibrated MODIS 16A was evaluated using RMSE. The results are shown in Section 3, Results.

2.3.4. Satellite-Based Water Yield Changes from Individual LULC Classes

The extraction of changes in total WY from individual LULC classes was carried out using a logistic function to establish the correlation between the WY from the water balance equation and those derived from satellite images of different types of LULCC. The developed models used in estimating WY from the LULC classes are detailed siEquation (9):

$$LULC_{WY} = AOOP + BL + FRST + OP + UB + WB, \tag{9}$$

where:

$LULCY_{TWY}$ = total WY from LULC classes;
$AOOP$ = WY from agriculture LULC class;
BL = WY from bare land class;
$FRST$ = WY from forest class;
OP = WY from oil palm class;
UB = WY from urban class;
WB = WY from water body class.

Therefore, the total WY of the study area was extracted from Landsat ETM+ and OLI using the developed model. This derivation process was carried out by extracting satellite-based WY (per pixel) from six LULC classes (agriculture other than oil palm, bare land, forest, oil palm, urban and water bodies). Therefore, to understand the developed total WY equation, the coefficients of the variables (which represent the individual LULC classes) are used as an indicator of the contribution of each LULC to the total WY.

The effects of LULC changes on WY are analysed at 5- and 10-year intervals, focusing on LULC changes in 2000–2005, 2005–2010, 2010–2015, and 2000–2015. The LULC changes to total WY (ΔWY) for the entire JRB from the changes in LULC classes were analysed using multiple regression analysis, given by (Equations (10)–(13)):

$$\Delta WY_{2000-2005} = b_1\Delta WY_{AOOP} + b_2\Delta WY_{BL} + b_3\Delta WY_{FRST} + b_4\Delta WY_{OP} \\ + b_5\Delta WY_{UP} + b_6\Delta WY_{WP} \tag{10}$$

$$\Delta WY_{2005-2010} = b_1\Delta WY_{AOOP} + b_2\Delta WY_{BL} + b_3\Delta WY_{FRST} + b_4\Delta WY_{OP} \\ + b_5\Delta WY_{UP} + b_6\Delta WY_{WP} \tag{11}$$

$$\Delta WY_{2010-2015} = b_1\Delta WY_{AOOP} + b_2\Delta WY_{BL} + b_3\Delta WY_{FRST} + b_4\Delta WY_{OP} \\ + b_5\Delta WY_{UP} + b_6\Delta WY_{WP} \tag{12}$$

$$\Delta WY_{2000-2015} = b_1\Delta WY_{AOOP} + b_2\Delta WY_{BL} + b_3\Delta WY_{FRST} + b_4\Delta WY_{OP} \\ + b_5\Delta WY_{UP} + b_6\Delta WY_{WP} \tag{13}$$

where:

$WY_{2000–2005}$ = changes total WY between 2000 and 2005;
$b_1, b_2, b_3, b_4, b_5,$ and b_6 are coefficients of WY changes due to the LULC classes of agriculture, bare land, forest, oil palm, urban, and water body, respectively.

The assessment of changes in WY for the changes in LULC changes was carried out, focusing first on the model-fitness (based on coefficients and significance level) on WY from each LULC class, and secondly on the validation (using RMSE) of these LULC WY.

2.3.5. Mann–Kendall Statistics for Trends Analysis

Mann–Kendall (MK) statistics [63] were employed to analyse the trends of the WY (decreasing or increasing) within the JRB. The MK test was selected because it has the following advantages: (i) it does not require normally distributed data, (ii) it has been widely used by the World Meteorological Organisation (WMO) and (iii) it can handle outliers. In the trend test of Mann–Kendall, individual data in a series are compared to all succeeding data in the series. In this method, the changes between each consecutive value are computed to show increasing (+1), decreasing (−1), and neutral (0) signs. The MK (S) for a given data series $x_1, x_2, x_3 \ldots ..,$ and xn (LULC WY) were computed using Equations (14)–(17):

$$S = \sum_{k=1}^{n-1} \sum_{i=k+1}^{n} sign(x_1 - x_k)$$ (14)

where

$$sign(x_1 - x_k) = \begin{cases} +1 \, if(x_1 - x_k) > 0 \\ 0 \, if(x_1 - x_k) = 0 \\ -1 \, if(x_1 - x_k) < 0 \end{cases}$$ (15)

where n is the number of data in the series while x_1 and x_k are the ranks for the data. The significance of the trends in the data series can be calculated using statistics Z in Equation (15), as below:

$$Z = \begin{cases} \frac{S-1}{\sqrt{Var(S)}} \, if \, S > 0 \\ 0 \, if \, S = 0 \\ \frac{S+1}{\sqrt{Var(S)}} \, if \, S < 0 \end{cases}$$ (16)

where Var (S) is the variance in S positive and negative Z values, indicating the direction of the trend that exists in the time series. In this expression, V(S) variance, and in case of tied ranks it is given as in Equation (16):

$$V(S) = \frac{n(n-1)(2n+5)\sum_{i=1}^{m} t_1(t_1 - 1)(2t_1 + 5)}{18}$$ (17)

where m is the number of tied groups and t_1 is several observations in the ith group. If there is no dependence, then V(S) can be obtained through Equation (16).

3. Results

There are two main results in this study. These include a comparison of satellite-based WY with in-situ-derived WY and a characterization of WY changes in relation to LULC within watersheds according to their corresponding LULC magnitude. However, validation and calibration results are first highlighted as valuable additions to further enhance the findings.

3.1. Calibration and Validation of Satellite Data

3.1.1. Calibration and Validation of TMPA

Table 4 presents the monthly observed rainfall data from nine rain-gauge locations compared with TMPA rainfall data. Three standard statistical indices were obtained, namely: coefficient of determination (R^2), Nash–Sutcliff efficiency (NSE) and Bias were obtained to evaluate the performance of the observed rainfall and TMPA data. They showed

good agreement with TMPA monthly rainfall at the nine observed locations in JRB, with R^2 in the range of 0.60–0.75, NSE values > 0.65 at six out of nine locations, and bias at less than 10% at all the locations.

Table 4. Evaluation of the performance of the monthly rainfall of TMPA versus the observation of the rain gauge in annual rainfall estimation at all nine catchments in the study area.

S/NO.	Stations	Lat. (N)	Long. (E)	Rain Gauge Obs. Annual Rainfall (mm/yr)	TMPA Obs. R2	NSE	BIAS%
1	Sg. Jengeli	01°57′00″	103°39′00″	2268.38	0.67	0.65	9
2	Sg. Johor	01°45′30″	103°50′00″	2158.92	0.65	0.63	10
3	Sg. Johor *	01°35′30″	103°56′30″	2592.13	0.69	0.67	8
4	Sg. Layang	01°32′30″	103°53′00″	2106.36	0.64	0.62	−9
5	Sg. Lebak	01°49′00″	103°48′00″	2323.95	0.73	0.71	−7
6	Sg. Linggiu	01°59′30″	103°40′30″	2085.75	0.60	0.60	−8
7	Sg. Sayong	01°52′30″	103°30′00″	2435.79	0.75	0.73	8
8	Sg. Seluyut	01°45′00″	104°00′00″	2388.05	0.70	0.69	4
9	Sg. Semangar	01°44′00″	103°40′00″	2487.75	0.68	0.66	9

* Upper Sg. Johor.

Figure 5 summarises the linear regression analysis used for the calibration of TMPA, showing the relation between the observed TMPA and the corresponding rainfall gauge data from 2000 to 2010, from the monthly rainfall of the nine (9) stations, obtained for 132 months in 11 years. The solid line displays the linear relationship, while the dotted line is a control indicator y = x.
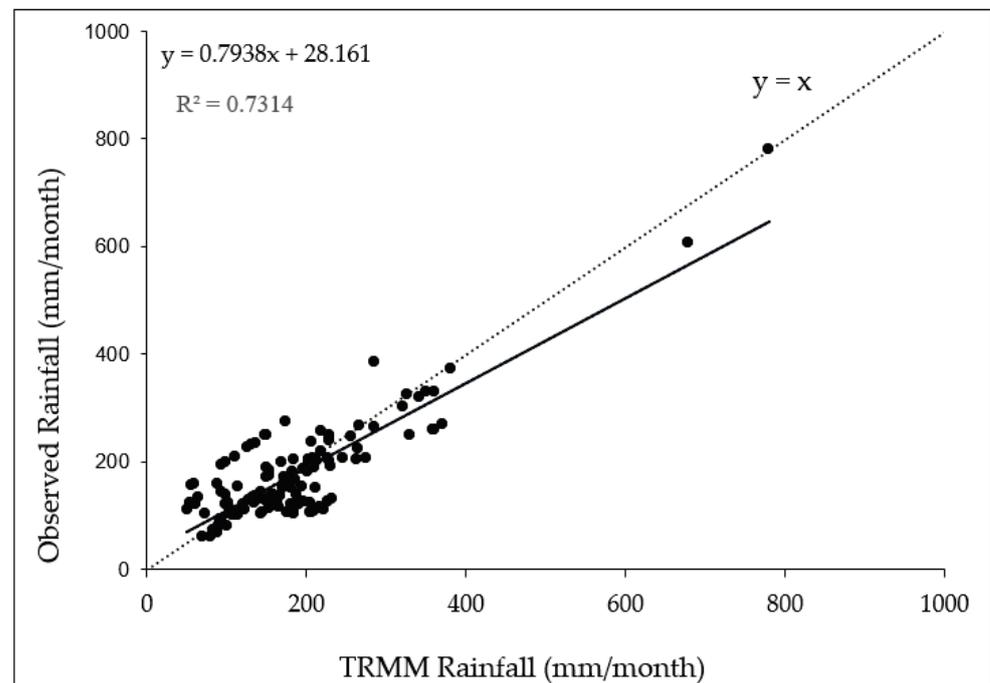
Figure 5. Comparison between observed rainfall and TMPA data (mm/month).

Figure 6 presents a comparison of uncalibrated and calibrated TMPA data using RMSE. The calibration minimises the random errors inherent in the TMPA observation, reducing the RMSE by up to 16.3%. The RMSE rainfall values estimated from TMPA calibrated data indicate that the rainfall data sets improved after calibration in terms of the RMSE values. This data calibration improvement pattern agrees with the temporal variation, where the difference between calibrated TMPA and ground measurement serves as evidence.
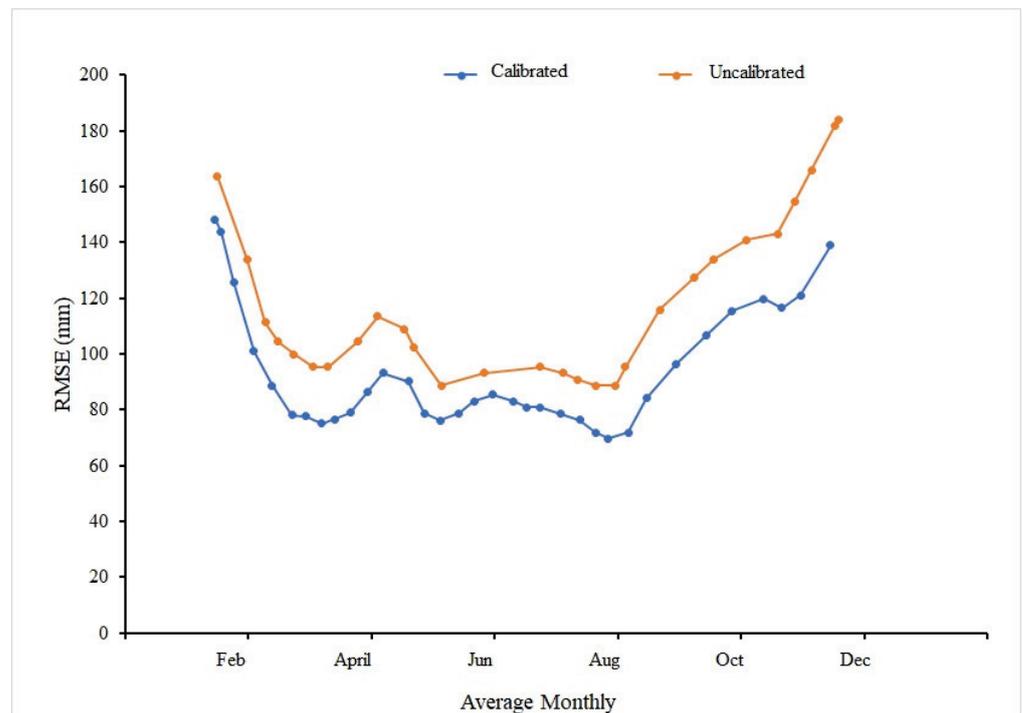
Figure 6. Average monthly root means square error for rainfall validation.

3.1.2. Calibration and Validation of MODIS ET

The calibration function derived from the regression of the ET MODIS 16A product against the corresponding ground ET at the Kluang station is presented in Figure 6. This study achieved accuracy using RMSE (±43 mm) and the determination coefficient ($R^2 = 0.642$). These assessments were realised using eleven (11) years' monthly mean of both satellite and in situ observations (Figure 7). RMSE plots for MODIS 16A ET and data from the Kluang meteorological station in 2007 and 2010 in various seasons, showing changes before and after the calibration, are presented in Figure 8. The variation in error ranged from 16 to 26 mm in the uncalibrated and calibrated MODIS ET. The calibration of MODIS ET led to improvements in WY monitoring, similarly to the research conducted by [18].
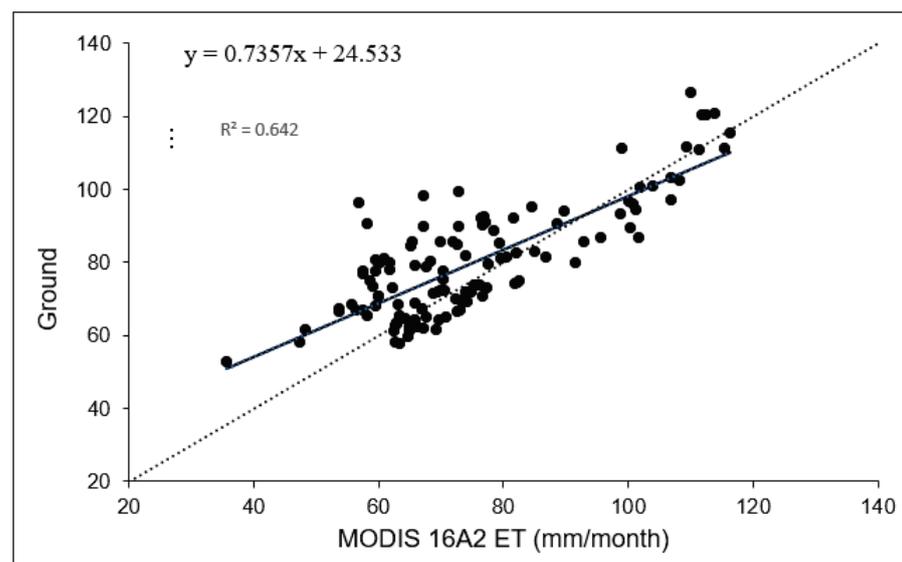
Figure 7. Calibration function derived from the regression of the ET MODIS 16A product versus the corresponding ground ET at the Kluang station.
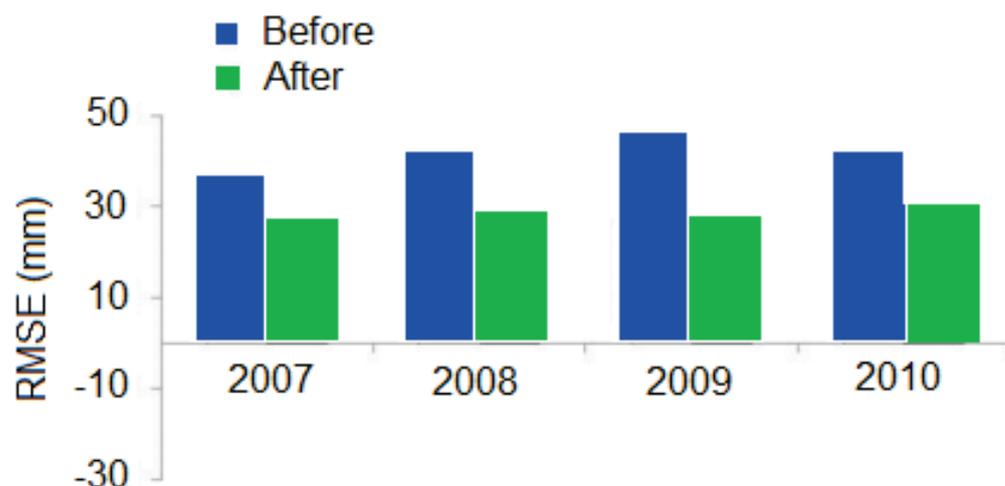
Figure 8. The plots of the root mean square error computed with the validation procedure for the MODIS 16A evapotranspiration (ET) and data from the Kluang meteorological station in 2007 and 2010 in various seasons show the changes before and after the calibration.

3.2. Assessment of Fully Satellite-Based Water Yield Compared to In-Situ-Derived Water Yield

3.2.1. Validation with River Flow Data

The satellite-based WY for JRB was also analysed to compare flood and drought occurrences. The time series graph of WY versus flood and drought is shown in Figure 9, where this catchment was reported to experience 100-year flooding in recent years [66]. The peak and lowest WY derived from this satellite-based method correspond well to the records of local flood and drought occurrences in JRB, respectively. The satellite-based and in-situ-derived WY is shown in Table 5.

3.2.2. Validation with Soil Water Assessment Tool (SWAT)

The observed and modelled monthly discharge at the Rantau Panjang discharge point is shown in Figure 8. The calibration period started from January 1985 to December 1999, while the validation period was from January 2000 to December 2009. The NSE, R^2 and PB values were (0.67, 0.67) and (-3.1), respectively, for the calibration period, whereas (0.63), (0.65) and (-1.9) are the values of NSE, R^2 and PB, respectively, for the validation period. Based on [67], for the calibration period, the NSE values demonstrate that the SWAT model for the JRB was deemed to have too good a performance for the calibration and validation times.

In Reference [68], the model demonstrates good discharge modelling performance in the Bukit Merah Reservoir, Malaysia, with an R^2 of 0.87 and 0.69 for the calibration and validation periods, respectively. For the NSE method, the SWAT output is 0.79 for calibration and 0.60 for validation periods. The performance of the SWAT model was considered 'very good' in the calibration period and 'satisfactory' in the validation period. Figure 10 demonstrated the observed and simulated means of monthly discharge recorded at the Rantau Panjang station.

The performance of the SWAT model for the period of calibration is better than that of validation; this might be because the temporal variations in the SWAT model parameters were not effectively considered. In addition, the hydrograph validation period demonstrates an overestimation of the simulated discharge during the southwest monsoon season, which was also found in other studies [66]. This may be because of the occurrence of extreme flood, where the model poorly matched the peak flow. The application of the model in Malaysia is still limited; hence, SWAT model calibration and validation in this study demonstrate that it is a consistent tool for hydrology cycle modelling in Peninsular Malaysia.
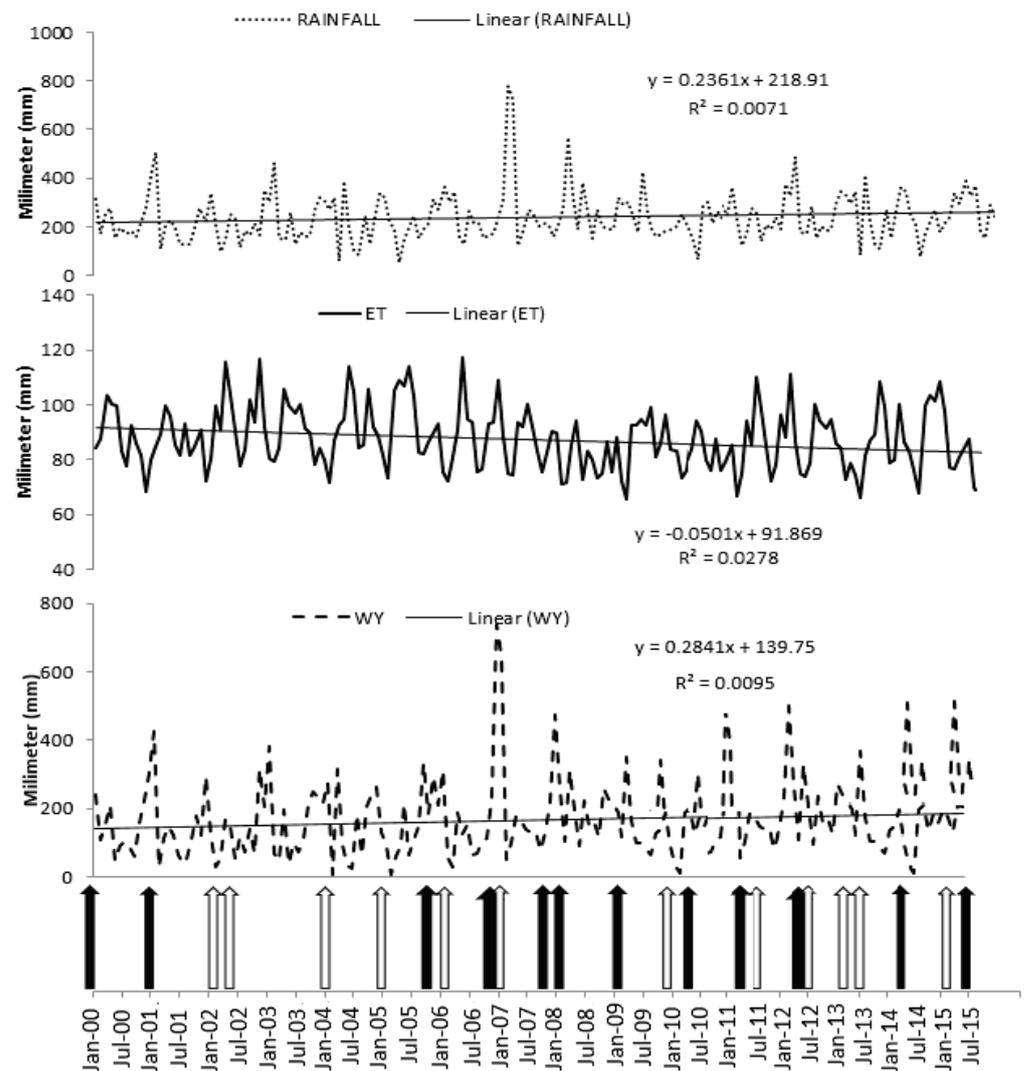
Figure 9. Monthly variations in water yield (2000–2015) and trends at the Rantau Panjang water level station. The bold and framed arrows indicated the occurrences of floods and droughts in JRB, respectively.

Table 5. The satellite-based and in-situ-derived water yield.

S/No.	Months	WY from Satellite-Based Water Balance Equation (mm/Month)				WY from SWAT (mm/Month)			
		2000	2005	2010	2015	2000	2005	2010	2015
1	January	165.76	175.52	124.29	215.86	169.23	135.93	157.87	200.67
2	February	119.49	233.95	250.14	128.24	103.88	247.23	226.17	122.59
3	March	103.86	100.66	118.17	161.44	139.47	55.66	126.02	173.01
4	April	122.57	85.26	94.83	133.98	143.26	42.46	100.00	117.4
5	May	82.38	79.35	62.40	97.51	88.06	83.47	69.59	160.63
6	June	35.17	43.82	54.49	70.20	38.08	34.09	72.12	82.49
7	July	64.88	64.98	66.97	96.81	76.48	82.08	103.29	110.53
8	August	63.60	103.94	75.91	86.81	83.36	118.55	88.80	93.97
9	September	77.49	82.92	84.60	122.44	72.36	124.4	77.35	146.65
10	October	65.63	171.61	86.52	254.29	78.96	189.34	79.73	229.29
11	November	94.40	133.05	294.25	230.85	83.35	174.51	256.78	217.80
12	December	228.87	161.15	219.50	367.89	194.90	171.16	155.32	317.74
Total WY (mm/yr)		1224.10	1436.21	1532.07	1966.32	1251.41	1458.88	1513.04	1972.77
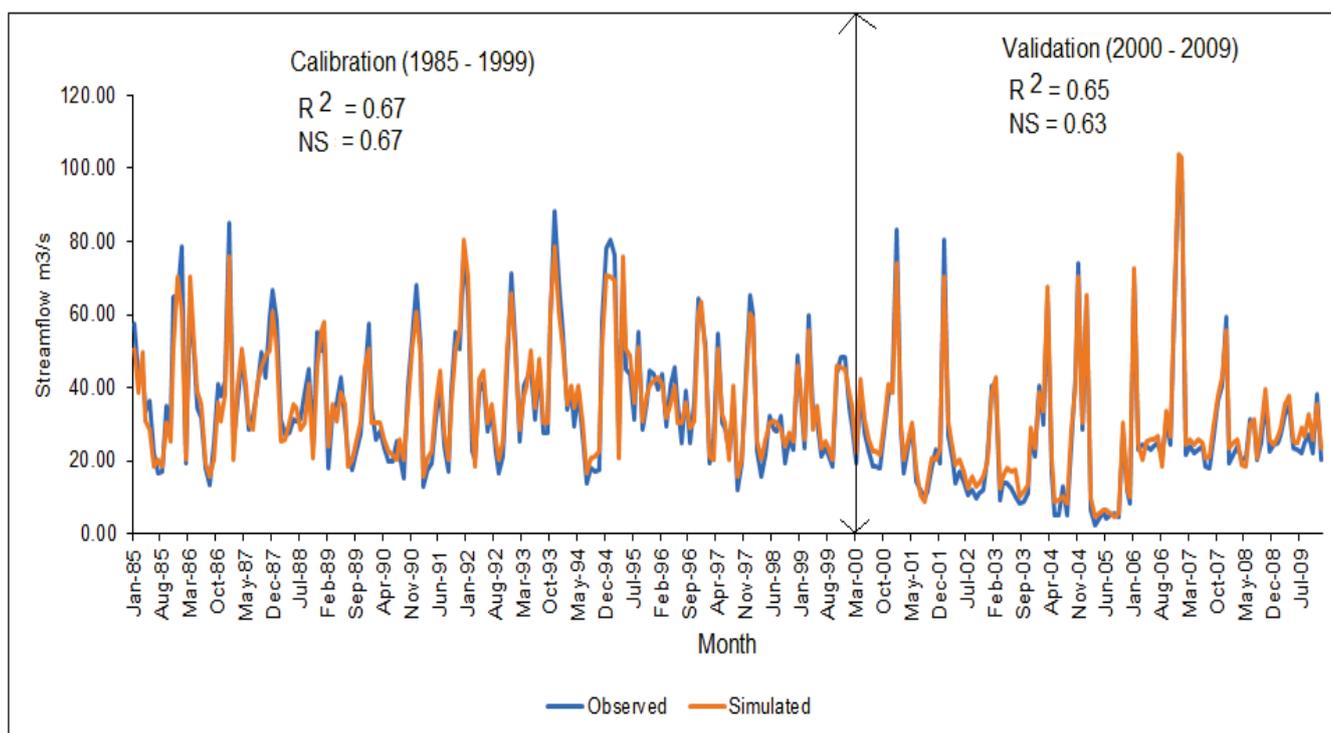
Figure 10. Observed and simulated means monthly discharge recorded at Rantau Panjang station.

3.2.3. Comparison of Similar Studies

The *WY* estimates for selected catchments in Malaysia, which have quite similar characteristics in terms of land use and rainfall, are compared with the present study (Table 6). Most of the analyses used the satellite to estimate the *WY*, except for the studies by [29] for the Johor River and by [69] for the Layang River, which combined SWAT and satellite data. The annual *WY* for various catchments ranges from 706 mm/year for JRB [29] to 1473 mm/year for Kenyir Lake [18]. For JRB, the *WY* estimates found in the earlier studies are quite similar to those in the present study.

Table 6. Estimate of water yield values from selected basins in Malaysia.

Watershed/Catchment	Type of Watershed/Catchment	Size (km²)	Mean Annual Precipitation (mm)	Total Water Yield (mm yr⁻¹)	References/Approach Satellite
Johor River	Oil palm plantation, forest, and semi-urban	2367.17	2500	710	This study/Satellite
Langat River	Semi-Urban	1257.70	2401	1207	[18]/Satellite
Layang	Semi-Urban	33.61	2690	1334	[69]/SWAT/Satellite
Pendang Terap	Forest	1032.30	2406	868	[18]/Satellite
Hulu Perak	Forest	857.30	2641	687	[18]/Satellite
Kenyir Lake	Forest	1260.00	2606	1473	[18]/Satellite
Johor River	Semi-Urban	2636.50		788	[18]/Satellite
Johor River	Semi-Urban	1652.00	2500	706	[29]/SWAT/Satellite
Hulu Langat	Forest	390.26	2453	742	[70]/SWAT/Satellite

3.3. Analysis of LULC Changes and Water Yield

Figure 11 presents the analysis of the change in LULC and WY in JRB between 2000 and 2015. Figure 11a depicts the spatio-temporal pattern of LULC classes, while Figure 11b presents their corresponding WY. The average WY for the entire catchment is presented in Figure 11c. The LULC distribution (Table 7) shows that oil palm and agriculture, other than oil palm, were the main land-cover classes, followed by urban, while there was a decline in the forested area. The accuracy of the LULC classification, summarised in Table 8, revealed high accuracy, which could not be unrelated to the use of only six classes in the classification. However, the results of the accuracy assessment show that the image classifications met the acceptance criteria.
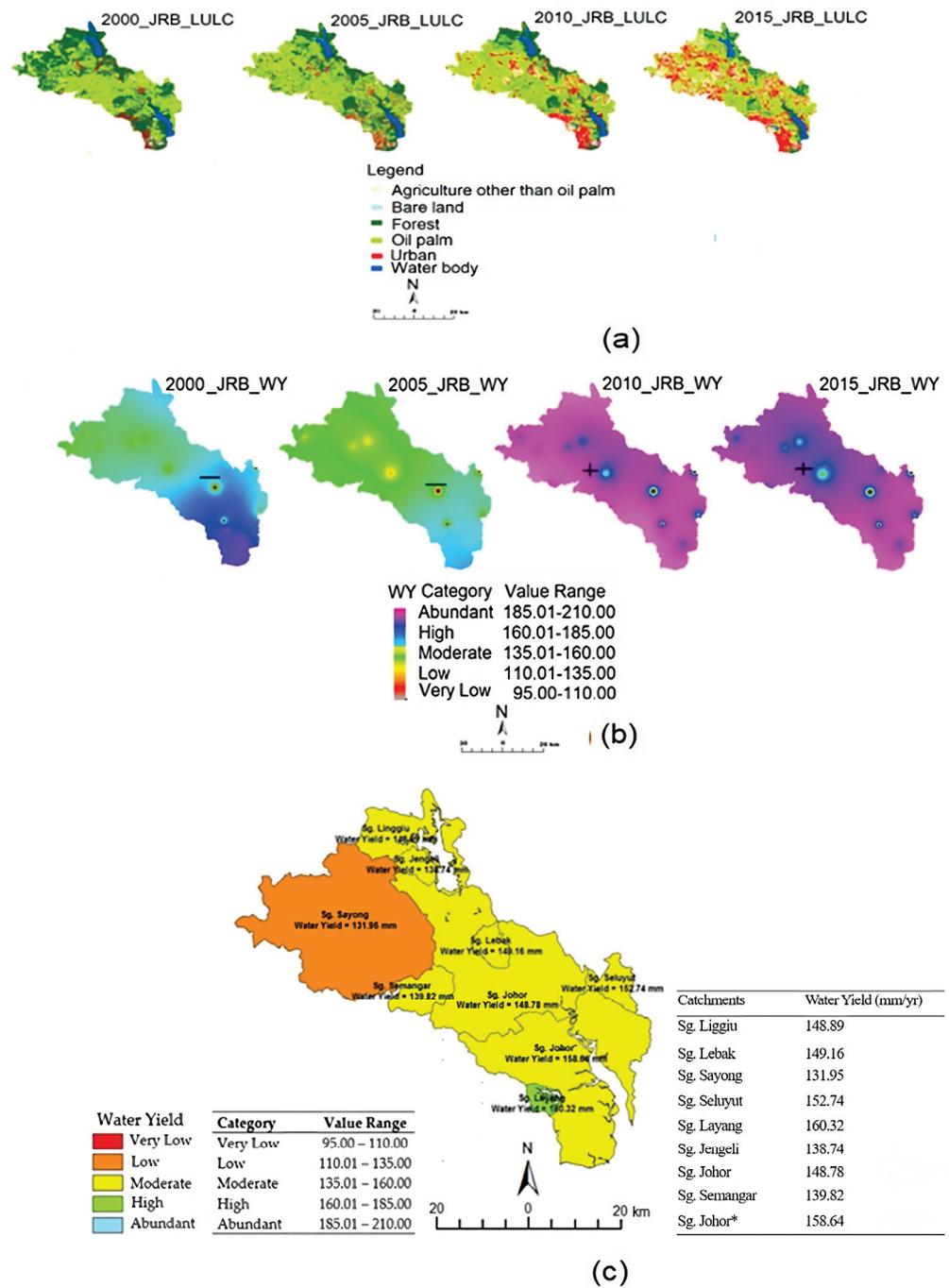
Figure 11. Spatial–temporal pattern of JRB: (**a**) LULC classes 2000, 2005, 2010 and 2015 (left to right); (**b**) water yield 2000, 2005, 2010 and 2015 (left to right), + and − signify increasing and decreasing water yield; (**c**) average WY 2000–2015 for the catchments. * Sg. Johor.

Table 7. Individual LULC classes' area water yield estimated for 2000, 2005, 2010, and 2015, as well as JRB WY for yearly comparison.

LULC Classes	Yr 2000		Yr 2005		Yr 2010		Yr 2015	
	Area (ha)	WY (mm/yr)	Area (ha)	WY (mm/yr)	Area (ha)	WY (mm/yr)	Area (ha)	WY (mm/mr)
Agriculture other than oil palm	11,943.5	535.28	24,412	717.18	33,480.8	879.59	38,169.3	983.06
Bare land	2885.4	856.7	2409.66	632.46	2145.87	314.88	5246.1	306.04

Table 7. *Cont.*

LULC Classes	Yr 2000		Yr 2005		Yr 2010		Yr 2015	
	Area	WY	Area	WY	Area	WY	Area	WY
	(ha)	(mm/yr)	(ha)	(mm/yr)	(ha)	(mm/yr)	(ha)	(mm/mr)
Forest	101,717	77.33	78,060.24	196.29	46,648.6	277.53	25,624.3	358.76
Oil palm	95,051.1	101.03	103,527	118.92	108,755.6	237.97	123,164	280.4
Urban	10,683	170.95	15,517.1	416.41	33,091	549.4	33,924.1	592.9
Water bodies	14,437.4	63.37	12,791.4	55.7	12,595.5	53.08	10,589.6	29.66
Total Water Yield (mm/yr) In JRB	606.32		756.21		803.07		1074.1	

Table 8. Classification accuracy of LULC maps of Johor River Basin.

LULU Categories	2000 (%)		2005 (%)		2010 (%)		2015 (%)	
	Prod.	User	Prod.	User	Prod.	User	Prod.	User
Agricultural area	86.33	92.64	91.33	85.50	86.33	87.83	92.00	89.46
Bare Land	89.67	89.67	88.00	91.10	91.33	92.92	90.26	87.55
Forest	93.00	88.48	85.00	85.93	98.00	88.91	90.68	92.45
Oil Palm	94.67	85.88	88.00	86.52	86.33	86.33	88.76	89.77
Urban	88.00	92.74	86.33	87.83	91.33	92.92	89.44	90.26
Waterbody	91.33	94.55	89.67	89.67	88.00	92.74	92.33	89.84
Overall Kappa	89.00		85.67		88.67		88.05	
Overall Accuracy	90.50		87.72		90.22		86.89	

The study found that between 2000 and 2015, WY from agricultural land other than oil palm, forest land, oil palm, urban areas, and water bodies increased by 20.58%, 12.93%, 8.24%, 19.39% and 1.55%, respectively, while WY from bare land decreased by 25.31%. Mann–Kendall statistics show that WY in JRB with respect to LULC showed an increasing trend between 2000 and 2005 but a decreasing trend between 2010 and 2015 (Figure 11b). The sub-basin with the highest average WY was Sg. Layang (160.32 mm/month) while the sub-basin with the lowest was Sg. Sayong (131.95 mm/month) (Figure 11c).

The results of the WY for all the catchments showed that the WY increased in 2015 compared to 2000, particularly for forest (Figure 12) and agriculture (other than oil palm). Oil palm and urban WY also increased, while bare land and water bodies WY decreased in their respective catchments from 2000 to 2015. However, the WY of individual LULC classes, estimated using a water balance equation in JRB for 2000, 2005, 2010, and 2015, respectively, is presented in Table 9.

3.4. Characterisation of WY in All Watersheds

The variation in WY intensities within LULC is shown in Table 8, and the changes in the LULC class versus WY changes are shown in Figure 13. Consequently, between 2000 and 2015, agricultural land increased by 18.78%, which led to an increase in WY by 20.58% in the same period. The bare land is reduced by 0.19%, leading to a reduction in WY of 25.31%. In contrast, a reduction in forest area to 33.40% increased the WY by 12.93%. Furthermore, an increase in oil palm area to 13.51% increased the WY by 8.24%, whereas an increase in urban areas to 18.50% increased the WY by 19.39%. The water bodies are reduced to −1.72%, while the WY extracted from the water bodies is also reduced by 1.55% (Table 10).

As outlined in Equation (1), in modelling the total WY attributed to LULC changes, it was found that the use of the AIC was the lowest at −394.78, suggesting the model as the optimum, as in Equation (18), and the full solution tabulated in Table 11. The AIK is a statistical method to evaluate how well a model fits the data from which it was generated.
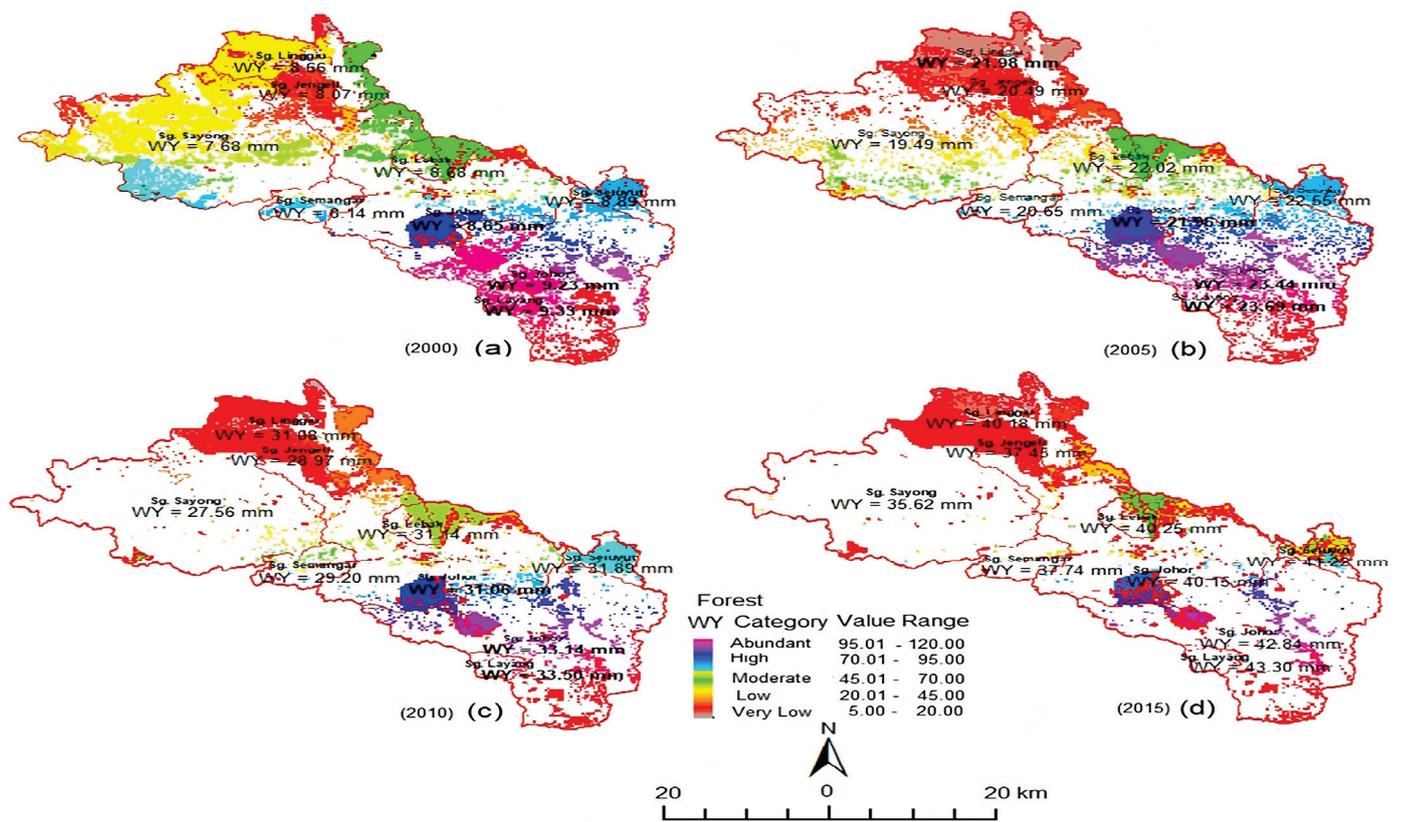
Figure 12. Variation in WY (mm/month) in forest LULC classes for 2000 (**a**), 2005 (**b**), 2010 (**c**) and 2015 (**d**).

Table 9. Individual LULC classes' water yield estimation using a water balance equation in JRB for 2000, 2005, 2010, and 2015.

S/No.	LULC Classes	2000 WY (mm/yr)	2005 WY (mm/yr)	2010 WY (mm/yr)	2015 WY (mm/yr)	Δ 2000–2015 (%)	*p*-Val.	RMSE
1	Agriculture other than oil palm	535.28	717.18	879.59	983.06	20.58	0.001	0.64
2	Bare land	856.7	632.46	314.88	306.04	25.31	0.001	0.58
3	Forest	77.33	196.29	277.53	358.76	12.93	0.001	0.56
4	Oil palm	101.03	118.92	237.97	280.4	8.24	0.001	0.6
5	Urban	170.95	416.41	549.4	592.9	19.39	0.001	0.62
6	Water body	63.37	55.7	53.08	29.66	1.55	0.001	0.59

Table 10. LULC changes 2000–2015 and WY changes.

* LULC Classes	LULC Changes 2000–2005		LULC Changes 2005–2010		LULC Changes 2010–2015		LULC Changes 2000–2015	
S/No.	Ha	%	Ha	%	Ha	%	Ha	%
1	12,455.73	5.26	9081.63	3.84	22,838.76	9.67	44,355.87	18.78
2	−482.94	−0.20	−256.59	−0.11	293.76	0.12	−448.11	−0.19
3	−23,590.98	−9.97	−31,476.87	−13.30	−23,920.47	−10.13	−78,897.78	−33.40
4	8429.49	3.56	5261.31	2.22	18,265.68	7.73	4644.45	13.51
5	4824.90	2.04	17,583.48	7.43	21,307.23	9.02	43,695.90	18.50
6	−1636.20	−0.69	−192.96	−0.08	−2253.60	−0.95	−4061.43	−1.72
Changes in Water yield (mm)	149.89	16.02	46.86	5.01	271.03	28.97	467.78	50.00

* Notes. 1 = agriculture other than OP; 2 = bareland; 3 = forest; 4 = oil palm; 5 = urban; and 6 = water body.
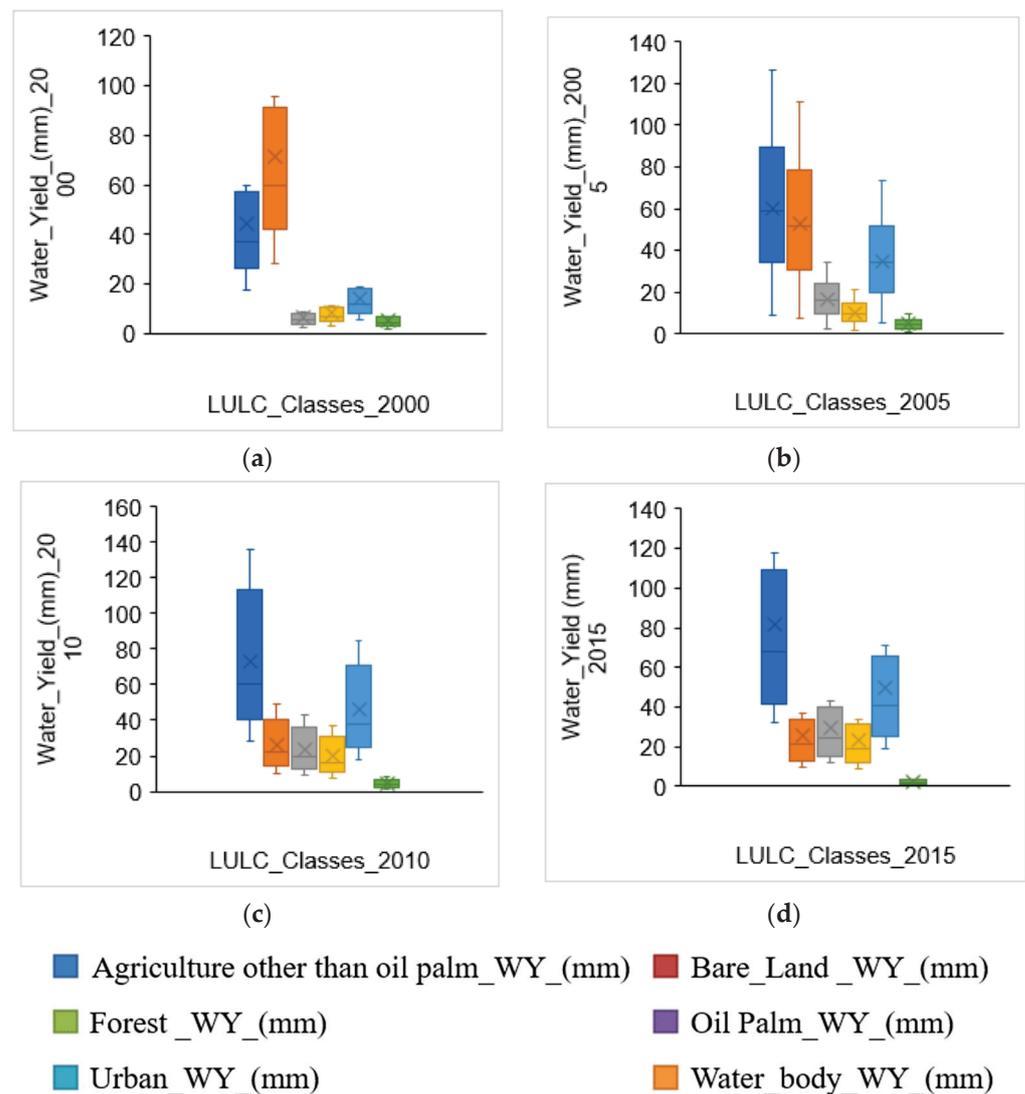
Figure 13. LULC classes' water yield changes box plot for the year: (**a**) 2000, (**b**) 2005, (**c**) 2010, and (**d**) 2015.

Table 11. The equation to best predict water yield.

LULC Classes	Estimate	Std. Error	*t* Value
Intercept	−0.004165 ***	0.001056	−3.942
Agriculture	1.643754 ***	0.040208	40.881
Bare land	1.093582 ***	0.060279	18.142
Forest	1.003378 ***	0.089682	11.188
Oil Palm	0.688391 ***	0.149676	4.599
Urban	0.752166 ***	0.016001	47.008

*** indicates a significant difference at $p < 0.001$.

However, the water body does not contribute significantly to WY. The strongest parameter that contributed to changes in WY was agriculture. This could be supported by the fact that the water balance equation used is based on the runoff model; hence, it is evident of there being no significant runoff over water bodies.

$$y_i = -0.004165 + 1.644x_1 + 1.094x_2 + 1.033x_3 + 0.688x_4 + 0.752x_5 + \varepsilon_i \qquad (18)$$

where yi is total water yield (in mm unit) for watershed; x_1, x_2, x_3, x_4, $x5$ is amount for WY (unit in mm) for land cover classes of agriculture (other than oil palm), bare land, forest, oil palm and urban, respectively, and *ε is error such that ε ∼ N(0, 0.037).*

However, there was no viable solution for Equation (17) in a further investigation of which LULC class changes affect the intensity of respective WY changes due to "limited LULC changes." The LULC changes do not show many changes within a short period, such as a less than a 5-year interval, except for the small percentage of urbanised areas. From 2000 to 2015, only four LULC maps were generated by classifying Landsat TM and OLI data acquired in 2000, 2005, 2010, and 2015. On the other hand, the average monthly satellite-based WY could be obtained from the daily WY derived from satellite precipitation and ET rates, producing adequate dependent parameters.

4. Discussion

Changes in WY have become a crucial issue in sustainable development across the globe. One of the obvious reasons for this change in WY is the LULC change. Satellite-based earth observations, as an important source of data for several aspects of the earth, have the key benefits of a synoptic view of the earth surface, regular and repeatable observation, monitoring of remote and inaccessible areas, and time series observation. Therefore, they can offer valuable data that can be applied to precisely estimate WY [8,71] and its changes over time due to LULC changes.

The analysis of WY in JRB between 2000 and 2015 revealed that changes in land use can have a significant impact on WY, as is the case in some forested ecosystems in the United States of America (USA) [72] due to the impact of landcover changes on evaporation, streamflow, and runoff [72–74]. However, in line with one of the recommendations by [72] to consider other landcovers, this study revealed that the highest increase in WY over the study area came from agricultural land other than oil palm, which increased by 18.78% and resulted in a 20.58% increase in WY. An increase of 13.51% in oil palm area also resulted in an 8.24% increase in WY. This demonstrates that an increase in agricultural activity can result in an increase in water resources and is likely due to a reduction in infiltration due to the replacement of natural vegetation cover with agricultural products and an increase in surface runoff. The increase of 12.93% in WY due to the reduction in forest area to 33.40% further emphasises the argument of an increase in WY due to an increase in agriculture, which could be due to the conversion of natural vegetation cover to agricultural fields and a reduction in infiltration. Nevertheless, the case of an increase of 19.39% WY due to an 18.50% increase in urban land cannot be unrelated to the increased compaction of the surface, reduced infiltration, and increased surface runoff, as highlighted in [75]. Finally, the 1.55% decrease in WY due to the reduction in water bodies is not surprising given the increased conversion of natural land for agriculture and the reduction in forest, which can expose shallower surface water to scorching evaporation, and, of course, lead to increased consumption due to increased urbanisation.

However, while the increase in water yield found in this study is a positive result, it is critical to recognise the potential negative consequences of natural vegetation conversion to agriculture. Land cover changes in the form of damage to endangered native vegetation can have a wide range of environmental implications [76], which should be carefully evaluated. The possible loss of biodiversity and ecosystem services offered by natural vegetation is one of the major drawbacks of this land cover change. Forests, for example, are home to many unique plant and animal species that contribute to the general health and stability of ecosystems. There is a risk of habitat fragmentation, loss of species richness, and the disruption of ecological processes when forests are converted to agricultural land. These ecological alterations may have long-term consequences for ecosystem functioning, such as diminished pollination, nutrient cycling, and natural pest control, among other things [77,78]. Furthermore, the conversion of natural vegetation to agriculture frequently involves the use of intensive farming practices, such as the application of artificial fertilisers and pesticides. The improper management of these inputs can lead to soil deterioration,

water contamination, and a loss of soil fertility over time. Furthermore, the removal of vegetation cover can enhance soil erosion, as already reported in the study area by [79], resulting in sedimentation in surrounding water bodies and potentially detrimental effects on water quality, as reported in [80]. It is also important to think about the social and economic implications of land cover changes. The conversion of natural vegetation to agricultural land may have consequences for local communities, particularly those that rely on forest resources for a living. Some of the social consequences that should be considered include the displacement of indigenous or marginalised communities, the loss of traditional knowledge, and changes in cultural practices.

Notwithstanding, this study highlights the importance of natural ecosystem conservation and minimising land conversion for other uses, as well as the importance of proper water body management in order to maintain their contributions to water resources. Overall, these findings indicate that careful consideration of LULC changes in the JRB is critical to ensure the basin's water resource sustainability and contributes to the development of a baseline for the effects of LULC changes on WY in the basin, which is highly significant for efficient land-use planning, ecological restoration, and management and guidance for regional socioeconomic development. Furthermore, the study can contribute significantly to related industries, such as the Public Utility Board, in deciding water resource assessment with regard to the quantity of water. The study contributes substantial support to understanding the degree of expansion in urban planning and the distribution of resources within the watershed.

5. Conclusions

In summary, the study reveals satellite-based techniques' ability to serve as a powerful tool to monitor the effects of spatiotemporal trend mapping and the estimation of LULC changes on WY. This effort was carried out between 2000 and 2015 in the JRB of Peninsular Malaysia. The effects of LULC changes on WY were achieved using a water balance equation technique that determines WY from the equilibrium of precipitation minus ET. The process was achieved by applying digital image processing to obtain WY from individual LULCs for the entire catchment. The method was validated using in situ measurement and derived values and reported good agreement with ground-based rainfall, ET, and river discharge. Between 2000 and 2015, agricultural areas other than oil palm increased to 11.07%, forest decreased to 32.15%, oil palm increased to 11.88%, urban areas increased to 9.82%, and WY increased to 15.76%. The findings will provide valuable information for water resource management and planning, as well as aiding policymakers in taking proactive steps to mitigate the effects of changes in LULC to WY. Consequently, the outcome will fast-track the attainment of the 2030 agenda of United Nations Sustainable Development Goal 6, targets 6.1 and 6.4.

Centre (USGS) for archive of Landsat Earth Observation Satellite Images Data provided for this research study. We acknowledge the tertiary education trust fund (Tedfund) Nigeria and Federal Polytechnic, Mubi, Nigeria, for providing a scholarship to Baiya Babangida.

References

1. Spruce, J.; Bolten, J.; Mohammed, I.N.; Srinivasan, R.; Lakshmi, V. Mapping Land Use Land Cover Change in the Lower Mekong Basin from 1997 to 2010. *Front. Environ. Sci.* **2020**, *8*, 21. [CrossRef]
2. Kundu, S.; Khare, D.; Mondal, A. Past, Present and Future Land Use Changes and Their Impact on Water Balance. *J. Environ. Manag.* **2017**, *197*, 582–596. [CrossRef] [PubMed]
3. Li, Y.; Chang, J.; Luo, L.; Wang, Y.; Guo, A.; Ma, F.; Fan, J. Spatiotemporal Impacts of Land Use Land Cover Changes on Hydrology from the Mechanism Perspective Using SWAT Model with Time-Varying Parameters. *Hydrol. Res.* **2019**, *50*, 244–261. [CrossRef]
4. Wang, L.; Pijanowski, B.; Yang, W.; Zhai, R.; Omrani, H.; Li, K. Predicting Multiple Land Use Transitions under Rapid Urbanization and Implications for Land Management and Urban Planning: The Case of Zhanggong District in Central China. *Habitat Int.* **2018**, *82*, 48–61. [CrossRef]
5. Yang, L.; Feng, Q.; Yin, Z.; Wen, X.; Si, J.; Li, C.; Deo, R.C. Identifying Separate Impacts of Climate and Land Use/Cover Change on Hydrological Processes in Upper Stream of Heihe River, Northwest China. *Hydrol. Process.* **2017**, *31*, 1100–1112. [CrossRef]
6. Lang, Y.; Song, W.; Zhang, Y. Responses of the Water-Yield Ecosystem Service to Climate and Land Use Change in Sancha River Basin, China. *Phys. Chem. Earth* **2017**, *101*, 102–111. [CrossRef]
7. Paz, I.; Tchiguirinskaia, I.; Schertzer, D. Rain Gauge Networks' Limitations and the Implications to Hydrological Modelling Highlighted with a X-Band Radar. *J. Hydrol.* **2020**, *583*, 124615. [CrossRef]
8. Kim, G.S.; Lim, C.H.; Kim, S.J.; Lee, J.; Son, Y.; Lee, W.K. Effect of National-Scale Afforestation on Forest Water Supply and Soil Loss in South Korea, 1971–2010. *Sustainability* **2017**, *9*, 1017. [CrossRef]
9. Nguyen, T. Optimal Ground Control Points for Geometric Correction Using Genetic Algorithm with Global Accuracy Optimal Ground Control Points for Geometric Correction Using. *Eur. J. Remote Sens.* **2017**, *48*, 101–120. [CrossRef]
10. Song, C.; Yuan, L.; Yang, X.; Fu, B. Ecological-Hydrological Processes in Arid Environment: Past, Present and Future. *J. Geogr. Sci.* **2017**, *27*, 1577–1594. [CrossRef]
11. Yuan, F.; Zhang, L.; Wah Win, K.W.; Ren, L.; Zhao, C.; Zhu, Y.; Jiang, S.; Liu, Y. Assessment of GPM and TRMM Multi-Satellite Precipitation Products in Streamflow Simulations in a Data Sparse Mountainous Watershed in Myanmar. *Remote Sens.* **2017**, *9*, 302. [CrossRef]
12. Gao, P.; Jiang, G.; Wei, Y.; Mu, X.; Wang, F.; Zhao, G.; Sun, W. Streamflow Regimes of the Yanhe River under Climate and Land Use Change, Loess Plateau, China. *Hydrol. Process.* **2015**, *29*, 2402–2413. [CrossRef]
13. Li, S.; He, F.; Zhang, X.; Zhou, T. Evaluation of Global Historical Land Use Scenarios Based on Regional Datasets on the Qinghai–Tibet Area. *Sci. Total Environ.* **2019**, *657*, 1615–1628. [CrossRef]
14. Mat Nazir, M.H.; Sulaiman, W.N.A.; Juahir, H. Hydrologic Response Characteristics of a Tropical Catchment to Land Use Changes: A Case Study of The Nerus Catchment. *Environ. Earth Sci.* **2015**, *73*, 7533–7545. [CrossRef]
15. Li, Q.; Wei, X.; Zhang, M.; Liu, W.; Fan, H.; Zhou, G.; Giles-Hansen, K.; Liu, S.; Wang, Y. Forest Cover Change and Water Yield in Large Forested Watersheds: A Global Synthetic Assessment. *Ecohydrology* **2017**, *10*, e1838. [CrossRef]
16. Soulis, K.X.; Dercas, N.; Papadaki, C. Effects of Forest Roads on the Hydrological Response of a Small-Scale Mountain Watershed in Greece. *Hydrol. Process.* **2015**, *29*, 1772–1782. [CrossRef]
17. Mateus, C.; Tullos, D.D.; Surfleet, C.G. Hydrologic Sensitivity to Climate and Land Use Changes in the Santiam River Basin, Oregon. *JAWRA J. Am. Water Resour. Assoc.* **2015**, *51*, 400–420. [CrossRef]
18. Hashim, M.; Reba, N.M.; Nadzri, M.I.; Pour, A.B.; Mahmud, M.R.; Yusoff, A.M.R.M.; Ali, M.I.; Jaw, S.W.; Hossain, M.S. Satellite-Based Run-Off Model for Monitoring Drought in Peninsular Malaysia. *Remote Sens.* **2016**, *8*, 633. [CrossRef]
19. Park, E.; Merino, E.; Lewis, Q.W.; Lindsey, E.O.; Yang, X. A Pathway to the Automated Global Assessment of Water Level in Reservoirs with Synthetic Aperture Radar (SAR). *Remote Sens.* **2020**, *12*, 1353. [CrossRef]
20. FAO. *WWC towards a Water and Food Secure Future Critical Perspectives for Policy-Makers*; FAO: Rome, Italy, 2015.
21. Hofste, R.W.; Reig, P.; Schleifer, L. 17 Countries, Home to One-Quarter of the World's Population, Face Extremely High Water Stress | World Resources Institute. Available online: https://www.wri.org/insights/17-countries-home-one-quarter-worlds-population-face-extremely-high-water-stress (accessed on 26 September 2021).
22. Department of Irrigation and Drainage Malaysia. *Review of the National Water Resources Study (2000–2050) and Formulation of National Water Resources Policy*; Department of Irrigation and Drainage Malaysia: Kuala Lumpur, Malaysia, 2011.
23. Rahman, S. The Struggle for Balance's Environmental Issues, Overlaps and Future. In *Johor: Abode for Development*; Hutchinson, F.E., Rahman, S., Eds.; ISEAS Publishing: Singapore, 2020; pp. 473–500.
24. Rahman, S. Johor's Forest City Faces Critical Challenges. In *Johor: Abode for Development*; Hutchinson, F.E., Rahman, S., Eds.; ISEAS Publishing: Singapore, 2020; pp. 447–472.

25. Pakiam, G.K.; Mui, H.; Terrace, K. *Trends in Southeast Asia Agriculture in Johor: What's Left*; ISEAS Publishing: Singapore, 2018; ISBN 978-981-4818-81-0.

26. Ewing, J.; Domondon, K. *Drought, Pollution and Johor's Growing Water Needs*; ISEAS Publishing: Singapore, 2016.

27. Wang, X.G.; Su, F.Z.; Zhang, J.J.; Cheng, F.; Hu, W.Q.; Ding, Z. Construction Land Sprawl and Reclamation in the Johor River Estuary of Malaysia since 1973. *Ocean Coast. Manag.* **2019**, *171*, 87–95. [CrossRef]

28. Tan, M.L.; Ibrahim, A.L.; Yusop, Z.; Duan, Z.; Ling, L. Impacts of Land-Use and Climate Variability on Hydrological Components in the Johor River Basin, Malaysia. *Hydrol. Sci. J.* **2015**, *60*, 873–889. [CrossRef]

29. Banjir, R.; Lembangan, D.; Johor, S.; Shakir, A.; Saudi, M.; Juahir, H.; Azid, A.; Azaman, F. Flood Risk Index Assessment in Johor River Basin. *Malays. J. Anal. Sci.* **2015**, *19*, 991–1000.

30. Tan, M.L.; Chua, V.P.; Li, C.; Brindha, K. Spatiotemporal Analysis of Hydro-Meteorological Drought in the Johor River Basin, Malaysia. *Theor. Appl. Climatol.* **2019**, *135*, 825–837. [CrossRef]

31. Bin, A.; Ali, M.D. *Flood Inundation Modeling and Hazard Mapping under Uncertainty in the Sungai Johor Basin, Malaysia*; Delft University of Technology: Delft, The Netherlands, 2018.

32. Tan, M.L.; Juneng, L.; Tangang, F.T.; Chan, N.W.; Ngai, S.T. Future Hydro-Meteorological Drought of the Johor River Basin, Malaysia, Based on CORDEX-SEA Projections. *Hydrol. Sci. J.* **2019**, *64*, 921–933. [CrossRef]

33. Hutchinson, F.E. Situating Johor. In *Johor: Abode for Development*; ISEAS Publishing: Singapore, 2020; pp. 5–25.

34. Shirazi, S.M.; Adham, M.I.; Saha, R.R.; Khan, M.Z.H.; Karim, M.R. Runoff Assessment of Watershed in South Johor of Malaysia. *World J. Sci. Eng.* **2017**, *2*, 1–22.

35. Ju, Z.; Leong Tan, M.; Samat, N.; Chang, C.K. Comparison of Landsat 8, Sentinel-2 and Spectral Indices Combinations for Google Earth Engine-Based Land Use Mapping in the Johor River Basin, Malaysia. *Malays. J. Soc. Space* **2021**, *17*, 30–46. [CrossRef]

36. Segal, D.; Harvard, U. Singapore's Water Trade with Malaysia and Alternatives. In *Desalination*; Wiley: Hoboken, NJ, USA, 2004.

37. Prusty, B.A.K.; Chandra, R.; Azeez, P.A. Wetland Science: Perspectives from South Asia. In *Wetland Science: Perspectives from South Asia*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 1–587. [CrossRef]

38. Wu, H.; Adler, R.F.; Tian, Y.; Gu, G.; Huffman, G.J. Evaluation of Quantitative Precipitation Estimations through Hydrological Modeling in IFloodS River Basins. *J. Hydrometeorol.* **2017**, *18*, 529–553. [CrossRef]

39. Kite, G.W.; Pietroniro, A. Remote Sensing Applications in Hydrological Modelling. *Hydrol. Sci. J.* **1996**, *41*, 563–591. [CrossRef]

40. Fayaz, A.; Shafiq, M.U.; Singh, H.; Ahmed, P. Assessment of Spatiotemporal Changes in Land Use/Land Cover of North Kashmir Himalayas from 1992 to 2018. *Model Earth Syst. Environ.* **2020**, *6*, 1189–1200. [CrossRef]

41. Sani, D.A.; Hashim, M.; Hossain, M.S. Recent Advancement on Estimation of Blue Carbon Biomass Using Satellite-Based Approach. *Int. J. Remote Sens.* **2019**, *40*, 7679–7715. [CrossRef]

42. Thompson, D.R.; Hochberg, E.J.; Asner, G.P.; Green, R.O.; Knapp, D.E.; Gao, B.C.; Garcia, R.; Gierach, M.; Lee, Z.; Maritorena, S.; et al. Airborne Mapping of Benthic Reflectance Spectra with Bayesian Linear Mixtures. *Remote Sens. Environ.* **2017**, *200*, 18–30. [CrossRef]

43. Ahmed, B.; Kamruzzaman, M.D.; Zhu, X.; Shahinoor Rahman, M.D.; Choi, K. Simulating Land Cover Changes and Their Impacts on Land Surface Temperature in Dhaka, Bangladesh. *Remote Sens.* **2013**, *5*, 5969–5998. [CrossRef]

44. Arnold, J.G.; Fohrer, N. SWAT2000: Current Capabilities and Research Opportunities in Applied Watershed Modelling. *Hydrol. Process.* **2005**, *19*, 563–572. [CrossRef]

45. Neitsch, S.L.; Arnold, J.G.; Kiniry, J.R.; Williams, J.R. *Soil and Water Assessment Tool Theoretical Documentation Version 2009*; Texas Water Resources Institute: College Station, TX, USA, 2011.

46. Guo, T.; Engel, B.A.; Shao, G.; Arnold, J.G.; Srinivasan, R.; Kiniry, J.R. Development and Improvement of the Simulation of Woody Bioenergy Crops in the Soil and Water Assessment Tool (SWAT). *Environ. Model. Softw.* **2019**, *122*, 104295. [CrossRef]

47. Arnold, J.G.; Moriasi, D.N.; Gassman, P.W.; Abbaspour, K.C.; White, M.J.; Srinivasan, R.; Santhi, C.; Harmel, R.D.; Van Griensven, A.; Van Liew, M.W.; et al. Swat: Model Use, Calibration, and Validation. *Trans. ASABE* **2012**, *55*, 1491–1508. [CrossRef]

48. Arnold, J.G.; Youssef, M.A.; Yen, H.; White, M.J.; Sheshukov, A.Y.; Sadeghi, A.M.; Moriasi, D.N.; Steiner, J.L.; Amatya, D.M.; Skaggs, R.W.; et al. Hydrological Processes and Model Representation: Impact of Soft Data on Calibration. *Trans. ASABE* **2015**, *58*, 1637–1660. [CrossRef]

49. Tanaka, T.; Tachikawa, Y.; Iachikawa, Y.; Yorozu, K. Impact Assessment of Upstream Flooding on Extreme Flood Frequency Analysis by Incorporating a Flood-Inundation Model for Flood Risk Assessment. *J. Hydrol.* **2017**, *554*, 370–382. [CrossRef]

50. Ning, T.; Li, Z.; Liu, W. Separating the Impacts of Climate Change and Land Surface Alteration on Runoff Reduction in the Jing River Catchment of China. *Catena* **2016**, *147*, 80–86. [CrossRef]

51. Wu, S.; Li, J.; Huang, G.H. A Study on DEM-Derived Primary Topographic Attributes for Hydrologic Applications: Sensitivity to Elevation Data Resolution. *Appl. Geogr.* **2008**, *28*, 210–223. [CrossRef]

52. Nilawar, A.P.; Waikar, M.L. Use of SWAT to Determine the Effects of Climate and Land Use Changes on Streamflow and Sediment Concentration in the Purna River Basin, India. *Environ. Earth Sci.* **2018**, *77*, 783. [CrossRef]

53. Li, Z.; Deng, X.; Wu, F.; Hasan, S.S. Scenario Analysis for Water Resources in Response to Land Use Change in the Middle and Upper Reaches of the Heihe River Basin. *Sustainability* **2015**, *7*, 3086–3108. [CrossRef]

54. Cheng, C. Spatial Climate Justice and Green Infrastructure Assessment: A Case Study for the Huron River Watershed, Michigan, USA. *GI_Forum* **2016**, *4*, 176–190. [CrossRef]

55. Chen, Y.; Ale, S.; Rajan, N.; Munster, C. Assessing the Hydrologic and Water Quality Impacts of Biofuel-Induced Changes in Land Use and Management. *GCB Bioenergy* **2017**, *9*, 1461–1475. [CrossRef]

56. Anand, J.; Gosain, A.K.; Khosa, R. Prediction of Land Use Changes Based on Land Change Modeler and Attribution of Changes in the Water Balance of Ganga Basin to Land Use Change Using the SWAT Model. *Sci. Total Environ.* **2018**, *644*, 503–519. [CrossRef]

57. Swat Input/Output File Documentation, Version 2012. Available online: https://swat.tamu.edu/docs/ (accessed on 28 September 2022).

58. Abbaspour, K.C.; Yang, J.; Maximov, I.; Siber, R.; Bogner, K.; Mieleitner, J.; Zobrist, J.; Srinivasan, R. Modelling Hydrology and Water Quality in the Pre-Alpine/Alpine Thur Watershed Using SWAT. *J. Hydrol.* **2007**, *333*, 413–430. [CrossRef]

59. Abbaspour, K.C.; Vaghefi, S.A.; Srinivasan, R. A Guideline for Successful Calibration and Uncertainty Analysis for Soil and Water Assessment: A Review of Papers from the 2016 International SWAT Conference. *Water* **2017**, *10*, 6. [CrossRef]

60. Abbaspour, K.C.; Faramarzi, M.; Ghasemi, S.S.; Yang, H. Assessing the Impact of Climate Change on Water Resources in Iran. *Water Resour. Res.* **2009**, *45*, 7615. [CrossRef]

61. Abbaspour, K.C. *SWAT-CUP-2012. SWAT Calibration and Uncertainty Program—A User Manual*; Institute of Aquatic Science and Technology: Dübendorf, Switzerland, 2012.

62. Iqbal, Z.; Shahid, S.; Ahmed, K.; Ismail, T.; Nawaz, N. Spatial Distribution of the Trends in Precipitation and Precipitation Extremes in the Sub-Himalayan Region of Pakistan. *Theor. Appl. Climatol.* **2019**, *137*, 2755–2769. [CrossRef]

63. Wang, L.; Zhang, J.; Liu, P.; Choo, K.K.R.; Huang, F. Spectral–Spatial Multi-Feature-Based Deep Learning for Hyperspectral Remote Sensing Image Classification. *Soft Comput.* **2017**, *21*, 213–221. [CrossRef]

64. Monteith, J.L. Evaporation and Envrionment. In *Symposia of the Society for Experimental Biology*; Cambridge University Press (CUP): Cambridge, UK, 1965; Volume 19.

65. Shah, S.M.H.; Mustaffa, Z.; Yusof, K.W. Disasters Worldwide and Floods in the Malaysian Region: A Brief Review. *Indian J. Sci. Technol.* **2017**, *10*, 1–9. [CrossRef]

66. Moriasi, D.N.; Arnold, J.G.; Liew, M.W.V.; 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]

67. Hamidon, N.; Hong, C.J.; Awang, M.; Rahman, M.A.A.; Ahmad, F.; Musa, K.; Yusof, F.M.; Adnan, S.H.; Mustafa, M.S.S. Future Flood Forecasting in Bukit Merah Using HEC-HMS Software. In Proceedings of the Third International Conference on Separation Technology 2020 (ICoST 2020), Johor, Malaysia, 15–16 August 2020; Volume 200, pp. 183–189. [CrossRef]

68. Hashim, M. Mapping the Daily Rainfall over an Ungauged Tropical Micro-Watershed: A Downscaling Algorithm Using GPM Data. *Water* **2020**, *12*, 1661. [CrossRef]

69. Memarian, H.; Balasundram, S.K.; Abbaspour, K.C.; Talib, J.B.; Boon Sung, C.T.; Sood, A.M. SWAT-Based Hydrological Modelling of Tropical Land-Use Scenarios. *Hydrol. Sci. J.* **2014**, *59*, 1808–1829. [CrossRef]

70. Zhao, Y.; Zhang, M.X.; Cao, H.W.; Yu, X.X.; Liu, B.; Zhu, S.B.; Cheng, C.; Yin, L.X.; Xie, G. Effect of Climatic Change and Afforestation on Water Yield in the Rocky Mountain Area of North China. *For. Syst.* **2015**, *24*, e014. [CrossRef]

71. Healey, N.C.; Rover, J.A. Analyzing the Effects of Land Cover Change on the Water Balance for Case Study Watersheds in Different Forested Ecosystems in the USA. *Land* **2022**, *11*, 316. [CrossRef]

72. Khand, K.; Senay, G.B. Runoff Response to Directional Land Cover Change across Reference Basins in the Conterminous United States. *Adv. Water Resour.* **2021**, *153*, 103940. [CrossRef]

73. Sharannya, T.M.; Venkatesh, K.; Mudbhatkal, A.; Dineshkumar, M.; Mahesha, A. Effects of Land Use and Climate Change on Water Scarcity in Rivers of the Western Ghats of India. *Environ. Monit. Assess.* **2021**, *193*, 820. [CrossRef]

74. Giri, S.; Arbab, N.N.; Lathrop, R.G. Water Security Assessment of Current and Future Scenarios through an Integrated Modeling Framework in the Neshanic River Watershed. *J. Hydrol.* **2018**, *563*, 1025–1041. [CrossRef]

75. Azadi, H.; Keramati, P.; Taheri, F.; Rafiaani, P.; Teklemariam, D.; Gebrehiwot, K.; Hosseininia, G.; Van Passel, S.; Lebailly, P.; Witlox, F. Agricultural Land Conversion: Reviewing Drought Impacts and Coping Strategies. *Int. J. Disaster Risk Reduct.* **2018**, *31*, 184–195. [CrossRef]

76. Shah, M.I.; Abbas, S.; Olohunlana, A.O.; Sinha, A. The Impacts of Land Use Change on Biodiversity and Ecosystem Services: An Empirical Investigation from Highly Fragile Countries. *Sustain. Dev.* **2023**, *31*, 1384–1400. [CrossRef]

77. He, X.; Liang, J.; Zeng, G.; Yuan, Y.; Li, X. The Effects of Interaction between Climate Change and Land-Use/Cover Change on Biodiversity-Related Ecosystem Services. *Glob. Chall.* **2019**, *3*, 1800095. [CrossRef] [PubMed]

78. Obaid, H.A.; Shahid, S. Soil Erosion Susceptibility of Johor River Basin. *Water Environ. J.* **2017**, *31*, 367–374. [CrossRef]

79. Mazilamani, L.S.; Annammala, K.V.; Nainar, A.; Najib, M.Z.M. Trace Element Concentrations in Fine Sediment and Linkages to Non-Point Pollution Source: Lower Johor River Basin. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *736*, 072005. [CrossRef]

80. Pak, H.Y.; Chuah, C.J.; Yong, E.L.; Snyder, S.A. Effects of Land Use Configuration, Seasonality and Point Source on Water Quality in a Tropical Watershed: A Case Study of the Johor River Basin. *Sci. Total Environ.* **2021**, *780*, 146661. [CrossRef] [PubMed]

Article

UAV-Based Hyperspectral Imaging for River Algae Pigment Estimation

Riley D. Logan [1], **Madison A. Torrey** [2], **Rafael Feijó-Lima** [3], **Benjamin P. Colman** [3], **H. Maurice Valett** [4] and **Joseph A. Shaw** [1,*]

[1] Optical Technology Center, Electrical and Computer Engineering Department, Montana State University, Bozeman, MT 59717, USA; riley.logan@montana.edu

[2] Civil Engineering Department, Montana State University, Bozeman, MT 59717, USA; madison.torrey@montana.edu

[3] Department of Ecosystem and Conservation Sciences, University of Montana, Missoula, MT 59812, USA; rafael.feijo@mso.umt.edu (R.F.-L.); ben.colman@mso.umt.edu (B.P.C.)

[4] Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA; maury.valett@umontana.edu

* Correspondence: joseph.shaw@montana.edu

Abstract: Harmful and nuisance algal blooms are becoming a greater concern to public health, riverine ecosystems, and recreational uses of inland waterways. Algal bloom proliferation has increased in the Upper Clark Fork River due to a combination of warming water temperatures, naturally high phosphorus levels, and an influx of nitrogen from various sources. To improve understanding of bloom dynamics and how they affect water quality, often measured as algal biomass measured through pigment standing crops, a UAV-based hyperspectral imaging system was deployed to monitor several locations along the Upper Clark Fork River in western Montana. Image data were collected across the spectral range of 400–1000 nm with 2.1 nm spectral resolution during two field sampling campaigns in 2021. Included are methods to estimate chl *a* and phycocyanin standing crops using regression analysis of salient wavelength bands, before and after separating the pigments according to their growth form. Estimates of chl *a* and phycocyanin standing crops generated through a linear regression analysis are compared to in situ data, resulting in a maximum R^2 of 0.96 for estimating fila/epip chl-a and 0.94 when estimating epiphytic phycocyanin. Estimates of pigment standing crops from total abundance, epiphytic, and the sum of filamentous and epiphytic sources are also included, resulting in a promising method for remotely estimating algal standing crops. This method addresses the shortcomings of current monitoring techniques, which are limited in spatial and temporal scale, by proposing a method for rapid collection of high-spatial-resolution pigment abundance estimates.

Keywords: remote sensing; algal pigment estimation; linear regression; imaging systems; hyperspectral imaging; drones; UAVs; river algae; algal blooms; water quality; inland waters

1. Introduction

1.1. Background

In 1983, the largest collection of US Environmental Protection Agency Superfund sites was established to include the Upper Clark Fork River (UCFR), a snowmelt-fed river located in western Montana, USA. The river's headwaters were the locus of over 100 years of mining and smelting activity, the byproducts of which were repeatedly deposited along the river and its adjacent floodplain via recurrent flooding. Along with heavy metal contamination, the UCFR suffers from anthropogenic nitrogen enrichment that, coupled with naturally high levels of phosphorous availability and insolation, leads to annual nuisance blooms of filamentous algae (*Cladophora glomerata*) during periods of enhanced net primary productivity [1–3]. Furthermore, as the growing season progresses, *Cladophora*

is replaced by bluegreen algal growth (Valett et al., in review). As these dense mats of algae grow, they can disrupt dissolved oxygen and pH [4,5], reduce benthic biodiversity [6], deteriorate water quality [4,7–9], and decrease recreational opportunities [4,7,9]. To understand the ecological status, composition, and health of water bodies (i.e., water quality) in which riverine algal blooms (RABs) occur, the standing crops of various algal pigments are commonly measured, rendering information on the extent and duration of the bloom. Monitoring RAB development is particularly important in Montana, which was one of the first states to adopt water quality standards based in part on numeric algal biomass, measured as standing crops of benthic algal chl *a* [10].

Photosynthesis in algae is supported by three major classes of pigments: chlorophyll, carotenoids, and biliproteins. Among the major classes of pigments, chlorophyll *a* (chl *a*) is commonly used to assess bloom characteristics for an array of algal types, while phycocyanin is used for cyanobacteria [11,12]. Methods for measurement generally require gathering organic material in situ and processing in the laboratory. While biomass abundance can be determined by burning the sample to obtain measures of ash-free dry mass (AFDM), pigment content is determined by extraction using various solvents [13]. Current methods of assessing algal abundance via pigment and organic matter standing crops include several shortcomings: time-intensive data collection, limited spatial distribution, and often long delays between sample collection and results. In response, remote sensing has been introduced as a promising method to overcome many of the shortcomings of these traditional techniques [14].

1.2. Existing Methods

Remotely sensed imagery, from either satellite or airborne platforms, allow for rapid data collection over large areas with relatively short data processing times [15]. Satellite remote sensing has been used extensively to monitor water quality and assess algal bloom dynamics in oceans [16–22] and to study algal pigments in large lakes [21–24]. However, satellite sensing of rivers and small lakes is prevented or impaired by disadvantages that include coarse spatial resolution, long revisit times, and obscuration by clouds. Unoccupied aerial vehicles (UAVs) have provided a means of capturing high-spatial-resolution image data while also maintaining precise control over the sampling location and time, with improved repeatability and data collection times [25].

UAV-based remote sensing systems have gained popularity in recent years for monitoring algal pigment concentrations in oceans [26], lakes and reservoirs [27–29], and rivers [30–32]. In smaller running water systems, UAV-based multispectral imagers have been used to estimate pigment concentrations using spectral indices such as the normalized difference vegetation index (NDVI), which was shown to be strongly correlated with chl *a* concentration [31]. Remote sensing systems have also been successfully deployed along larger water systems, such as the UCFR, where a small UAV-based color imager with red, green, and blue (RGB) channels was used to assess percent algal cover [30]. UAV-based multispectral imagers also have been used with machine learning techniques to measure chl *a* concentration and total suspended solids in reservoirs and artificial lakes [32].

Spectral indices are commonly used to estimate biological markers in vegetation, ranging from assessing forest health and stress [33] to estimating leaf chlorophyll [34]. These indices vary from simple ratios of two spectral bands to more complex formulations involving several wavelengths. Though spectral indices have been developed for estimating chl *a* concentration in coastal waters [20], inland lakes [35], and streams [31], they have been developed to estimate concentrations within a water body (i.e., per unit volume of water) rather than per unit area. In addition, several established spectral indices have been shown to have decreased performance when generalized across systems with differing water properties [36]. Furthermore, no known studies have assessed water quality through estimation of algal standing crops (i.e., the mass of algal pigment in a given area) using UAV-based hyperspectral imaging systems over shallow rivers.

1.3. Ecological Context

While many rivers and streams are heavily canopied and their food webs are subsidized by inputs of terrestrial organic matter [37–42], others, such as streams in semi-arid systems [37] or grasslands [43], and mid-order rivers with open, well-lit channels [44], rely on autochthonous organic matter production and processing [45–47]. In these types of streams, autotrophic organisms (i.e., benthos) typically colonize and grow on the stream bottom and proliferate when warm temperatures, ample insolation, and abundant nutrients support growth. The proliferation of benthic algae influences rates of primary production and respiration, trophic relationships among consumers such as insects and fish, and alters water quality. As systems dominated by the flow of water, measures of ecosystem metabolism [48,49] and nutrient cycling in streams and rivers reflect transport and processing that typically occur over kilometers of river distance upstream from points of assessment. Accordingly, the ability to assess the spatial and temporal distributions of benthic algae is critical for understanding in-stream influences over riverine process. Current methods for quantifying algal standing crops are constrained to scales smaller (cm^2–m^2) than are needed to couple structure to functional attributes like metabolism or nutrient uptake that extend over larger scales (i.e., 100–1000 m) reflecting the influences of flow. Such constraint makes typical measures susceptible to site-specific conditions (i.e., local variation in growth conditions), rendering extrapolation to larger scales inappropriate.

1.4. Objectives

Here, this methodological knowledge gap is addressed by presenting an approach for quantifying algal abundance from image data collected by a UAV-based hyperspectral imaging system to determine standing crops of chl *a* and phycocyanin. Data were collected from the UCFR during the annual algal bloom dominated by the nuisance filamentous green algae, *Cladophora*, but include colonists that live on algal filaments (i.e., epiphytic forms) and more adenate forms associated with rock surfaces (i.e., epilithic algae), and bluegreen cyanobacteria that contribute to periphyton composition as the bloom progresses (Valett et al., in revision). Concurrently, ground-truth data were collected from designated plots to superimpose in situ sampling locations and corresponding image pixels. Using these data sets, the following research objectives were pursued:

1. present a framework for collecting high-spatial and -spectral image data using a UAV-based hyperspectral imaging system along a clear and shallow river;
2. determine optimal linear relationships between spectral band ratios (i.e., the ratio of reflectance values at two wavelengths) and the algal pigments chl *a* and phycocyanin present among filamentous, epiphytic, and epilithic forms of benthic algae;
3. compare the performance of the derived optimal linear relationships to select existing spectral indices developed for monitoring chl *a*.

This paper expands on a conference paper that provided a preliminary summary of this work [50], by including expanded discussions of the ecology of the Upper Clark Fork River; algae phenology; ground-truth data collection; imager calibration; the theoretical basis for the employed method; algorithmic development; results of ground-truth data collection; expanded modeling by separating by algal growth form, including another sampling location, and estimating phycocyanin abundance; an uncertainty analysis to aid in assessments of generalizability; and performance comparisons with several established spectral band ratios.

2. Materials and Methods

2.1. Study Sites

The UCFR is a clear and shallow cobble-bed river that forms at the confluence of Warm Springs Creek and Silver Bow Creek, with a long history of eutrophication and heavy algal growth [1–3]. Data were collected at two sites on the main stem of the river (Figure 1). The field sites, Gold Creek (46.59°N, 112.93°W) and Bear Gulch (46.70°N, 113.43°W),

feature channels of comparable width (30 and 28 m) and depth (35 cm) during the time of field sampling and both have been shown to contain elevated nutrients, likely from wastewater and other anthropogenic activities in the area [51]. Coupled to little shading from the surrounding riparian canopy, such enrichment creates conditions conducive to RAB proliferation [2].
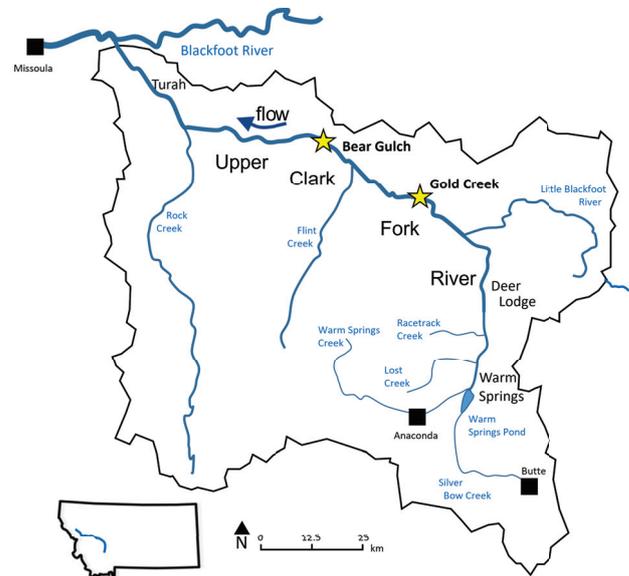
Figure 1. The Upper Clark Fork River originates at Warms Spring and flows approximately 200 km downstream to its confluence with the Blackfoot River. Stars indicate sample sites located on the main stem at Gold Creek and Bear Gulch.

2.2. RABs and Algal Phenology

Growth of *Cladophora* in the UCFR is controlled by many factors, but is typically initiated after spring runoff events in May or June [52]. Peak runoff scours benthic organic matter through shear forces created by elevated flow. Algal biomass then accrues as flow declines (Valett et al., in press) [53], temperatures increase, and *Cladophora* standing crops become visually evident [54]. The UCFR typically approaches base flow by early July, after which *Cladophora* blooms reach peak biomass [53]. As blooms progress, diatoms begin to accumulate on *Cladophora* filaments, as seen in other river systems experiencing RABs dominated by *Cladophora* [55,56]. Self-shading, grazing pressures, and thermal stress can lead to rapid extirpation and *Cladophora* decline. During this process, *Cladophora* blooms experience a shift in their surface appearance, transitioning from light green (early phases) to yellow (later stages with epiphytic coloration) to a darker brown (heavily colonized or senescent) [57]. As *Cladophora* blooms senesce in late summer and early fall, they give way to mats of bluegreen algae (Valett et al., in press), which uniquely employ the accessory pigment phycocyanin. The progression of RABs in the UCFR is spatially variable with stages of the bloom differing among locations. The visible change in *Cladophora* blooms, including accrual of epiphytic algae and change in physiological status, along with the shift to bluegreen bacteria and its exclusive pigment, phycocyanin, form the basis for remotely sensing algal standing crops of different composition and character.

2.3. Imaging System

2.3.1. Imager

Hyperspectral images were gathered using a Pika L hyperspectral imaging system (Resonon Inc., Bozeman, MT, USA), with a spectral range of approximately 387–1023 nm and nominal band spectral resolution of 2.1 nm, resulting in 300 spectral channels. During data collection, the Pika L was fitted with a 17 mm objective lens, resulting in a 17.6° across-

track full-angle field of view. This pushbroom imager was combined with an airborne imaging system that included a system control unit with onboard storage, GPS and inertial measurement unit, and downwelling irradiance sensor, forming the payload for the UAV.

2.3.2. UAV

The airborne hyperspectral system described above was mounted to a DJI Matrice 600 Pro hexacopter (DJI, Shenzhen, China) via a DJI Ronin-MX gimbal. The Matrice 600 Pro measures $1668 \times 1518 \times 727$ mm, with all frame arms, propellers, and GPS antennae unfolded, and a total weight of approximately 9.5 kg. With a total takeoff weight of 15 kg, the UAV was able to carry a payload of approximately 5.5 kg. Operation of the UAV was handled via a remote controller with a 2.4 GHz radio link and maximum transmission distance of approximately 5 km coupled with a tablet computer, allowing for autonomous flight using DJI Ground Station Pro software. In addition to the GPS system incorporated into the imaging system, location information is supplied through the onboard flight control system on the UAV via three additional GPS antennae.

2.4. Flight Overview

All UAV flights were conducted either by, or under the supervision of, a pilot certified through the Federal Aviation Administration (FAA) Part 107 licensure during each data collection campaign on 17 August and 9 September 2021. The airspace above each sampling site was verified as Class D using the FAA-certified AirMap software (AirMap, Santa Monica, CA, USA). To minimize changes in lighting conditions, flights were scheduled within two hours of solar noon, or between approximately 11:00–15:00 MDT on mostly clear days with calm wind to minimize image blurring and errors in calibration techniques caused by tilting of the UAV. Prior to flying, both the UAV and airborne imaging system were inspected and calibrated according to manufacturer specifications [58,59] to ensure proper operation and precise heading and location information. After inspection and calibration, reflectance tarps were laid out along the river bank within the field of view of the imaging system.

Flight plans were created prior to each mission using DJI Ground Station Pro software. Each flight was conducted in 300–500 m straight-line segments over the center of the river channel at an above ground level of 120 m, the maximum allowed by the FAA, and a flight speed of approximately 2.5 m/s. Flight parameters were chosen to maximize the amount of river channel captured in each image while minimizing image distortion caused by increased flying speeds and frequent aerial movements.

The airborne hyperspectral imaging system and downwelling irradiance sensor were configured prior to each flight using Resonon Ground Station software. Exposure settings on the imager were held constant throughout each flight after being set through an auto exposure routine using the 11% reflectance target illuminated by direct sunlight. Frame rate was calculated by inputting the flight height and speed into the Ground Station software. The pushbroom hyperspectral imager was set to capture a new data cube after 2000 image lines were captured, at which time the spectrometer would capture a downwelling spectral irradiance measurement.

2.5. Ground-Truth Data Collection

To build a relationship between image data collected by the hyperspectral imager and algal pigment abundance, ground-truth data were collected immediately prior to UAV flights on 17 August and 9 September 2021. On 17 August, prior to UAV flights, seven plots were established randomly on the stream bottom at each sample site. Circular plots were delineated with 112 cm sections of white hose filled with sand to ensure that spectral and pigment data were derived from the same locations. Sampling on 9 September was conducted at the Gold Creek site with 20 plots placed in four sets of five organized transects, purposefully moving from areas of low to high growth (Figure 2).

Figure 2. Sample plots and transects at the Gold Creek site on 9 September 2021. Five hoops (small white circles) were placed along four transects (red ovals, two of five shown) to establish plots moving from areas of low to high algal growth.

After image data were collected, water depth was measured and all benthic substrata (rocks, etc.) were collected within each plot. Pigment abundance was determined as areal standing crops for three growth forms: filamentous algae, epiphytes, and epilithon. Filamentous algae were removed from the surface of stones within each plot, rinsed and squeezed, and retained independently. Rinse from cleaning (i.e., squeezing and rinsing) filamentous algae was retained separately to quantify epiphytic pigments. Epilithic algae were removed using steel brushes and the resultant slurry was processed following Steinmann et al. [13]. All samples were bagged and placed in a cooler to avoid direct sunlight and keep samples chilled.

Samples were analyzed for chl *a* using 90% buffered acetone for pigment extraction and spectrophotometric assessment with acidification following the Ritchie method [60]. Phycocyanin standing crops extracted from epilithic and epiphytic samples were obtained through the use of a handheld fluorimeter (Aquafluor 8000, Turner Designs, Sunnyvale, CA, USA) following Moulton et al. [61] and Cremella et al. [62]. Calculations for final standing crops included corrections for sub-sampling and were normalized to sampling area.

For each plot, benthic organic matter (BOM) standing stocks were measured for each growth form as AFDM (g/m^2) following standard protocols [13]; dry mass was measured following 24 h at 60 °C, and ash mass after sample ignition for one hour at 550 °C, cooling and re-weighing. For slurries of epilithon, subsamples were filtered through pre-weighed Whatman GF/F (0.7 μm pore size) filters. For filamentous algae, subsamples of wet weight were used for drying and combustion. One corrupted data point was removed from the 17 August Gold Creek data set.

2.6. Image Preprocessing

2.6.1. Image Calibration

Hyperspectral data captured by the imaging system have a native pixel format of raw digital counts. Calibrating the pixel values to a common physical scale is key to achieving accurate image processing and comparison among imaging systems. Image calibration was performed using two methods: reflectance targets laid out along the river bank and a downwelling irradiance sensor mounted on the UAV.

The first of these methods used two G8T portable fabric reflectance targets (Group 8 Technology Inc., Provo, UT, USA). The reflectance standards had spectral reflectances of approximately 6% and 11% from 400–1000 nm, with sizes of 1.2 × 1.2 m and 2 × 2 m, respectively. Both tarps were laid out in flat areas along the river bank in locations that would be captured by the imager during UAV flights. As the tarps were contained in only a

portion of a single image, the pixels containing the tarps were used as a reflectance standard from which all other image pixels were calibrated throughout the flight using Equation (1):

$$\rho(\lambda)_{scene} = \frac{L(\lambda)_{scene} - L(\lambda)_{dark}}{L(\lambda)_{tarp} - L(\lambda)_{dark}} \rho(\lambda)_{tarp}, \tag{1}$$

where $\rho(\lambda)_{scene}$ is the spectral reflectance of the scene, $L(\lambda)_{scene}$ is the spectral radiance of the imaged scene, $L(\lambda)_{tarp}$ is the spectral radiance of the reflectance standard, $L(\lambda)_{dark}$ is the dark current signal on the image sensor, and $\rho(\lambda)_{tarp}$ is the spectral reflectance of the standard. Though this method is an accepted calibration technique, it relies on several assumptions, such as constant lighting conditions throughout the flight and pixel-to-pixel response consistency.

The second method removed the assumptions implicit in the first method by using an Ocean Insight Flame UV-VIS spectrometer (Ocean Insight, Orlando, FL, USA) fitted with a cosine corrector and calibrated to measure downwelling spectral irradiance from 400–1000 nm with nominal spectral resolution of 1.34 nm. The spectrometer was configured to synchronously measure spectral downwelling irradiance with each hyperspectral data cube captured by the imager. This method allowed each hyperspectral data cube to be converted to reflectance with precise illumination information throughout the flight and thereby removed the assumption of unchanging illumination, using

$$\rho(\lambda)_{scene} = \frac{L(\lambda)_{scene} - L(\lambda)_{dark}}{\frac{E(\lambda)_{downwelling}}{\pi} - L(\lambda)_{dark}}, \tag{2}$$

where $E(\lambda)_{downwelling}$ is the downwelling spectral irradiance and π is the hemispheric projected solid angle.

2.6.2. Pixel Selection

After the hyperspectral images were converted from digital counts to reflectance, spatial pixels within each plot that were not saturated by sun glint were averaged using Resonon Spectronon software. For a 120 m flight altitude with a 17 mm focal length lens, the across-track image contained 900 pixels in a 17.6-degree field of view, such that each plot contained an average of approximately 40 pixels after flying over each plot. The spectra contained in each of these spatial pixels was averaged to create a single spectrum for each plot.

2.7. Image Processing

Hyperspectral image data were processed by isolating the signal reflected from the stream bed harboring algal growth, converting image data to physical units of reflectance, and developing and applying spectral indices following the steps outlined below.

2.7.1. Theoretical Basis for Isolating Stream Bed Reflectance

The goal of the image processing method presented was to relate the upwelling radiance measured by the hyperspectral imaging system to the pigment abundance contained within benthic algal blooms. Following a development similar to that presented by Legleiter, et al. [63,64], total upwelling spectral radiance [W/(m² sr nm)] measured by the UAV-based imaging system can be written as

$$L_{tot}(\lambda) = L_b(\lambda) + L_{wb}(\lambda) + L_{ws}(\lambda) + L_{atm}(\lambda), \tag{3}$$

where L_{tot} is the total upwelling spectral radiance, L_b is the spectral radiance reflected from the benthos, L_{wb} is the spectral radiance scattered to the imager from the water body, L_{ws} is the spectral radiance reflected from the water's surface, L_{atm} is the spectral radiance scattered from the atmosphere, and all terms are a function of wavelength, λ. Each term can also have variable polarization, which creates radiometric errors if the polarization-sensitive instrument response is ignored [65,66]; however, this becomes negligible for the small incidence angles (less than 8°) employed in this work. Of the terms addressed in

Equation (3), L_b is of primary interest when relating image data to a difference in benthic algal pigment contents. Owing to the nature of the UCFR and UAV-based data collection, several assumptions can be made to simplify Equation (3) and isolate the signal of interest.

With the exception of stormflow and snowmelt conditions, the UCFR is clear and shallow, meaning the majority of the downwelling solar irradiance will be reflected from the stream bed itself, making L_b the dominant signal reaching the imager. Upwelling radiance from optical scattering within the water body prior to reaching the bottom substrate, L_{wb}, largely depends on the water's inherent optical properties. These generally include the absorption coefficient, scattering coefficient, and volume scattering function, which are functions of both depth and dissolved and suspended matter present in the stream [67,68]. Owing to the clarity and shallowness of the UCFR, radiance generated through L_{wb} can be considered negligible when compared to the relatively large signal of L_b. The signal generated by L_{ws} largely depends on the reflectance of the water's surface, which is a function of surface conditions and stream morphology. The reflectance of the water's surface has been measured as low as 0.03 between 400–800 nm, with reflectance values approaching 0 for calm surfaces [63], making L_{ws} small compared to L_b. The final term, L_{atm}, typically becomes negligible when optical signals take relatively short paths through a clear atmosphere, such as UAV operation at 120 m on a clear day. However, under smoky conditions, which often occur during the summer in Montana, atmospheric radiance may increase. Radiative transfer simulations show that elevated atmospheric scattering driven by smoke generates a spectrally-flat reduction in contrast from 400–1000 nm; however, the increased L_{atm} is still much smaller than the signal generated by L_b.

Taking these assumptions and simplifications into account, total upwelling radiance reaching the imager can be considered to be dominated by upwelling radiance from the benthos and Equation (3) can be rewritten only in terms of L_b as

$$L_{tot} \approx L_b = \frac{E_d T_{aw} T_{wa} R_b e^{-2\kappa d}}{\pi}, \tag{4}$$

where E_d is the downwelling irradiance [W/m^2]; T_{aw} is the air-water interface transmittance; T_{wa} is the water-air interface transmittance; R_b is the reflectance of the stream bed; κ is the optical extinction coefficient; d is the depth of the water body, which must be considered on both its downwelling and upwelling path through the water; and the factor of pi represents the bidirectional reflectance distribution function for a Lambertian surface. Transmission across the air-water interface occurs twice, once when the optical signal enters the water (T_{aw}) and again when the scattered sunlight leaves the water (T_{wa}). For small incidence angles, the transmittance across this interface is equal when moving between the media; however, for large incidence angles these terms may need to be considered separately.

As the primary signal of interest for identifying algal pigment standing crops is the stream bed reflectance, Equation (4) can be rewritten as

$$R_b = \frac{R_{tot}}{T_{aw} T_{wa} e^{-2\kappa d}}, \tag{5}$$

where R_{tot} is the reflectance measured at the imager, or L_{tot}/E_d. Equation (5) provides a basis for isolating the reflectance of the stream bed to accurately assess changes in pigment abundance.

Calculating or measuring the extinction coefficient, κ, enables correction for the effects of optical absorption and scattering within the water as a function of wavelength and depth. Accounting for these losses may allow for the use of wavelengths that are strongly absorbed by water, which contain important pigment information. However, correcting for the effects of water extinction requires continuous measurement of the depth of the studied water body, making the method largely inaccessible for most remote sensing applications. The results presented here do not include an extinction correction because it is less important for shallow water and it is preferred to present a method that is more readily implemented by others using the equipment presented herein.

2.7.2. Development of Spectral Indices and Application to Algal Standing Crops

Using methods similar to those presented in earlier works [36,64], a brute-force analysis was performed to relate hyperspectral reflectance data to in situ samples of chl *a* and phycocyanin. Reflectance data across the spectral range of 400–850 nm from each measurement plot were extracted to form all possible band ratio combinations of the form $R_{\lambda_1}/R_{\lambda_2}$, where R_λ represents the reflectance at a single 2.1 nm-wide spectral channel. Data were omitted for wavelengths longer than 850 nm, where the signal-to-noise ratio declines rapidly because of increasing water absorption and decreasing detector responsivity. The resulting values were matched to corresponding in situ pigment measurements and linear regressions performed to relate pigment standing crops to each band ratio's spectral signal. The goodness of fit for each model was assessed via the coefficient of determination (R^2) for each data set. The approach generated a large number (45,796) of band ratios and R^2 values for each data set analyzed. Data sets containing the band ratios and the resultant R^2 values were analyzed with frequency distributions to characterize the spread of potentially useful band ratios. A matrix of R^2 values for each band ratio combination was developed to depict correlation intensity across all band ratio combinations. The matrix was used to highlight areas sensitive to the target pigment, from which the band ratio with the highest R^2 value was saved as the optimal band ratio for the data set.

With this approach, two major factors influencing the identity of the appropriate spectral signal (i.e., band ratio) and the quality of the linear models employed to estimate pigment standing crops were addressed. First, independent models were generated for different growth forms including: (i) filamentous plus epiphytic (fila/epip), (ii) epilithic, and (iii) total (filamentous, epiphytic, and epilithic). Models estimating phycocyanin abundance were restricted to epiphytic and epilithic growth, as no bluegreen bacteria were associated with *Cladophora* filaments. Second, the influence of RAB development over time was addressed by comparing combined and independent models derived from multiple dates (17 August, 9 September). Finally, the combined influence of form and stage was analyzed by generating models specific to individual growth forms for different sampling dates.

2.8. Uncertainty Analysis

The optimal band ratios selected by the brute-force method were based on correlating a small number of pigment observations (n = 33) to spectral models with a large number of potential explanatory variables (n = 45,796), creating conditions that could generate spurious relationships. To test the generalizability and uncertainty around all band ratios, including those identified as optimal by the brute-force method, a resampling approach was used. In this approach, the 33 pigment observations were subsampled into groups of 26 (representing 80% of the observations) and these subsamples were repeated 1000 times without replacement. This resampling approach was chosen as it would allow comparison between consistently sized subsamples of the data. For each subsample, a linear regression was performed between each of the 45,796 possible band ratios and the subsample of pigments, generating 45,796,000 linear models and R^2 values (1000 for each band ratio). The average R^2 value was calculated for each band ratio across all 1000 resamples. With this method, the performance of each of the optimal band ratios identified by the brute-force analysis could be assessed against each pigment subsample, giving a distribution of R^2 values and a standard deviation around the mean R^2 value for each pigment growth form. The subsampling analysis was run for total standing crops of chl *a* and phycocyanin, as well as after separating by growth form.

3. Results

3.1. RAB Characteristics

When plots were randomly established at both Gold Creek and Bear Gulch, total chl *a* standing crop ranged from a high of 316.5 mg/m^2 to a low of 69.3 mg/m^2 and averaged 191.4 ± 33.1 mg/m^2 and 183.5 ± 31.8 mg/m^2 at Gold Creek and Bear Gulch, respectively

(Table 1). When samples at Gold Creek were purposefully distributed to include transects into the primary RAB, total chl *a* standing crops ranged from a high of 377.8 mg/m^2 to a low of 46.1 mg/m^2 and averaged 186.0 ± 30.3 mg/m^2. The majority of chl *a* was found in standing crops of filamentous growth across all field sites and sampling dates. Epilithic chl *a* standing crops tended to be greater than those of epiphytic chl *a* at Gold Creek, whereas Bear Gulch contained similar levels of each. Total standing crop of BOM was highest on 17 August at Gold Creek (97.4 g AFDM/m^2), including a large proportion (88%) of filamentous growth. Between 17 August and 9 September, total BOM decreased at Gold Creek, largely due to a decrease in filamentous forms. Bear Gulch showed the lowest standing stocks of filamentous BOM, but larger standing stocks of epiphytic and epilithic BOM compared to the Gold Creek data sets. Total phycocyanin standing crops were similar between field sites; however, abundance was distributed almost evenly between epiphytic and epilithic growth forms at Bear Gulch whereas the majority of phycocyanin standing crops were contained in epilithon at Gold Creek.

Table 1. Standing crop of chlorophyll a (chl *a*), benthic organic matter (BOM), and phycocyanin at Gold Creek and Bear Gulch. Data are means ± standard error; n = 6 at Gold Creek on 17 August, n = 7 at Bear Gulch on 17 August, and n = 20 at Gold Creek on 9 September.

Field Site	Field Date	Growth Form	chl *a* (mg/m^2)	BOM (g FDM/m^2)	Phycocyanin (μg/m^2)
Gold Creek	17 August	Filamentous	158.9 ± 30.3	85.8 ± 13.3	NA
		Epiphytic	7.3 ± 2.8	4.2 ± 0.4	6.5 ± 2.2
		Epilithic	25.2 ± 2.2	8.8 ± 13.9	31.3 ± 4.6
		Total	191.4 ± 33.1	97.4 ± 13.9	37.8 ± 4.6
	9 September	Filamentous	149.5 ± 24.0	67.4 ± 11.0	NA
		Epiphytic	10.0 ± 2.5	5.0 ± 1.7	7.2 ± 2.4
		Epilithic	26.6 ± 3.8	9.9 ± 1.3	27.2 ± 4.6
		Total	186.0 ± 30.3	82.4 ± 12.1	34.4 ± 5.5
Bear Gulch	17 August	Filamentous	124.8 ± 25.8	55.5 ± 14.8	NA
		Epiphytic	26.8 ± 9.9	17.2 ± 3.7	20.8 ± 5.2
		Epilithic	31.9 ± 5.2	15.4 ± 2.5	21.7 ± 2.6
		Total	183.5 ± 31.8	88.1 ± 18.1	42.5 ± 5.3

3.2. Spectral Models

Estimative modeling began by running the brute-force analysis using the combined data set, composed of pigment data collected across all field sites and dates and averaged spectra from corresponding plots. Each plot consisted of a single spectrum obtained by averaging the spectra contained in all hyperspectral pixels within the plot (Figure 3). The model was run using total chl *a* standing crops (i.e., sum of all growth forms) and total phycocyanin abundance, generating three output products (Figure 4): a heat map of coefficient of determination (R^2) values for every possible band ratio (Figure 4a), a histogram showing the counts of R^2 values (Figure 4b), and a linear regression between the optimal band ratio and the targeted pigment (Figure 4c).

The heat maps for each regression represent the resulting R^2 value generated for every possible band ratio, created from a particular wavelength as denominator (y-axis) and numerator (x-axis). The heat maps allow for a quantifiable means of selecting the optimal band ratio (i.e., The band ratio with the highest R^2 value) and determining spectral regions which are sensitive to the targeted pigment. Next, the histograms show the salience of the regressions generated by the brute-force analysis by displaying the number of band ratios which generated the R^2 values shown in the heat maps. Finally, the linear regressions for each data set are a graphical representation of the fit between the best-performing band ratio, selected by the brute-force algorithm, and the pigment under analysis. For each linear

regression, the field location and date from which the ground-truth pigment standing crop was measured is designated with a unique marker and color.
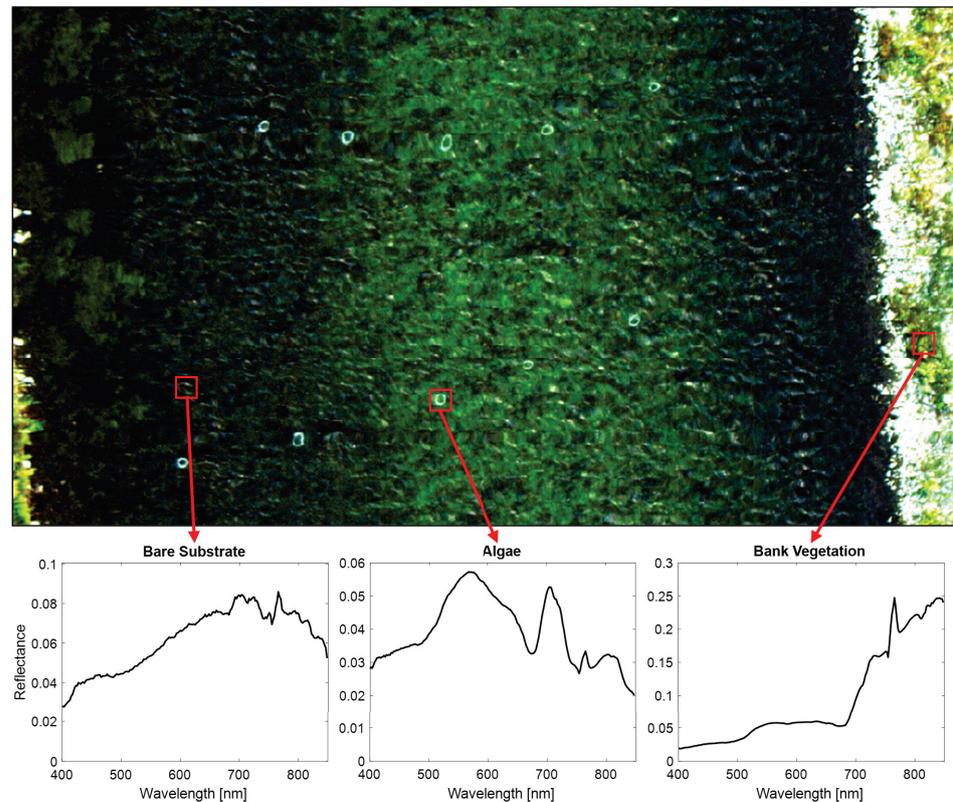

Figure 3. RGB image created from one of the hyperspectral image cubes gathered using the UAV system at Gold Creek. Examples of the complete spectra contained at each pixel are shown for bare river substrate, Cladophora algae, and bank vegetation for the spectral range of 400–850 nm. Note the varying reflectance axes and strong water absorption beyond approximately 725 nm that suppresses the strong near-infrared reflectance of the submerged algae.

A band ratio of 684/674 nm was selected as the optimal ratio for estimating chl *a* standing crops, which is likely based on the known chl *a* absorption line near 665 nm [69,70]. Of the 45,796 possible band ratios, the optimal band ratio was one of eight which produced an R^2 above 0.40 (Figure 4b). Analysis of the combined data set showed promising results when estimating chl *a* standing crops, with an $R^2 = 0.57$ and root-mean-square error (RMSE) = 66.29 mg/m^2. Regression analysis of phycocyanin standing crops resulted in worse estimation performance, with an $R^2 = 0.41$ and RMSE = 15.54 mg/m^2. An optimal band ratio of 753/824 nm was selected, highest among 10 band ratios which resulted in an R^2 above 0.35 (Supplementary Figure S1).

3.2.1. Separation by Growth Form

Though the combined data set showed promising results, the optical signal received by the imaging system was likely dominated by either large mats of filamentous algae obscuring the epilithic growth or areas of solely epilithic growth with minimal filamentous algae. To explore this possibility, the pigment abundances related to the filamentous algae and epiphytes were separated from the abundances generated from epilithic sources, and each data set was used as input to the brute-force analysis using the same imagery, with zero-valued points removed. As epiphytes tend to collect and grow on the surface of filamentous algae, the optical signals for these growth forms were likely conflated; therefore, the chl *a* standing crops associated with these two forms were combined into a single group.

(a)

(b)

(c)
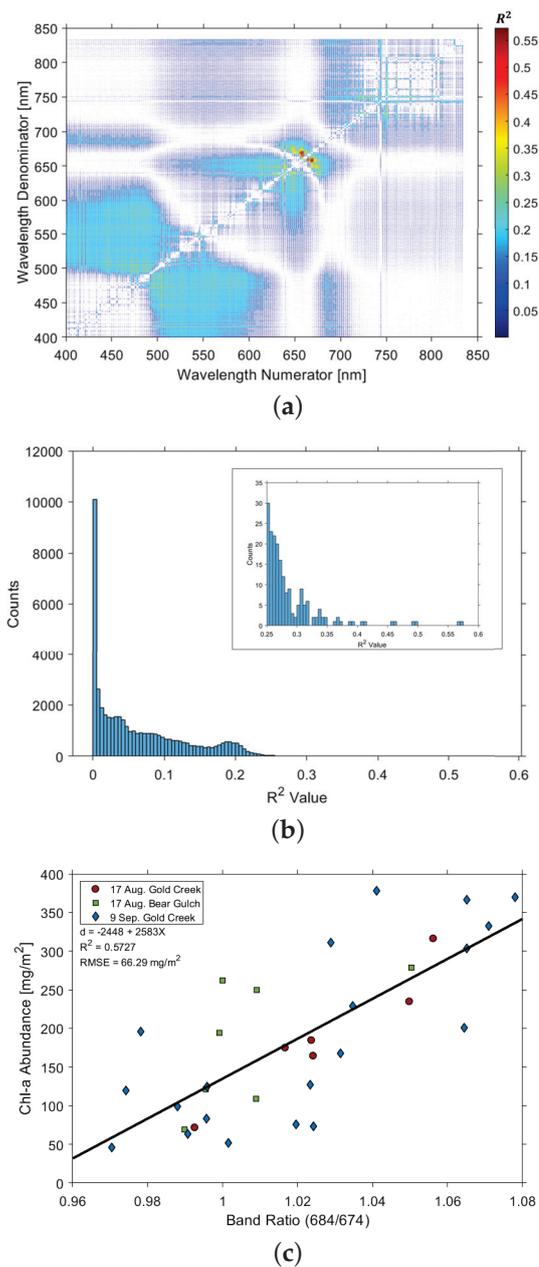
Figure 4. Analytics from the regression analysis of total chl *a* standing crops (sum of filamentous, epiphytic, and epilithic sources) from the combined (all field sites) data set; (**a**) Heat map of R^2 values generated by fitting each band combination against total chl *a* standing crops. Colors within the heat map represent the R^2 value, with wavelength numerators and denominators shown on the x- and y-axes, respectively; (**b**) Histogram of the R^2 values generated using the brute-force analysis. The insert shows counts beyond an $R^2 = 0.25$, showing that relatively few band ratios generate fits in this region; (**c**) Linear regression analysis between chl *a* abundance (y-axis) and optimal band ratio (x-axis). Data collected on 17 August 2021 at the Gold Creek and Bear Gulch site represented by red circles and green squares, respectively, while data collected on 9 September 2021 are shown with blue diamonds.

Filamentous Plus Epiphytic

After separating each data set, the strength of relationships between target pigment abundance and image data were compared to the combined data set's performance when estimating total pigment standing crops. The difference in model performance is described as ΔR^2, which is positive for increases in R^2 and negative for decreases. The performance

of the regression analysis modestly improved when estimating fila/epip chl *a* standing crops individually, with a ΔR^2 of 0.05, increasing from 0.57 to 0.62, while maintaining an optimal band ratio of 684/674 nm. With improved performance, the salience of this ratio also increased; the optimal band ratio was one of six generating an R^2 above 0.50 (Supplementary Figure S2).

Estimation of phycocyanin standing crops was limited to epiphytic and epilithic growth forms as phycocyanin was not associated with the filamentous growth forms present in the UCFR. Estimative performance of epiphytic phycocyanin standing crops decreased compared to the combined data set estimates, with a ΔR^2 of −0.08. The decreased performance was associated with an optimal band ratio similar to those selected for chl *a* estimation, with a band ratio of 688/655 nm (Supplementary Figure S3).

Epilithic

Estimation of epilithic chl *a* standing crops decreased compared to the combined data set, with a loss in R^2 resulting in a ΔR^2 of −0.13, decreasing from 0.57 to 0.44 while the optimal band ratio shifted to longer wavelengths of 729/809 nm (Supplementary Figure S4). Performance for estimating epilithic phycocyanin standing crops improved compared to the combined data set, with R^2 values increasing from 0.41 to 0.50, resulting in a ΔR^2 of 0.09 while using an optimal band ratio of 738/804 nm (Supplementary Figure S5).

The improved strength of the relationship for estimating chl *a* in fila/epip growth forms separately and decrease in performance for epilithic chl *a* suggests that the optical signal from filamentous and epiphytic algal growth dominates the imaged scene during a large bloom. The increased performance for epilithic phycocyanin is likely due to the larger abundance of phycocyanin commonly found in epilithic algal growth. Those pixels which contain epilithic growth, which are not obscured by filamentous growth, are likely better estimated when separated by growth form.

3.2.2. Separation by Phenology

Sources of algal biomass change throughout the summer growing season, with the ratio of biomass sourced from epilithon, epiphytes, and filamentous algae shifting as a function of stream conditions. This change has been observed to depend on sampling location along the UCFR, with field sites farther downstream progressing through their life cycle prior to those closer to the headwaters. Compared to both Gold Creek sites, the Bear Gulch site visually appeared to contain elevated epiphyte colonization relative to filamentous organic matter. To explore the role phenology had in band ratio selection, the data collected from the two field sites were separated, forming a Bear Gulch data set (n = 7) and Gold Creek data set (n = 26). These two data sets were used as inputs to the brute-force analysis to estimate standing crops from all pigment sources, epiphytic and filamentous sources, and epilithon.

Bear Gulch
Total

Separation based on phenology led to improvements in the overall strength of the regression relationship, with performance increasing for all pigments and sources. When estimating total chl *a*, the brute-force analysis selected a band ratio of 554/536 nm, resulting in a ΔR^2 of 0.38, increasing from the combined model's R^2 value of 0.57 to 0.95 (Supplementary Figure S6). Estimation of total phycocyanin at Bear Gulch shifted the optimal band ratio from 753/824 nm in the combined model to 625/619 nm (Supplementary Figure S7). The associated ΔR^2 was 0.51, with R^2 values increasing from 0.41 to 0.92.

Filamentous plus Epiphytic

Improvements in the estimation of the fila/epip and epilithic groups were also seen after separating the data by location. When estimating fila/epip chl *a* at Bear Gulch, the

brute-force analysis selected a shorter wavelength ratio of 554/536 nm with a ΔR^2 of 0.39 associated with an R^2 of 0.96 (Supplementary Figure S8). Similarly, the estimation of epiphytic phycocyanin improved, with a ΔR^2 of 0.53, after R^2 values increased from 0.41 to 0.94 and selecting an optimal band ratio of 748/740 nm (Supplementary Figure S9).

Epilithic

Epilithic pigment estimation accuracy also increased for both chl *a* and phycocyanin, with R^2 values improving from 0.57 to 0.86 and 0.41 to 0.89, respectively, leading to ΔR^2 values of 0.29 and 0.48 compared to the combined model (Supplementary Figure S10). Estimation of epilithic chl *a* from the Bear Gulch data set relied on shorter-wavelength ratios as compared to the combined data set, selecting a band ratio of 573/589 nm. The ratios selected for epilithic phycocyanin also shifted to shorter wavelengths, using a band ratio of 525/519 nm (Supplementary Figure S11).

Gold Creek

Total

Analyzing the Gold Creek data separately resulted in improved estimates compared to the combined data set across all growth forms. The strength of the relationship between total chl *a* and image data at Gold Creek improved from an R^2 of 0.57 to 0.63 ($\Delta R^2 = 0.06$) while for total phycocyanin, R^2 values increased from 0.41 to 0.46, resulting in a ΔR^2 of 0.05. The optimal band ratio for estimating total chl *a* from the Gold Creek data set remained the same between the data sets, maintaining a band ratio of 684/674 nm (Supplementary Figure S12). The optimal band ratio for total phycocyanin showed a near-inverse relationship compared to the combined data set, shifting from 753/824 nm to 804/748 nm (Supplementary Figure S13).

Filamentous plus Epiphytic

The strength of the linear regression was highest at Gold Creek when estimating fila/epip chl *a*, with a ΔR^2 of 0.11, after R^2 values increased from 0.57 to 0.68 (Figure 5). Notably, the optimal band ratio remained the same for the combined data set and the Gold Creek data set. This band ratio is the vicinity of known chl *a* absorption near 665 nm, suggesting the brute-force analysis relied on a change in chl *a* abundance when estimating total and fila/epip chl *a* standing crops. Estimation of epiphytic phycocyanin also showed modest improvements, increasing from an R^2 of 0.41 to 0.43, an improvement associated with a ΔR^2 of 0.02 and the selection of a band ratio of 778/774 nm (Supplementary Figure S14).

Epilithic

The regression analysis for estimating epilithic pigmentation also improved for chl *a* (Supplementary Figure S15) and phycocyanin (Supplementary Figure S5), with R^2 values increasing from 0.57 to 0.62 and 0.41 to 0.59, respectively, leading to ΔR^2 values of 0.05 and 0.18. The optimal band ratio for estimating epilithic phycocyanin shifted to longer wavelengths, selecting an optimal band ratio of 804/740 nm.

The correlation maps for the Gold Creek data set show that wavelengths surrounding chl *a* absorption lines are favored when estimating total chl *a* abundance, fila/epip chl *a*, and epiphytic phycocyanin. However, longer-wavelength band ratios are preferred when estimating the remaining pigments and forms, often selecting wavelengths over 700 nm.

3.2.3. Results Summary

Separating the combined data set by growth form (pigment source) and phenological state (field site) generally resulted in improved performance for estimating both chl *a* (Table 2) and phycoycanin (Table 3) standing crops, with the exception of epilithic chl *a* for which performance dropped after growth form separation. The strongest relationship between band ratio and pigment abundance was found after separating by phenological state and estimating pigment standing crops at the Bear Gulch. Although the linear

regression models at the Bear Gulch site produced R^2 values above 0.86, this field site was limited to seven data points, which may introduce problems such as spurious fits. Optimal band ratio selection was fairly robust when estimating chl *a* abundance at Bear Gulch, converging on a band ratio of 554/536 nm for total and fila/epip chl *a*; however, optimal band ratio selection for phycocyanin abundance was more variable.

Table 2. Model performance and optimal band ratio selected for chl *a* abundance after separating data sets by growth form and phenological state. The ΔR^2 value describes the change in R^2 after separating the data set, relative to the performance of the combined data set estimating total pigment abundance.

Data Set	Growth Form	Optimal Band Ratio	R^2	ΔR^2
Combined	Total	684/674	0.57	-
	Fila/epip	684/674	0.62	0.05
	Epilithic	729/809	0.44	−0.13
Bear Gulch	Total	554/536	0.95	0.38
	Fila/epip	554/536	0.96	0.39
	Epilithic	573/589	0.86	0.29
Gold Creek	Total	684/674	0.63	0.06
	Fila/epip	684/674	0.68	0.11
	Epilithic	804/738	0.62	0.05

Table 3. Model performance and optimal band ratio selected for phycocyanin abundance after separating data sets by growth form and phenological state. The ΔR^2 value describes the change in R^2 after separating the data set, relative to the performance of the combined data set estimating total pigment abundance.

Data Set	Growth Form	Optimal Band Ratio	R^2	ΔR^2
Combined	Total	753/824	0.41	-
	Epiphytic	688/655	0.33	−0.08
	Epilithic	738/804	0.50	0.09
Bear Gulch	Total	625/619	0.92	0.51
	Epiphytic	748/740	0.94	0.53
	Epilithic	525/519	0.89	0.48
Gold Creek	Total	804/748	0.46	0.05
	Epiphytic	778/774	0.43	0.02
	Epilithic	804/740	0.59	0.18

The strongest relationship using the larger Gold Creek data set was found after separating the data sets by growth form and analyzing epilithic phycocyanin and fila/epip chl *a* standing crops (Figure 5). Estimation of epilithic phycocyanin abundance produced an R^2 of 0.59 and utilized a band ratio of 804/740 nm (Figure 5b,d,f). Analysis of fila/epip chl *a* produced an R^2 of 0.68 and converged on the optimal band ratio of 684/674 nm, found in several chl *a* abundance measurements for the Gold Creek and combined data sets (Figure 5a,c,e). When applied at Bear Gulch, the optimal band ratio for the combined and Gold Creek data produced reasonable fits when estimating chl *a* standing crops. Fitting to total and fila/epip chl *a* standing crops, the band ratio of 684/674 nm produced R^2 values of 0.37 and 0.36, respectively, meaning this band ratio may carry meaningful information across field sites (i.e., phenological states).
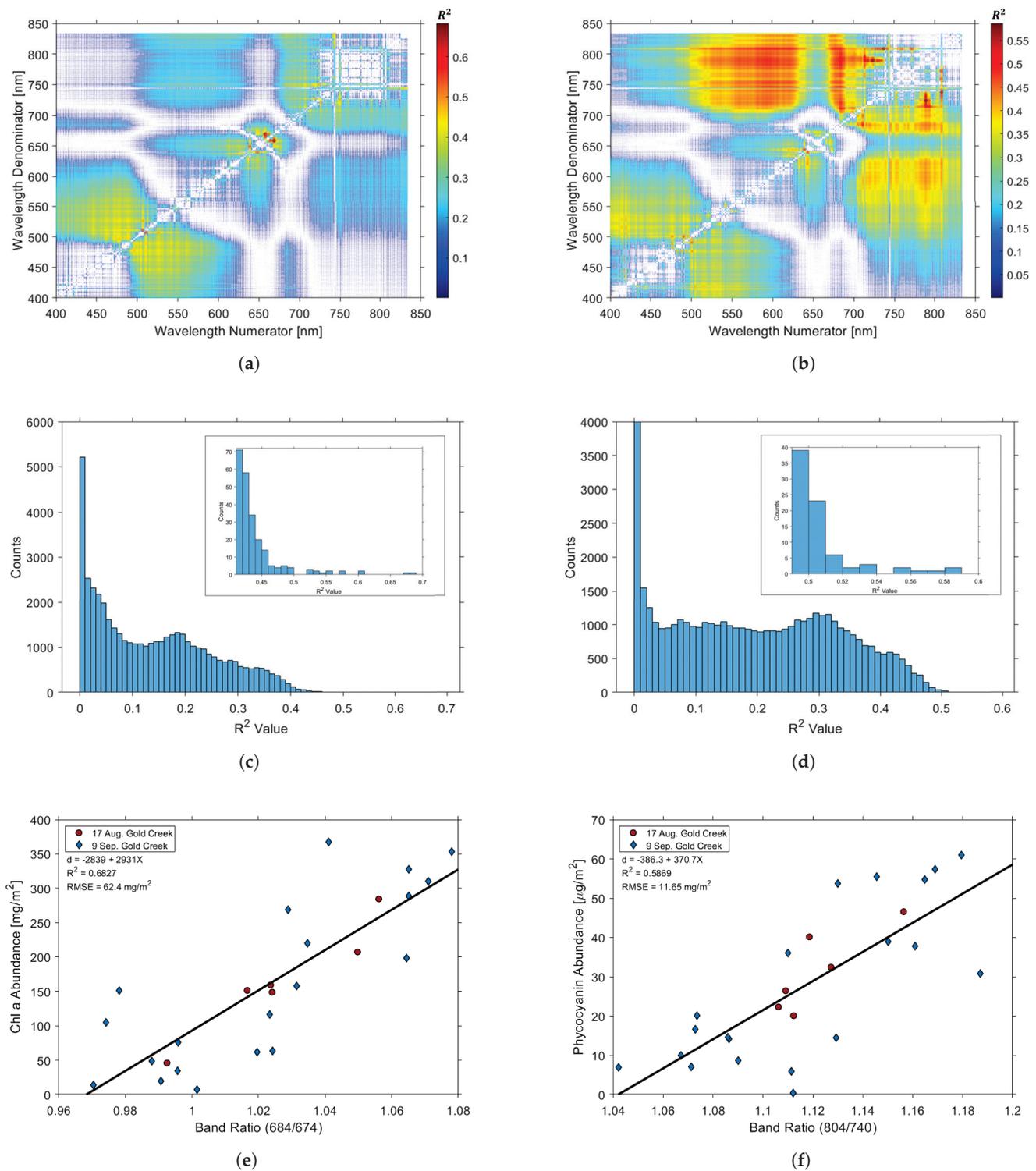
Figure 5. Linear regressions, correlation maps, and histograms for the Gold Creek data set. Results gathered via the brute-force analysis for estimating fila/epip chl *a* abundance (left column) and epilithic phycocyanin abundance (right column); (**a**) Gold Creek fila/epip chl *a* correlation map; (**b**) Gold Creek epilithic phycocyanin correlation map; (**c**) Gold Creek fila/epip chl *a* histogram; (**d**) Gold Creek epilithic phycocyanin histogram; (**e**) Gold Creek fila/epip chl *a* linear regression; (**f**) Gold Creek epilithic phycocyanin linear regression.

Due to its strong performance and frequent selection as optimal, the linear equation obtained from the band ratio of 684/674 was used to estimate pigment abundances for the top-performing data set, found for fila/epip chl *a* measurements at the Gold Creek field site, and compared to the measured abundances (Figure 6). In general, the model showed promising results, but tended to overestimate low chl *a* abundances (below $\approx 100 \text{ mg/m}^2$) and underestimate elevated abundances (above $\approx 250 \text{ mg/m}^2$).
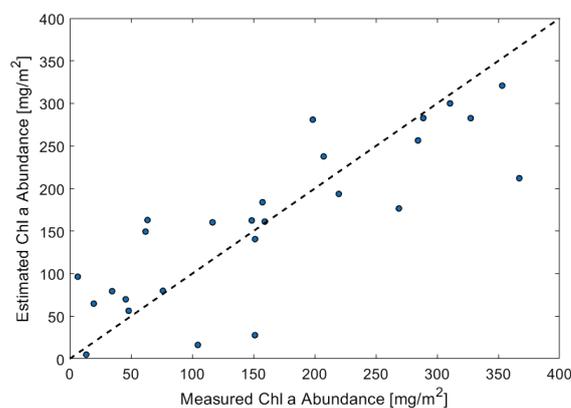
Figure 6. Estimated versus measured fila/epip chl *a* abundance at the Gold Creek field site using the regression analysis generated by the brute-force method. Points falling along the 1:1 line (dashed) represent perfect performance.

3.3. Generalizability of Band Ratios

The resampling approach identified the same optimal band ratios and similar average R^2 values as the brute-force method for four out of the six combinations of pigment and growth form, and identified the inverse band ratio for one. For all of the chl *a* growth forms, the same optimal band ratios were identified by both the brute-force method and resampling 80% of the observations 1000 times, and the mean R^2 value from the resampling approach was identical to that identified by the brute-force method on all of the samples (Table 4). Additionally, the resulting data distribution of the R^2 values across all 1000 resamples for the optimal band ratios were tightly constrained (Figure 7), with standard deviations between 0.05 and 0.06. For phycocyanin, the resampling approach selected the same optimal band ratio for total abundance, the inverse band ratio for epilithic (i.e., brute force, 738/804; resampling, 804/738), while both band ratios had R^2 values that were nearly identical between the resampling and brute-force approaches and small standard deviations. In contrast, the resampling approach identified a different optimal band ratio for the epiphytic growth form for phycocyanin (i.e., brute force, 688/655; resampling, 684/674) and the R^2 value was higher for the resampling approach for this alternative band ratio. If the R^2 is examined from the resampling approach for the optimal band ratio identified by the brute-force method, a mean R^2 of 0.18 ± 0.19 is generated.

Table 4. Optimal band ratio, mean R^2 value, and standard deviation identified after running the uncertainty analysis on total pigment abundance and after separating by growth form.

Pigment	Growth Form	Optimal Band Ratio		R^2	
		Brute Force	**Resampling**	**Brute Force**	**Resampling**
chl *a*	Total	684/674	684/674	0.57	0.57 ± 0.06
	Fila/epip	684/674	684/674	0.62	0.62 ± 0.05
	Epilithic	729/809	729/809	0.44	0.44 ± 0.06
Phycocyanin	Total	753/824	753/824	0.41	0.42 ± 0.09
	Epiphytic	688/655	684/674	0.33	0.59 ± 0.22
	Epilithic	738/804	804/738	0.50	0.51 ± 0.06
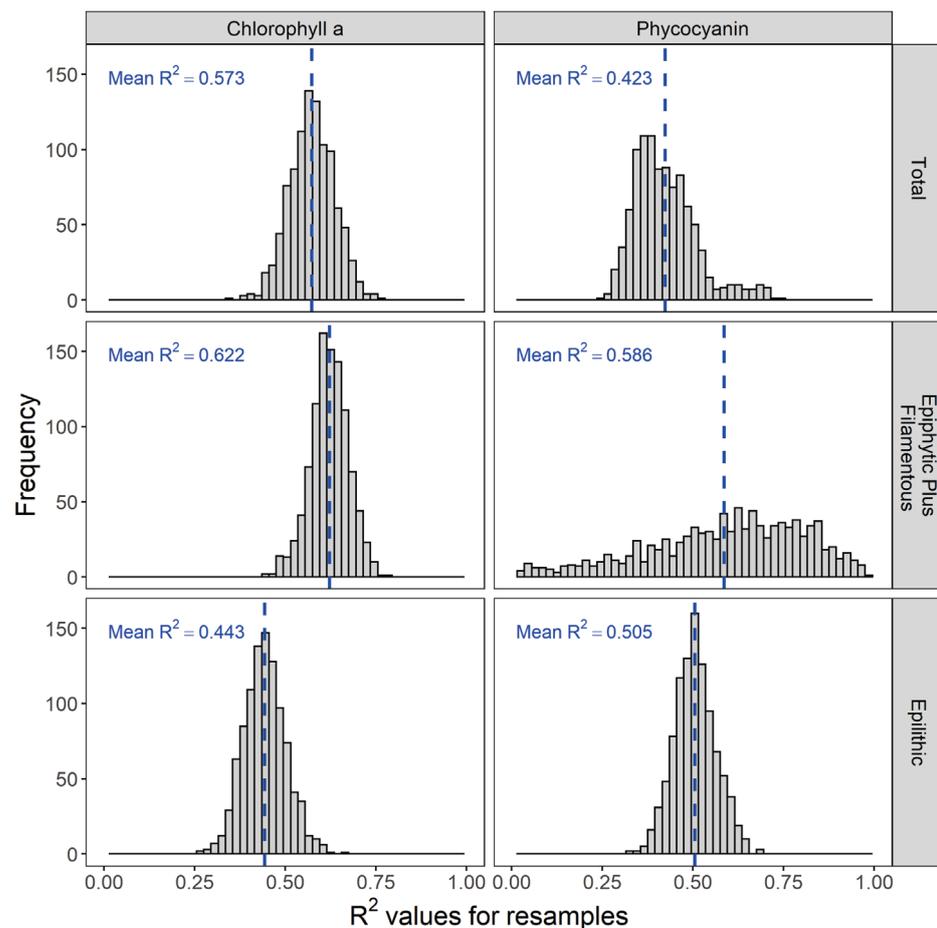
Figure 7. Histograms of R^2 values generated by running linear regressions on each iteration of pigment subsampling against the optimal band ratios identified for estimating pigment abundance for each growth form.

3.4. Comparison of Brute-Force Method with Existing Indices

To provide context for the performance of the presented linear relationships, the derived optimal band ratio for estimating total chl *a* abundance from all field data (n = 33) was compared to several existing band ratio methods for monitoring chl *a*, including the Ha Index [71], Shafique Index [72], and NDVI [73]. All three methods employ the use of band ratios for estimating chl *a* concentration in systems such as tropical lakes using Sentinel 2A imagery [71], large rivers using airborne hyperspectral imagery [72], and general vegetated systems [73]. The selected band ratios were generated using UCFR hyperspectral data, matching the spectral resolution reported in the original work, where NDVI was calculated by integrating across the reported spectral width of Landsat 8 bands [74].

When compared to the optimal band ratio identified by the brute-force method, the selected spectral band ratios performed poorly when tasked with estimating total chl *a* abundance in the UCFR, with a maximum R^2 value of 0.13 (Table 5). Surprisingly, although many of the band ratios used similar wavelengths to those selected by the brute-force method, their performance was notably lower, with NDVI having almost no correlation with total chl *a* abundance. The poor performance associated with NDVI is likely attributable to the near-infrared band centered at 865 nm, a spectral region that is strongly absorbed by water. Previous research has shown that the efficacy of NDVI is a function of water depth, with NDVI becoming ineffective for detecting submerged vegetation at depth greater than 30 cm [75].

To explore the effects of spectral resolution on pigment abundance estimations, the spectral channels used to generate NDVI were reduced to 10 nm and 2.1 nm then fit

to pigment abundance. When reduced to 10 nm, the R^2 modestly improved to 0.05 (*p*-value = 0.23) while a spectral resolution of 2.1 nm generated similar results (R^2 = 0.05, *p*-value = 0.20), suggesting spectral bandwidth did not play a major role in performance in this case. Finally, the brute-force method was set to search for the optimal normalized difference index across the spectral region of 650–850 nm, selecting the same wavelengths used in the simple ratio identified by the brute-force method ($\frac{684 - 674}{684 + 674}$), with nearly identical performance (R^2 = 0.57, *p*-value < 0.001).

Table 5. Comparison of selected spectral band ratios for estimating total chl *a* abundance in the UCFR from all field sites (n = 33). Listed wavelengths represent band centers for each spectral ratio while regression equations are presented as linear equations in the form of y = mx + b.

Band Ratio	Center Wavelengths (nm)	Spectral Resolution (nm)	Regression Equation	R^2	*p*-Value
Ha Index	$\frac{559.8}{664.6}$	30, 35	y = 177.4X − 61.3	0.12	0.048
Shafique Index	$\frac{705}{675}$	5	y = 154.5X − 47.4	0.13	0.041
NDVI	$\frac{865 - 655}{865 + 665}$	30	y = -235.5X + 125.8	0.04	0.25
Brute-Force Method	$\frac{684}{674}$	2.1	y = 2583X − 2448	0.57	<0.001
Brute-Force Method, Normalized Difference	$\frac{684 - 674}{684 + 674}$	2.1	y = 5276X + 135.6	0.57	<0.001

4. Discussion

The optimal band ratio for estimating total and fila/epip chl *a* standing crops in the combined and Gold Creek data sets is likely explained by a change in the reflectance spectra near a chl *a* absorption region driven by variations in pigment abundance. The denominator selected by the brute-force method is near the spectral trough of this region whereas the numerator is along the red edge, a wavelength pair which is likely sensitive to changes in chl *a* abundance. The shift to shorter wavelengths at Bear Gulch is likely driven by a larger proportion of total BOM derived from epiphytic and epilithic growth, representing later stages of bloom progression. The results presented here suggest that late-stage algae are characterized by a decrease in NIR reflectance, shortening of the red edge, and increased reflectance across visible wavelengths, all of which likely reduce the information contained in the vicinity of the 665 nm chl *a* absorption line. During this progression, spectral information near the visible reflectance peak between approximately 530–600 nm carries more information, leading the brute-force method to this region.

The optimal band ratios selected by the brute-force method show promising generalizability across pigment growth forms assessed in the resampling analysis. The results of this analysis showed elevated maximum mean R^2 for the optimal band ratio identified for each growth form, as well as low standard deviation. Though this analysis is still data-limited, these early results show that the optimal band ratios may be a promising tool for assessing pigment abundance in clear and shallow rivers similar to the UCFR.

When compared to existing spectral indices developed for estimating chl *a* concentration, the band ratio selected by the brute-force method showed superior performance to all others, with NDVI performing quite poorly. The poor performance of the existing band ratios may be attributed to their general development for measuring volumetric chl *a* concentration, broader spectral resolution, which tends to smear small spectral features [71,72], or, in the case of NDVI, the use of longer wavelengths that are strongly absorbed by water [73,75]. Overall, the derived band ratios suggest that the spatially averaged spectra (average of all pixels contained in each plot) contain unique and useful information for estimating the abundance of each growth form present.

5. Conclusions

A new method using a UAV-based hyperspectral imager for quantitatively characterizing algal blooms in narrow, shallow rivers has been presented. The results from this analysis suggest that spectral band ratios are a promising method for estimating chl *a* and phycocyanin standing crops contained within, near, and on the surface of blooms of the filamentous nuisance algae, *Cladophora glomerata*, growing in the UCFR. The brute-force analysis used to generate spectral band ratios for estimating chl *a* and phycocyanin standing crops from filamentous, epiphytic, and epilithic growth forms often selected wavelengths near known absorption bands of the pigments [69,70]. The absorption bands represent the spectral nature of primary production carried out both by eukaryotic algae and cyanobacteria and reflects their structural and functional character derived from evolutionary selection for light harvesting and carbon fixation. These features are robust and consistent over the time frames applicable to ecological investigations addressing algal abundance; however, the data presented here are limited to the species and conditions present along the UCFR in 2021. Additionally, the scope of this work is limited to the exploration of linear relationships between spectral band ratios and pigment abundance. In cases where the linear fit is not highly correlated (i.e., low R^2 values), a nonlinear fit may explain more of the observed variance in pigment abundance, but also raises the risk of over-fitting. Questions still remain about the generalizability of the selected optimal band ratios, meaning that more data must be collected and analyzed, though preliminary analyses show strong agreement across data sets.

The brute-force method converged on two optimal band ratios when estimating fila/epip chl *a* standing crops across data sets, selecting a ratio of 684/674 nm for both the combined and Gold Creek data sets and a ratio of 554/536 nm at the Bear Gulch site. The brute-force analysis appeared to identify other important wavelengths for estimating chl *a* for algae in different phenological states, selecting shorter wavelengths (554, 536 nm) when analyzing *Cladophora* blooms with significant diatom coverage, such as those seen at the Bear Gulch site. The method also illuminates regions sensitive to phycocyanin standing crops, often selecting wavelengths over 700 nm (Table 3). Interestingly, the brute-force analysis selected similar band ratios for epilithic chl *a* and phycocyanin abundance, indicating that these signals may be conflated. That is, the presence of one pigment may be related to the presence (or absence) of the other.

The brute-force analysis outlined here represents the early framework for developing a network of low-cost sensors to estimate pigment standing crops over large stretches of riverine systems affected by algal bloom activity. The adoption of a sensor network based on the proposed method would allow for more frequent data collection, while also providing information on a spatial scale large enough to understand ecological factors such as algal metabolism and nutrient uptake. While current measurement methods are constrained to small spatial scales (cm^2–m^2), the proposed method allows for much broader spatial coverage (100–1000 m), allowing for greater understanding of the relationship between RAB development and functional attributes related to water quality. Expanding access to RAB detection and monitoring techniques would help maintain water quality standards, especially in areas where standards are based on algal biomass or nutrient standards, or lead to remediation in contaminated waterways.

In future work, a larger data set collected in the UCFR in 2022 will be analyzed and new data will be collected in the Gallatin River in Montana to begin assessing how well the results presented here apply to other times, conditions, and rivers. These additional data also will aid in the design of multispectral imaging systems capable of RAB detection, along with the development of algal classification maps leading to percent cover estimates.

References

1. Watson, V.; Berlind, P.; Bahls, L. Control of Algal Standing Crop by P and N in the Clark Fork River. In Proceedings of the Clark Fork River Symposium, Missoula, MT, USA, 20–20 April 1990; pp. 47–62.
2. Dodds, W.K.; Smith, V.H.; Zander, B. Developing nutrient targets to control benthic chlorophyll levels in streams: A case study of the Clark Fork River. *Water Res.* **1997**, *31*, 1738–1750. [CrossRef]
3. Suplee, M.W.; Watson, V.; Dodds, W.K.; Shirley, C. Response Of Algal Biomass To Large-Scale Nutrient Controls In The Clark Fork River, Montana, United States. *J. Am. Water Resour. Assoc.* **2012**, *48*, 1008–1021. [CrossRef]
4. Whitton, B.A. Biology of Cladophora in Freshwaters. *Water Res.* **1970**, *4*, 457–476. [CrossRef]
5. Biggs, B.J.; Price, G.M. A survey of filamentous algal proliferations in new zealand rivers. *N. Z. J. Mar. Freshw. Res.* **1987**, *21*, 175–191. [CrossRef]
6. Timoshkin, O.A.; Samsonov, D.P.; Yamamuro, M.; Moore, M.V.; Belykh, O.I.; Malnik, V.V.; Sakirko, M.V.; Shirokaya, A.A.; Bondarenko, N.A.; Domysheva, V.M.; et al. Rapid ecological change in the coastal zone of Lake Baikal (East Siberia): Is the site of the world's greatest freshwater biodiversity in danger? *J. Great Lakes Res.* **2016**, *42*, 487–497. [CrossRef]
7. Whitman, R.L.; Shively, D.A.; Pawlik, H.; Nevers, M.B.; Byappanahalli, M.N. Occurrence of Escherichia coli and enterococci in Cladophora (Chlorophyta) in nearshore water and beach sand of Lake Michigan. *Appl. Environ. Microbiol.* **2003**, *69*, 4714–4719. [CrossRef] [PubMed]
8. Verhougstraete, M.P.; Byappanahalli, M.N.; Rose, J.B.; Whitman, R.L. Cladophora in the Great Lakes: Impacts on beach water quality and human health. *Water Sci. Technol.* **2010**, *62*, 68–76. [CrossRef]
9. Heuvel, A.V.; McDermott, C.; Pillsbury, R.; Sandrin, T.; Kinzelman, J.; Ferguson, J.; Sadowsky, M.; Byappanahalli, M.; Whitman, R.; Kleinheinz, G.T. The Green Alga, Cladophora , Promotes Escherichia coli Growth and Contamination of Recreational Waters in Lake Michigan. *J. Environ. Qual.* **2010**, *39*, 333–344. [CrossRef]

10. Montana Department of Environmantal Quality. 17.30.631 Numeric Algal Biomass And Nutrient Standards. 2002. Available online: https://rules.mt.gov/gateway/ruleno.asp?RN=17%2E30%2E631 (accessed on 31 May 2023).

11. Richardson, L.L. Remote Sensing of Algal Bloom Dynamics. *BioScience* **1996**, *46*, 492–501. [CrossRef]

12. Li, L.; Sengpiel, R.E.; Pascual, D.L.; Tedesco, L.P.; Wilson, J.S.; Soyeux, A. Using hyperspectral remote sensing to estimate chlorophyll-a and phycocyanin in a mesotrophic reservoir. *Int. J. Remote Sens.* **2010**, *31*, 4147–4162. [CrossRef]

13. Steinman, A.D.; Lamberti, G.A.; Leavitt, P.R.; Uzarski, D.G. Biomass and Pigments of Benthic Algae. In *Methods in Stream Ecology*; Elsevier: Amsterdam, The Netherlands, 2017; Volume 1, pp. 223–241.

14. Topp, S.N.; Pavelsky, T.M.; Jensen, D.; Simard, M.; Ross, M.R. Research trends in the use of remote sensing for inland water quality science: Moving towards multidisciplinary applications. *Water* **2020**, *12*, 169. [CrossRef]

15. Gholizadeh, M.H.; Melesse, A.M.; Reddi, L. A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors* **2016**, *16*, 1298. [CrossRef] [PubMed]

16. Wei, G.F.; Tang, D.L.; Wang, S. Distribution of chlorophyll and harmful algal blooms (HABs): A review on space based studies in the coastal environments of Chinese marginal seas. *Adv. Space Res.* **2008**, *41*, 12–19. [CrossRef]

17. Sathyendranath, S.; Cota, G.; Stuart, V.; Maass, H.; Platt, T. Remote sensing of phytoplankton pigments: A comparison of empirical and theoretical approaches. *Int. J. Remote Sens.* **2001**, *22*, 249–273. [CrossRef]

18. Sathyendranath, S.; Prieur, L.; Morel, A. A three-component model of ocean colour and its application to remote sensing of phytoplankton pigments in coastal waters. *Int. J. Remote Sens.* **1989**, *10*, 1373–1394. [CrossRef]

19. Brando, V.E.; Dekker, A.G. Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1378–1387. [CrossRef]

20. Ekstrand, S. Landsat TM based quantification of chlorophyll-a during algae blooms in coastal waters. *Int. J. Remote Sens.* **1992**, *13*, 1913–1926. [CrossRef]

21. Matthews, M.W.; Bernard, S.; Robertson, L. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sens. Environ.* **2012**, *124*, 637–652. [CrossRef]

22. Pahlevan, N.; Smith, B.; Binding, C.; Gurlin, D.; Li, L.; Bresciani, M.; Giardino, C. Hyperspectral retrievals of phytoplankton absorption and chlorophyll-a in inland and nearshore coastal waters. *Remote Sens. Environ.* **2021**, *253*, 112200. [CrossRef]

23. Papenfus, M.; Schaeffer, B.; Pollard, A.I.; Loftin, K. Exploring the potential value of satellite remote sensing to monitor chlorophyll-a for US lakes and reservoirs. *Environ. Monit. Assess.* **2020**, *192*, 808. [CrossRef]

24. Simis, S.G.H.; Peters, S.W.M.; Gons, H.J. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. *Limnol. Oceanogr.* **2005**, *50*, 237–245. [CrossRef]

25. Anderson, K.; Gaston, K.J. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Front. Ecol. Environ.* **2013**, *11*, 138–146. [CrossRef]

26. Cheng, K.H.; Chan, S.N.; Lee, J.H. Remote sensing of coastal algal blooms using unmanned aerial vehicles (UAVs). *Mar. Pollut. Bull.* **2020**, *152*, 110889. [CrossRef]

27. Kwon, Y.S.; Pyo, J.C.; Kwon, Y.H.; Duan, H.; Cho, K.H.; Park, Y. Drone-based hyperspectral remote sensing of cyanobacteria using vertical cumulative pigment concentration in a deep reservoir. *Remote Sens. Environ.* **2020**, *236*. [CrossRef]

28. Wu, D.; Li, R.; Zhang, F.; Liu, J. A review on drone-based harmful algae blooms monitoring. *Environ. Monit. Assess.* **2019**, *191*, 211. [CrossRef]

29. King, T.; Hundt, S.; Hafen, K.; Stengel, V.; Ducar, S. Mapping the probability of freshwater algal blooms with various spectral indices and sources of training data. *J. Appl. Remote Sens.* **2022**, *16*, 044522. [CrossRef]

30. Flynn, K.F.; Chapra, S.C. Remote sensing of submerged aquatic vegetation in a shallow non-turbid river using an unmanned aerial vehicle. *Remote Sens.* **2014**, *6*, 12815–12836. [CrossRef]

31. Kim, E.J.; Nam, S.H.; Koo, J.W.; Hwang, T.M. Hybrid approach of unmanned aerial vehicle and unmanned surface vehicle for assessment of chlorophyll-a imagery using spectral indices in stream, South Korea. *Water* **2021**, *13*, 1930. [CrossRef]

32. Kupssinskü, L.S.; Guimarães, T.T.; De Souza, E.M.; Zanotta, D.C.; Veronez, M.R.; Gonzaga, L.; Mauad, F.F. A method for chlorophyll-a and suspended solids prediction through remote sensing and machine learning. *Sensors* **2020**, *20*, 2125. [CrossRef] [PubMed]

33. Kureel, N.; Sarup, J.; Matin, S.; Goswami, S.; Kureel, K. Modelling vegetation health and stress using hyperspectral remote sensing data. *Model. Earth Syst. Environ.* **2022**, *8*, 733–748. [CrossRef]

34. Main, R.; Cho, M.A.; Mathieu, R.; O'Kennedy, M.M.; Ramoelo, A.; Koch, S. An investigation into robust spectral indices for leaf chlorophyll estimation. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 751–761. [CrossRef]

35. Binding, C.; Greenberg, T.; Bukata, R. The MERIS Maximum Chlorophyll Index; its merits and limitations for inland water algal bloom monitoring. *J. Great Lakes Res.* **2013**, *39*, 100–107. [CrossRef]

36. Huang, Y.; Jiang, D.; Zhuang, D.; Fu, J. Evaluation of hyperspectral indices for chlorophyll-a concentration estimation in Tangxun lake (Wuhan, China). *Int. J. Environ. Res. Public Health* **2010**, *7*, 2437–2451. [CrossRef] [PubMed]

37. Fisher, S.G.; Gray, L.J.; Grimm, N.B.; Busch, D.E. Temporal Succession in a Desert Stream Ecosystem Following Flash Flooding. *Ecol. Monogr.* **1982**, *52*, 93–110.

38. Minshall, G.W.; Petersen, R.C.; Cummins, K.W.; Bott, T.L.; Sedell, J.R.; Cushing, C.E.; Vannote, R.L. Interbiome comparison of stream ecosystem dynamics. *Ecol. Monogr.* **1983**, *53*, 1–25. [CrossRef]

39. McClain, M.E.; Richey, J.E.; Brandes, J.A.; Pimentel, T.P. Dissolved organic matter and terrestrial-lotic linkages in the Central Amazon Basin of Brazil. *Glob. Biogeochem. Cycles* **1997**, *11*, 295–311. [CrossRef]

40. Garcia, R.D.; Reissig, M.; Queimaliños, C.P.; Garcia, P.E.; Dieguez, M.C. Climate-driven terrestrial inputs in ultraoligotrophic mountain streams of Andean Patagonia revealed through chromophoric and fluorescent dissolved organic matter. *Sci. Total Environ.* **2015**, *521*, 280–292. [CrossRef]

41. Vannote, R.L.; Minshall, G.W.; Cummins, K.W.; Sedell, J.R.; Cushing, C.E. The River Continuum Concept. *Can. J. Fish. Aquat. Sci.* **1980**, *37*, 130–137. [CrossRef]

42. Thorp, J.H.; Delong, M.D. The Riverine Productivity Model: An Heuristic View of Carbon Sources and Organic Processing in Large River Ecosystems. *Oikos* **1994**, *70*, 305–308. [CrossRef]

43. Tate, C.M. Patterns and controls of nitrogen in tallgrass prairie streams. *Ecology* **1990**, *71*, 2007–2018. [CrossRef]

44. Valett, H.M.; Peipoch, M.; Poole, G.C. Nutrient processing domains: Spatial and temporal patterns of material retention in running waters. *Freshw. Sci.* **2022**, *41*, 195–214. [CrossRef]

45. Minshall, G.W. Autotrophy in stream ecosystems. *BioScience* **1978**, *28*, 767–771. [CrossRef]

46. Hall Jr, R.O.; Yackulic, C.B.; Kennedy, T.A.; Yard, M.D.; Rosi-Marshall, E.J.; Voichick, N.; Behn, K.E. Turbidity, light, temperature, and hydropeaking control primary productivity in the c olorado river, g rand c anyon. *Limnol. Oceanogr.* **2015**, *60*, 512–526. [CrossRef]

47. Finlay, J.C. Patterns and controls of lotic algal stable carbon isotope ratios. *Limnol. Oceanogr.* **2004**, *49*, 850–861. [CrossRef]

48. Bernhardt, E.; Heffernan, J.; Grimm, N.; Stanley, E.; Harvey, J.; Arroita, M.; Appling, A.P.; Cohen, M.; McDowell, W.; Hall, R., Jr.; et al. The metabolic regimes of flowing waters. *Limnol. Oceanogr.* **2018**, *63*, S99–S118. [CrossRef]

49. Battin, T.J.; Lauerwald, R.; Bernhardt, E.S.; Bertuzzo, E.; Gener, L.G.; Hall Jr, R.O.; Hotchkiss, E.R.; Maavara, T.; Pavelsky, T.M.; Ran, L.; et al. River ecosystem metabolism and carbon biogeochemistry in a changing world. *Nature* **2023**, *613*, 449–459. [CrossRef]

50. Logan, R.D.; Hamp, S.M.; Torrey, M.A.; Feijo de Lima, R.; Colman, B.P.; Valett, H.; Shaw, J.A. UAV-Based Hyperspectral Imaging for River Algae Pigment Estimation and Development of a Low-Cost Multispectral Imager. In Proceedings of the SPIE Future Sensing Technologies Proceedings, Yokohoma, Japan, 22 May 2023.

51. Ingman, G.L.; Kerr, M.A. Nutrient Sources in the Clark Fork River Basin. In Proceedings of the Clark Fork River Symposium, Missoula, MT, USA, 20–20 April 1990.

52. Lohman, K.; Priscu, J.C. Physiological Indicators of Nutrient Deficiency in Cladophora Chlorophyta. *J. Phycol.* **1992**, *28*, 443–448. [CrossRef]

53. Banish, N.J. Factors Influencing Cladophora Biomass Abundance in the Upper Clark Fork River, Montana. Ph.D Thesis, University of Montana, Missoula, MT, USA, 2017.

54. Suplee, M.W.; Watson, V.; Teply, M.; McKee, H. How Green is Too Green? Public Opinion of What Constitutes Undesirable Algae Levels in Streams 1. *JAWRA J. Am. Water Resour. Assoc.* **2009**, *45*, 123–140. [CrossRef]

55. Power, M.; Lowe, R.; Furey, P.; Welter, J.; Limm, M.; Finlay, J.; Bode, C.; Chang, S.; Goodrich, M.; Sculley, J. Algal mats and insect emergence in rivers under Mediterranean climates: Towards photogrammetric surveillance. *Freshw. Biol.* **2009**, *54*, 2101–2115. [CrossRef]

56. Power, M.E.; Bouma-Gregson, K.; Higgins, P.; Carlson, S.M. The thirsty Eel: Summer and winter flow thresholds that tilt the Eel River of northwestern California from salmon-supporting to cyanobacterially degraded states. *Copeia* **2015**, *103*, 200–211. [CrossRef]

57. Furey, P.C.; Lowe, R.L.; Power, M.E.; Campbell-Craven, A.M. Midges, Cladophora, and epiphytes: Shifting interactions through succession. *Freshw. Sci.* **2012**, *31*, 93–107. [CrossRef]

58. DJI. *Matrice 600 Pro—User Manual*; Technical Report; DJI: Shenzhen, China, 2022.

59. Resonon. *Airborne User Manual*; Technical Report; Resonon: Bozeman, MT, USA, 2022.

60. Ritchie, R.J. Universal chlorophyll equations for estimating chlorophylls a, b, c, and d and total chlorophylls in natural assemblages of photosynthetic organisms using acetone, methanol, or ethanol solvents. *Photosynthetica* **2008**, *46*, 115–126. [CrossRef]

61. Moulton, T.P.; Souza, M.L.; Walter, T.L.; Krsulović, F.A. Patterns of periphyton chlorophyll and dry mass in a neotropical stream: A cheap and rapid analysis using a hand-held fluorometer. *Mar. Freshw. Res.* **2009**, *60*, 224–233. [CrossRef]

62. Cremella, B.; Huot, Y.; Bonilla, S. Interpretation of total phytoplankton and cyanobacteria fluorescence from cross-calibrated fluorometers, including sensitivity to turbidity and colored dissolved organic matter. *Limnol. Oceanogr. Methods* **2018**, *16*, 881–894. [CrossRef]

63. Legleiter, C.J.; Roberts, D.A.; Marcus, W.A.; Fonstad, M.A. Passive optical remote sensing of river channel morphology and in-stream habitat: Physical basis and feasibility. *Remote Sens. Environ.* **2004**, *93*, 493–510. [CrossRef]

64. Legleiter, C.J.; Roberts, D.A.; Lawrence, R.L. Spectrally based remote sensing of river bathymetry. *Earth Surf. Process. Landforms* **2009**, *34*, 613–628. [CrossRef]

65. Shaw, J.A. The effect of instrument polarization sensitivity on sea surface remote sensing with infrared spectroradiometers. *J. Atmos. Ocean. Technol.* **2002**, *19*, 820–827. [CrossRef]

66. Logan, R.D.; Venkatesulu, E.; Shaw, J.A. Comparing the Polarimetric Response of Hyperspectral Imagers. In Proceedings of the Polarization Science and Remote Sensing X, San Diego, CA, USA, 1–5 August 2021; Volume 11833, pp. 90–97.

67. Haltrin, V.I. Absorption and Scattering of Light in Natural Waters. In *Light Scattering Reviews: Single and Multiple Light Scattering*; Kokhanovsky, A.A., Ed.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 445–486. [CrossRef]

68. Buiteveld, H.; Hakvoort, J.H.M.; Donze, M. Optical properties of pure water. *Ocean. Opt. XII* **1994**, *2258*, 174–183. [CrossRef]
69. Ylöstalo, P.; Kallio, K.; Seppälä, J. Absorption properties of in-water constituents and their variation among various lake types in the boreal region. *Remote Sens. Environ.* **2014**. *148*, 190–205. [CrossRef]
70. Karcz, D.; Boroń, B.; Matwijczuk, A.; Furso, J.; Staroń, J.; Ratuszna, A.; Fiedor, L. Lessons from chlorophylls: Modifications of porphyrinoids towards optimized solar energy conversion. *Molecules* **2014**, *19*, 15938–15954. [CrossRef] [PubMed]
71. Ha, N.T.T.; Thao, N.T.P.; Koike, K.; Nhuan, M.T. Selecting the best band ratio to estimate chlorophyll-a concentration in a tropical freshwater lake using sentinel 2A images from a case study of Lake Ba Be (Northern Vietnam). *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 290. [CrossRef]
72. Shafique, N.A.; Fulk, F.; Autrey, B.C.; Flotemersch, J. Hyperspectral Remote Sensing of Water Quality Parameters for Large Rivers in the Ohio River Basin. In Proceedings of the 1st Interagency Conference on Research in the Watershed, Benson, AZ, USA, 27–30 October 2003; pp. 216–221.
73. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]
74. Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* **2016**, *185*, 46–56. [CrossRef]
75. Cho, H.J.; Kirui, P.; Natarajan, H. Test of Multi-spectral Vegetation Index for Floating and Canopy-forming Submerged Vegetation. *Int. J. Environ. Res. Public Health* **2008**, *5*, 477–483. [CrossRef] [PubMed]

Article

Implementation of Ground-Based Lightning Locating System Using Particle Swarm Optimization Algorithm for Lightning Mapping and Monitoring

Kamyar Mehranzamir [1], Amin Beiranvand Pour [2,3,*], Zulkurnain Abdul-Malek [4], Hadi Nabipour Afrouzi [5], Seyed Morteza Alizadeh [6] and Mazlan Hashim [3]

[1] Department of Electrical and Electronic Engineering, Faculty of Science and Engineering, University of Nottingham Malaysia, Jalan Broga, Semenyih 43500, Selangor, Malaysia; kamyar.mehranzamir@nottingham.edu.my

[2] Institute of Oceanography and Environment (INOS), Universiti Malaysia Terengganu (UMT), Kuala Nerus 21030, Terengganu, Malaysia

[3] Geoscience and Digital Earth Centre (INSTeG), Research Institute for Sustainable Environment, Universiti Teknologi Malaysia, Johor Bahru 81310, Johor, Malaysia

[4] Institute of High Voltage & High Current, School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Johor, Malaysia

[5] Faculty of Engineering, Computing and Science, Swinburne University of Technology Sarawak, Kuching 93350, Sarawak, Malaysia

[6] Engineering Institute of Technology, Melbourne, VIC 3000, Australia

* Correspondence: beiranvand.pour@umt.edu.my; Tel.: +60-9-6683824; Fax: +60-9-6692166

Abstract: Cloud-to-ground (CG) lightning is a natural phenomenon that poses significant threats to human safety, infrastructure, and equipment. The destructive impacts of lightning strikes on humans and their property have been a longstanding concern for both society and industry. Countries with high thunderstorm frequencies, such as Malaysia, experience significant fatalities and damage due to lightning strikes. To this end, a lightning locating system (LLS) was developed and deployed in a 400 km^2 study area at the University Technology Malaysia (UTM), Johor, Malaysia for detecting cloud-to-ground lightning discharges. The study utilized a particle swarm optimization (PSO) algorithm as a mediator to identify the best location for a lightning strike. The algorithm was initiated with 30 particles, considering the outcomes of the MDF and TDOA techniques. The effectiveness of the PSO algorithm was found to be dependent on how the search process was arranged. The results of the detected lightning strikes by the PSO-based LLS were compared with an industrial lightning detection system installed in Malaysia. From the experimental data, the mean distance differences between the PSO-based LLS and the industrial LLS inside the study area was up to 573 m. Therefore, the proposed PSO-based LLS would be efficient and accurate to detect and map the lightning discharges occurring within the coverage area. This study is significant for researchers, insurance companies, and the public seeking to be informed about the impacts of lightning discharges.

Keywords: lightning locating system; particle swarm optimization; VLF and VHF sensors; GPS antennas; lightning mapping; environmental monitoring

1. Introduction

Precise determination of a cloud-to-ground (CG) lightning strike location is essential for various applications utilizing lightning locating system (LLS) data. Moreover, a reliable LLS system can function as an efficient means to monitor lightning discharges in real-time. The capacity to promptly detect flashes can help facilitate early warning systems for severe weather events and enable timely and effective disaster response. Lightning is a naturally occurring phenomenon that often takes place during thunderclouds and rainfall. The electromagnetic radiation emitted by a lightning discharge spans a broad spectrum, from

very low frequency (VLF) to very high frequency (VHF). Furthermore, lightning discharges can deliver a substantial amount of current, typically in the tens of kiloampere range, to objects or the ground [1,2]. Hence, this charge transfer is a serious worry for humans and the industry. Cloud to cloud, cloud to air, inter-cloud, and cloud to ground (CG) are different types of lightning discharges. The CG discharges are comprised of fast/very fast impulse current discharges to the ground with return stokes [3–5].

Governments establish meteorological stations to track the frequency of lightning events annually. The majority of lightning strike-related damage and disruption stems from the rapid discharge of electrical energy in the form of flashes between clouds and the ground. In addition to the damage to properties, many people are killed by the cloud-to-ground discharges in the world each year [6–8]. These severe negative impacts of CG discharges to people and devices led to the designing of protective systems such as lightning protection systems [1]. Power system outages, forest fires, and livestock deaths are a few examples of lightning-related incidents. Nowadays, the behavior of lightning activities is monitored by various sensors and systems, which are satellite-based or ground-based networks [9–11]. A lightning locating system (LLS) is designed to accurately locate lightning strikes and provide valuable information for both private and governmental sectors. Moreover, LLSs assist researchers in extracting essential characteristics and parameters of lightning strikes. Precise estimation of CG strike location is critical for applications that utilize LLS information. It should be emphasized that various characteristics of lightning discharges can be extracted from the captured data depending on the type of LLS employed [12]. Lightning locating systems (LLSs) accumulate extensive data on lightning activities, which can be applied in various practical contexts, including weather monitoring and prediction, geophysical research, and national power grid protection.

Single station lightning detection sensors, such as lightning discharge counters, are only capable of detecting and tallying the instances of lightning discharges, without accurately locating the source of the radiation. However, the advent of multi-sensor lightning detection systems, which have been in use since the 1920s, has reduced the issues associated with single station LLSs. Russian physicist Popoff conducted the first measurements in 1895 to study the electromagnetic fields created by lightning at a distance. Other scientists such as Appleton, Watson-Watt, and Herd also conducted research on radiation fields [13]. In subsequent years, more sensors were utilized to capture the entirety of the lightning channel and provide a more precise LLS. Rakov noted in 2013 that two commonly used lightning locating techniques for long-baseline systems operating in the LF/VLF ranges are the time difference of arrival (TDOA) and magnetic direction finding (MDF) methods [14]. Most ground-based LLSs require more than one station (sensor) to accurately locate the lightning discharges [15]. The accurate lightning locating system based on different technologies inevitably includes several sensors.

1.1. Lightning Detection Techniques

Lightning tracking and mapping system has many applications in both academic and industrial sectors. The detected weather changes by the LLSs are used for understanding the physics of lightning discharges and global electric movements. Magnetosphere, ionosphere, and NOx generation studies also use the captured data by the LLSs [16–19].

Lightning mapping systems are closely monitored by weather forecasters and researchers to study storm patterns. The aviation industry is a major user of LLSs, with air traffic controllers redirecting flights as soon as weather hazards are detected. LLS alarms are also equipped in fire detectors in forestry areas, outdoor activities, and sports facilities to warn the public [20,21]. Many outdoor activities such as recreation centers and outdoor facilities such as transmission lines can be influenced by lightning [22–25].

Various lightning detection sensors and systems possess their own advantages and drawbacks. For instance, some are capable of detecting cloud-to-ground lightning (CG) and other cloud activities, while others have limitations on their coverage areas and require an optimal number and location of measuring stations. Nevertheless, the advancements

in lightning detection techniques have helped researchers locate lightning discharges with minimal detection errors. In the past three decades, lightning locating systems have undergone significant improvements, resulting in more accurate systems [26]. Modern ground-based LLSs are capable of not only locating lightning, but also determining its peak current and tracking cloud movements. Different working frequencies of LLSs reveal different flash characteristics, which is why they are installed in various frequency ranges to provide complementary information about thunderstorms. To establish a correlation between annual lightning discharges and landform distribution, such as topography and latitude, extensive research is necessary. It has been observed that in certain regions, the number of return strokes and lightning current amplitude increase with altitude.

Two widely used lightning detection methods are magnetic direction finding (MDF) and time difference of arrival (TDOA) methods, which have specific baselines based on the distances and locations of the stations relative to each other. It is known that VLF sensors have a longer wavelength compared to VHF sensors, resulting in VLF radiations being able to travel farther distances than VHF radiations. VHF emissions are commonly emitted from clouds due to electrical breakdown, whereas VLF emissions are propagated by return strokes or k-changes and localize the CG activities [27]. The magnetic crossed-loop antenna is used to obtain the direction of the horizontal component of the captured magnetic field that is generated by a lightning discharge. Based on the construction of this antenna, two vertical and orthogonal loops while the planes are oriented NS (North-South) and EW (East-West) are designed to measure the magnetic fields and, therefore, find the direction of the discharge [28,29].

Loop antennas are widely used in various applications, such as survival communications, oil exploration, and geophysical research, as low-frequency sensors. These antennas can detect sensing magnetic fields, which induce a voltage in the loops of the crossed-loop antennas. Due to their sensitivity to magnetic fields and not electric fields, these antennas are particularly useful for measuring magnetic field variations. The induced voltages in the two loops are proportional to the derivative of the induced magnetic field components perpendicular to the areas of the loops, multiplied by the area of the antenna (A).

The induced voltages in the loop are measured to obtain the magnetic field components. The magnetic field in parallel to the ground plane is generated by the vertical lightning discharge channel [30].

According to the method based on Faraday's law, the direction of the magnetic field source can be determined by measuring the ratio of the induced voltages in two loops. The line connecting the point of the lightning channel is at a right angle to this magnetic field. The output voltage of each loop is proportional to the cosine of the angle between the magnetic field vector and normal vector to the plane of the loop [31,32]. The first few microseconds of the induced magnetic fields by the return strokes are employed to estimate the location for three reasons. First, the few microseconds of the return strokes are generated by the last few hundred meters of the lightning channels. Second, it is known that one of the strongest events in lightning discharge is the return stroke, and it can be detected at large distances. Moreover, it is the return stroke that is of interest in connection with lightning protection. The third reason for using the first few microseconds of a return stroke is to avoid the errors caused by ionosphere reflections [33].

According to López et al. [34] in 1991, Ortéga [35] in 2007, and Chen et al. [36] in 2013, errors are included in calculating the azimuth of arrival waveform to the crossed-loop antenna. Position and topography of the surrounding terrain, such as close metallic objects and tall structures are several possible site errors [37]. The time difference of arrival (TDOA) is a commonly used method to detect lightning strikes. This method involves installing synchronized measuring stations to capture lightning flashes in different frequency bands, including VLF, LF, HF, and VHF ranges. However, as these systems capture various types of cloud activities, a separate algorithm is needed to differentiate cloud-to-ground discharges from other types of cloud activities. By using the TDOA-based LLS in the VLF range, which is typically around 10 kHz, cloud-to-ground discharges can be detected exclusively [38–40].

The precise global positioning system (GPS) records the time of the detected lightning strikes when receiving the signals. The time differences of the detected waveforms are used to create hyperbolas around the measurement stations, and the points of intersection of these hyperbolas determine the location of the lightning. Mathematical calculations require a minimum of four measurement stations to obtain an unambiguous point for each calculation. Another lightning detection technique is interferometry, which uses phase measurement in a small bandwidth to determine the direction of lightning discharge [41]. At least two detecting antennas with distance are needed in such a system. Identical narrowband filters are connected to the receivers. The outputs of the receivers are processed by the phase detector, which generates a voltage. This voltage level is proportional to the phase differences of the signals that define a plane to locate the radiation source. A minimum of three measuring antennas and two orthogonal baselines is required to calculate the azimuth and elevation of the source. Recent lightning locating systems combine more than one detection technique to provide surplus information that localizes the CGs with lower distance error [41].

As MDF computes the triangulation area of the radiation direction and TDOA determines the intersection of hyperbolas, a combination of MDF and TDOA is one of the best choices for a combined LLS. Each technique can provide independent results and, finally, it is possible to find more accurate output by employing optimization techniques such as particle swarm optimization (PSO) [42]. The localization of lightning protection systems usually relies on analyzing the CG lightning flashes captured by the lightning locating system. However, the accuracy of the MDF method for determining the direction of field incidence can be affected by factors such as site errors and local sensor conditions, including nearby objects, metal fences, or other conducting materials [34,36,43].

1.2. Main Lightning Detection Networks in Malaysia

In Malaysia, lightning is a significant cause of power outages. To address this issue, automatic lightning detection networks have been established, which enable the analysis of lightning strikes at both regional and local levels. The PSO-based LLS in this study utilizes a designed map that is automatically accessed by the LLS program from online maps as needed. One advantage of the designed LLS in this research is the use of a dynamic and online map, unlike many other lightning locating systems like BOLTEK that rely on a static photo as the screen map. To obtain a suitable map, companies must be contacted and given the exact installation location of the antenna [44]. Static map photos are unreliable for locating lightning strikes, as the points on them may not be accurate. A dynamic map, which allows zooming in and out to real places, is preferred. Lightning detection networks are typically owned by weather forecasting companies or well-equipped private research institutes due to their size and cost. Public individuals seldom own such networks. It is crucial to warn the public of lightning strikes detected by LDNs, particularly those engaged in outdoor activities. The aim of this work is to address the aforementioned gaps. The Lightning Detection System implemented by the Malaysian Meteorological Department (MMD) comprises two sub-systems, namely the wide area lightning localization sub-system (SAFIR) and the precision lightning methodology warning sub-system (PLWS). The detection stations of these sub-systems report lightning data, detection station status, and auto-test data at intervals of every second and every 5 min, respectively [39].

This research utilizes data from the lightning detection system (LDS) of the Malaysian Meteorological Department (MMD) to determine lightning occurrences in Peninsular Malaysia. The LDS comprises eight sensors located at Bt. Tangga (Kedah), Bertam (Penang), Parit (Perak), Kuala Krai (Kelantan), Jengka (Pahang), Klang (Selangor), Jasin (Melaka), and Mersing (Johor). However, certain sites that are noisy and have obstructions, as well as uneven sensor spacing, are identified as weaknesses of the lightning detection system.

The accuracy of a lightning detection system is influenced by various factors, including the types and number of sensors used, their placement in relation to each other, the topography of the sites, and the presence of obstructions and structures in the environment. At least

three sensors are required to compute lightning data, and they should be placed no farther apart than their nominal range (200 km for VHF sensors). Placing the sensors in a straight line is not recommended. In 1994, TNB Research (TNB-R) installed a lightning detection network (LDN) in Peninsular Malaysia to mitigate the negative impacts of lightning on TNB equipment. TNB-R has installed IMPACT ESP sensors at eight different locations in Malaysia for lightning detection and fault location estimation. Continuous monitoring of sensor performance is carried out by the lightning detection system laboratory operated by TNB-R. Lightning events and their locations are recorded by the Lightning Processor 2000 (LP 2000) system, developed by Vaisala. However, these lightning locating systems do not provide public access to the captured lightning data. The objective of this research is to tackle the issues associated with lightning locating systems utilizing TDOA and MDF techniques. To achieve this, a hybrid TDOA and MDF method employing a particle swarm optimization algorithm is proposed as a PSO-based LLS, which aims to reduce errors associated with both TDOA and MDF techniques.

2. The Study Area and Dataset

Malaysia is situated close to the Equator and is divided into two main regions, namely Peninsular Malaysia or West Malaysia, and East Malaysia or Malaysia Timur, which are not connected to each other. The largest states in East Malaysia are Sarawak and Sabah [45]. Malaysia's eastern region is geographically separated from Peninsular Malaysia by the South China Sea. Both regions lie within the tropical latitudes, characterized by consistently high temperatures and humidity levels throughout the year. In addition to experiencing heavy rainfall, Malaysia also has distinct monsoon seasons in the western part of the country from May to October. The annual temperature ranges from 25 °C to 35 °C [46]. The monsoon regime results in frequent rainfall throughout the year across the entire country. Peninsular Malaysia has a mean annual rainfall of around 2540 mm, but each region has its own unique characteristics. Figure 1 displays the average yearly distribution of lightning discharges per square kilometer using NASA's lightning imaging sensor [47]. Malaysia is among the countries with the highest number of lightning strikes in the world.
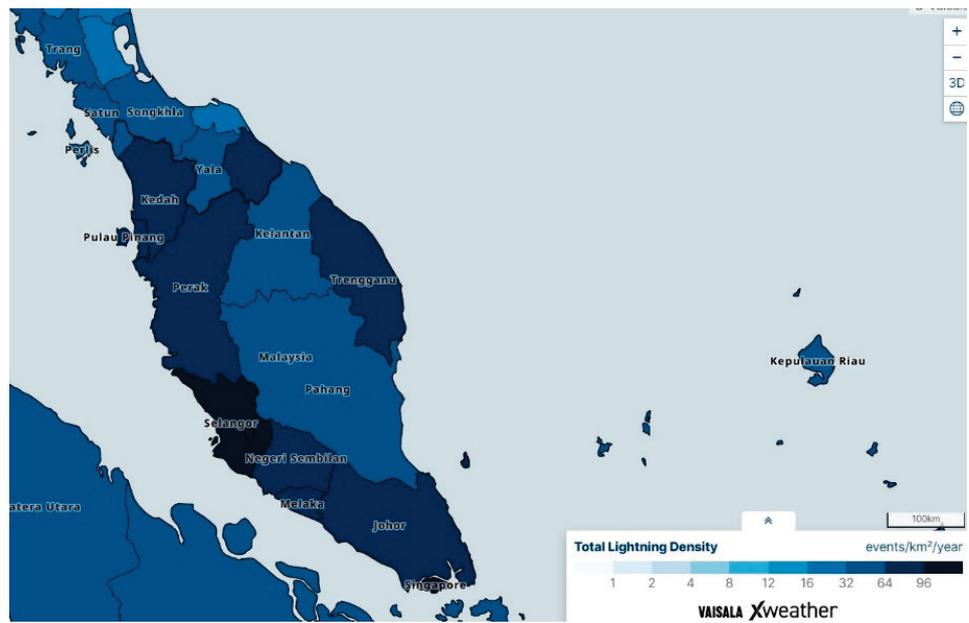

(a)
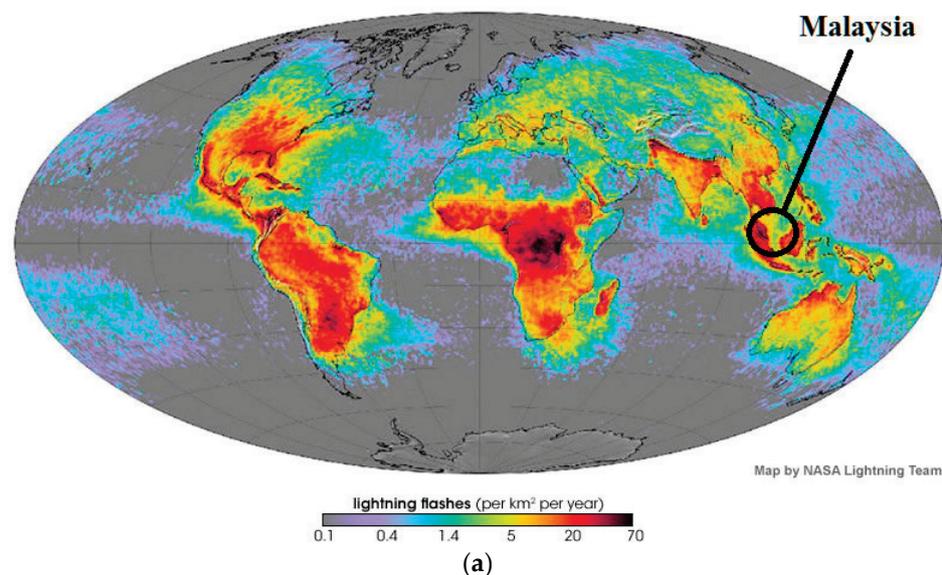
Figure 1. *Cont.*

(**b**)

Figure 1. (**a**) World lightning map; (**b**) Total lightning density of Peninsular Malaysia in (**a**) from Vaisala [47].

Figure 2 shows Peninsular Malaysia and the geographical location of the case study located in Skudai, Johor province, the southern part of Peninsular Malaysia [48,49]. The objective of this research is to analyze the waveforms obtained from measuring stations to determine the occurrence of CG discharges using an implemented LLS. The state of Johor, where the study was conducted, has a climate similar to other states in Malaysia, characterized by frequent rainfall throughout the year. The wettest months in this area are April, October, November, and December. The measurement stations include IVAT station (1.560447, 103.643542), B11 station (1.557839, 103.635694), and VAN station (1.565997, 103.633469).

Figure 2. Peninsular Malaysia and geographical location of the case study located in Skudai, Johor province [50].

Lightning locating systems have been used by many countries to collect data on lightning for the purpose of mapping CG flashes. The coverage area of an LLS depends on the number of stations and frequency ranges used. For this research, a 400 km^2 area centered on the IVAT station (1.560447, 103.643542) was chosen as the study area, as shown in Figure 3. The latitude and longitude coordinates of the vertices of the study area are listed below:
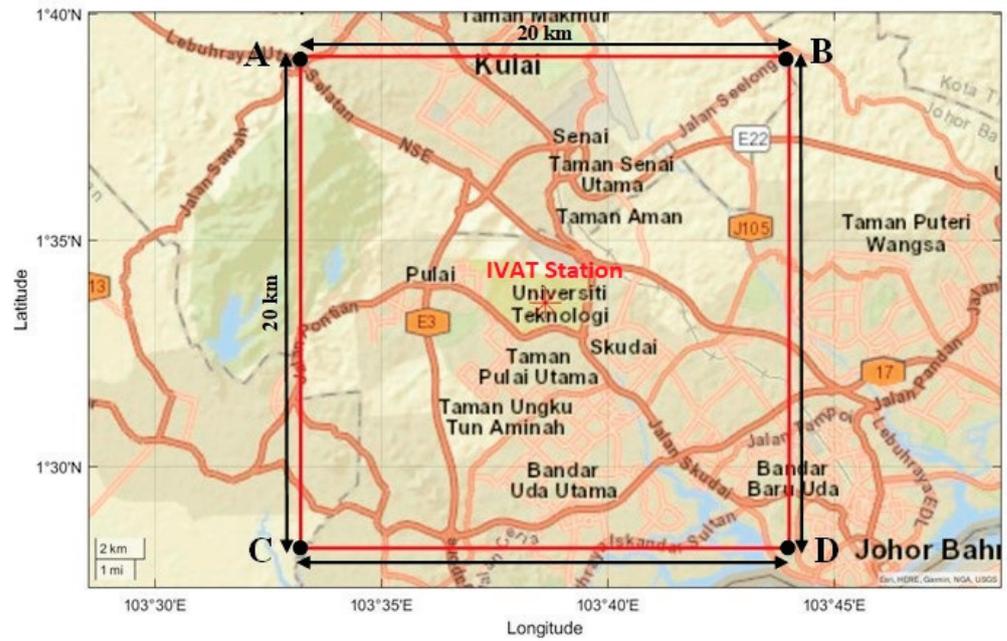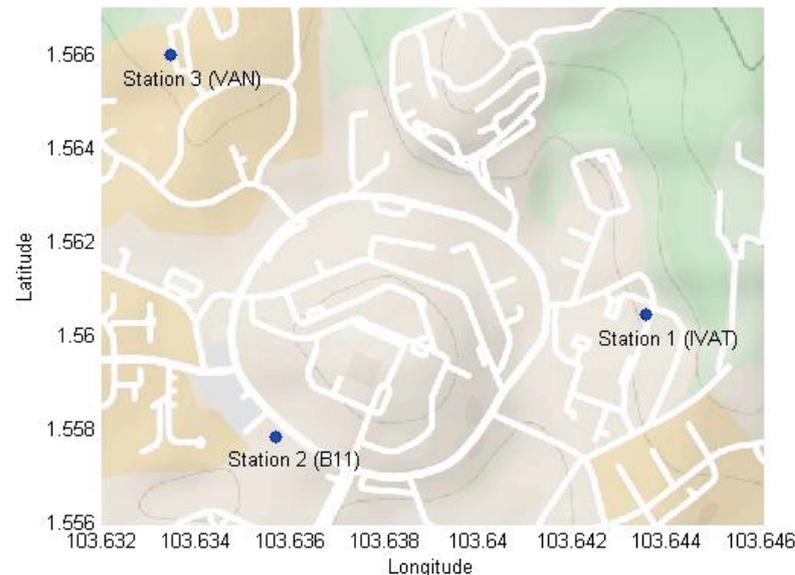
Figure 3. Coverage area of designed lightning locating system with respect to IVAT station (1.560447, 103.643542).

A = 1.650912, 103.553678,
B = 1.650912, 103.733406,
C = 1.469982, 103.553678,
D = 1.469982, 103.733406.

Kulai Toll Plaza in the Kulai region is where Position A is located. Position B is situated in the Tebrau area, which is on the east side of Senai International Airport and 10 km away from it. Jeram Batu area is where Point C is located, adjacent to the Sungai Pulai river and

near Kampung Jelutong Tengah. Johor City is where Location D is situated. IVAT station at the University Technology Malaysia (UTM) is considered the central location.

Calibration of Lightning Locating Sensors

To ensure that the lightning detection equipment was uniform, all three lightning locating sensor systems were placed at the IVAT station for calibration purposes. The goal of this calibration test was to confirm that all the hardware equipment used for lightning detection was identical. To achieve this, three sets of crossed-loop antennas with their amplifiers and three sets of parallel plate antennas with buffer circuits were installed at the IVAT station. The timestamps were saved and synchronized using GPS cards (model: GPS180PEX) to ensure accurate time and location data for the TDOA method. The GPS cards were chosen for their ability to provide more precise time and location information.

The system used GPS cards with a resolution of 100 ns, which provided high-precision time synchronization. The three crossed-loop antennas had the same size, shape, and materials, and the VLF amplifiers were designed with identical components to ensure consistency. It is important to mention that in a lightning locating system, the length of cables and wires from the antennas to data loggers should be the same at all stations. Figure 4 illustrates the diagram of the calibration system installed for this study. The primary objective of the calibration process was to ensure that all the VLF and VHF sensors could detect the CG lightning activities at the same time with minimal delay. The installed system at the IVAT station was tested using incoming lightning strikes.
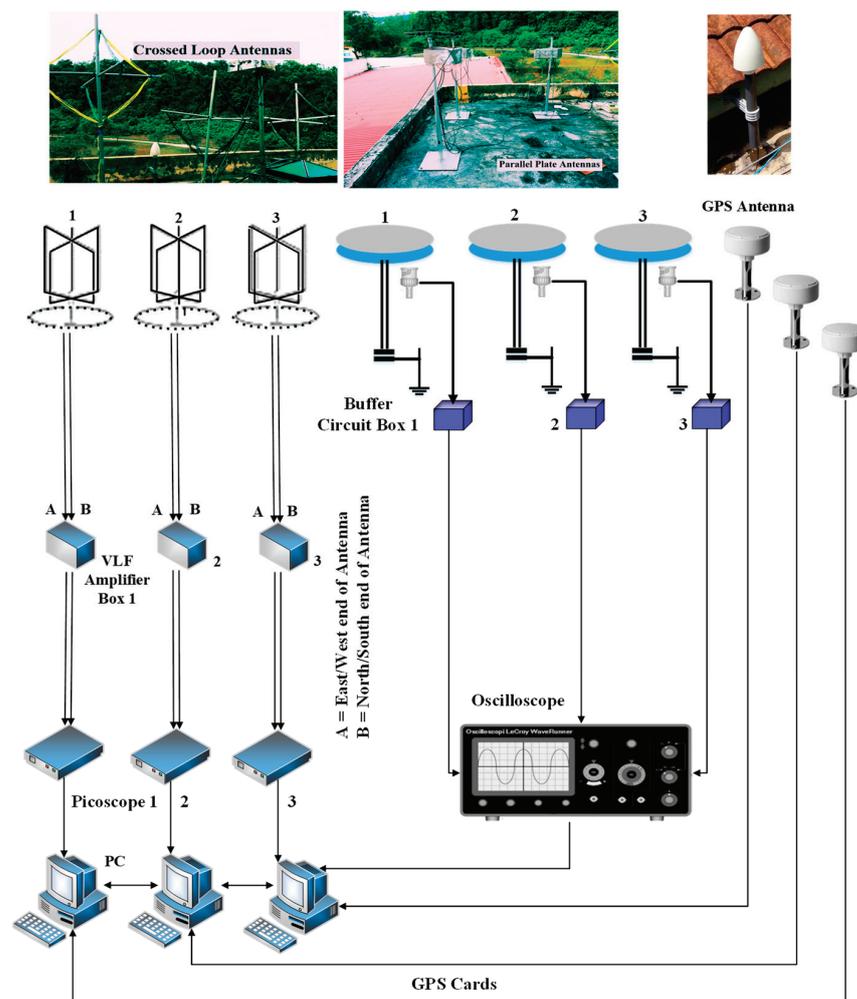
Figure 4. Schematic of calibration process of the sensors at IVAT station installed at Universiti Teknologi Malaysia, Johor (VHF and VLF bands).

In this study, the VLF system captured six waveforms using three crossed-loop antennas, while the VHF system captured electric fields using three parallel plate antennas. These waveforms were then analyzed in MATLAB software to determine the signal delays.

Figure 5 presents three electric field records of a cloud-to-ground lightning flash with 500 ms of time frame detected by antennas 1, 2, and 3 at 15:47:11.2534810 (hour:minute:second 100 ns of time precision). All three GPS cards saved the same timestamps, which means the possible delay time is below 100 ns (GPS resolution). The delay time of 1 ns indicates that the signal has traveled 0.3 m in distance. Hence, if the delay time is 100 ns, it means there is a distance error of 30 m. Therefore, MATLAB software was utilized to analyze the signals and determine the precise time delay.

(**a**)

(**b**)

(**c**)
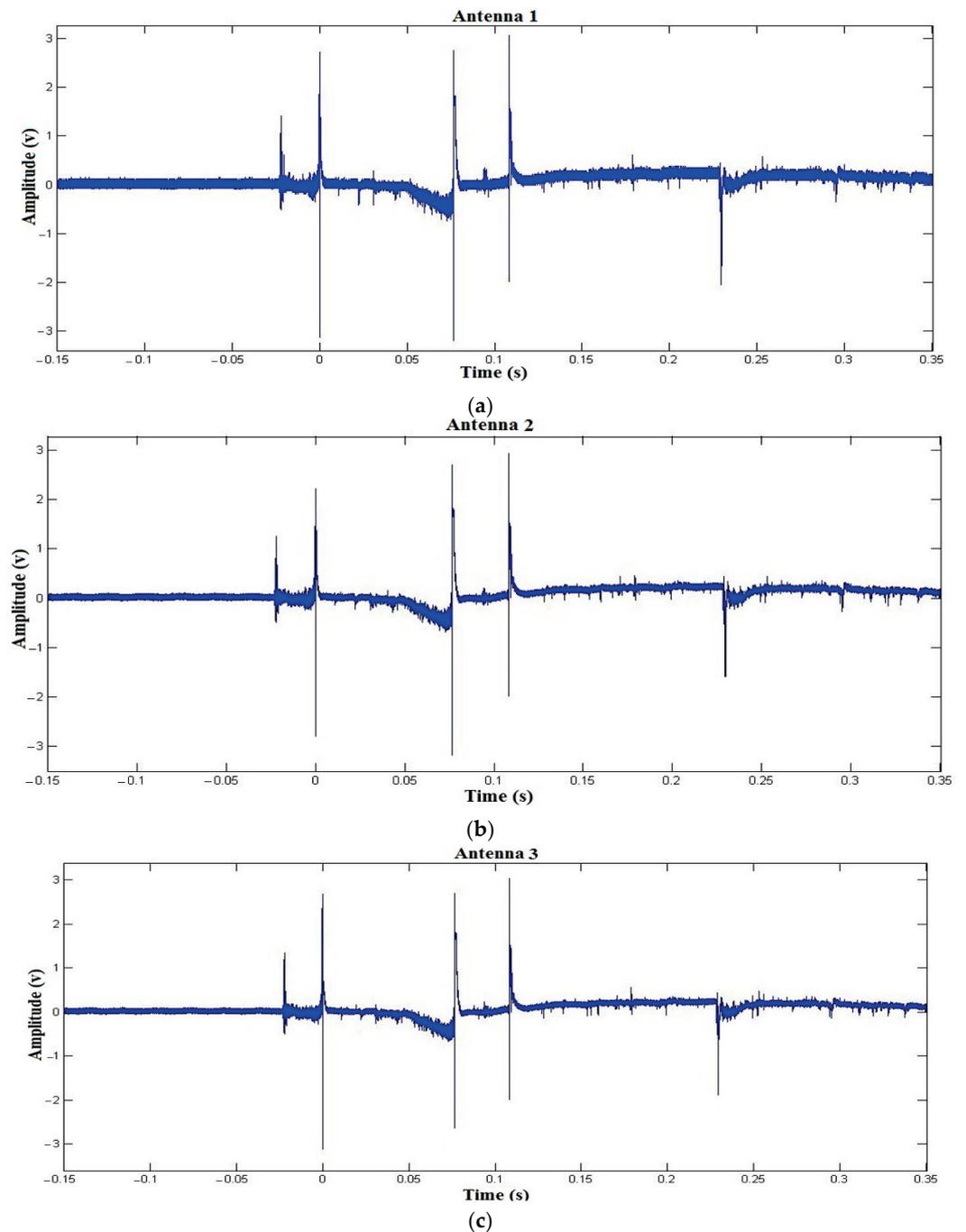
Figure 5. Electric field record of a cloud-to-ground lightning flash with 500 ms of time frame detected by antennas 1 (**a**), 2 (**b**), and 3 (**c**), respectively.

The zoom in of the first return stroke (time zero) of three captured electric fields in Figure 5 is presented in Figure 6 to check the possible delay between the waveforms. In Figure 5, the delay between the captured waveforms by antennas 1–2, 1–3, and 2–3 were 50 ns, 55 ns, and 50 ns, respectively. Corresponding magnetic fields to all electric fields in Figure 5 at a similar time were captured by crossed-loop antennas. The delay time between each pair of six magnetic field signals was calculated and analyzed using MATLAB software. Mathematical calculations were used to find the exact delay time between the signals, as the waveforms had a remarkable resemblance to each other.
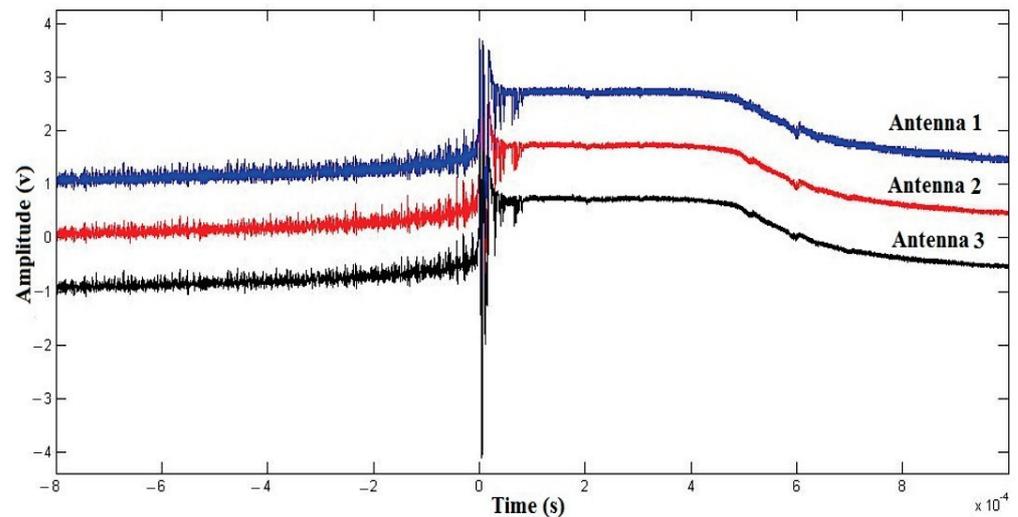
Figure 6. Zoom in of first return stroke of three detected electric fields by antennas 1, 2, and 3 in Figure 5 (shifted in Y axis to show three waveforms).

The calibration test has been repeated for 20 CG lightning events and an average delay time of 50 ns was reported. Therefore, it can be said that the additional delay time was due to different system characteristics of the stations, electronic component's characteristics, and other inconsistencies such as differences in the cable length between the employed equipment at the stations. The delay time observed was very low, and it was less than the resolution of the GPS cards used, which was 100 ns. Hence, this delay time was ignored in this research.

3. Methodology

3.1. Combined Techniques

The time of occurrence of the first return stroke is extracted by high-resolution GPS antennas. The GPS cards are responsible for capturing accurate time and these timestamps are used in the TDOA method. The photos of the GPS card and its antenna are presented in Figure 7. This PCI Express slot card was the best choice for adding a highly accurate time base to the server and workstations. GPS180PEX is a low-profile board for computers with a PCI express interface. The rear slot cover integrates the antenna connector, a BNC connector for modulated time codes, a 9-pin D_SUB male connector, and two status LEDs. The resolution of pulse outputs is 100 ns.

The MDF method can provide the location of the discharge with a minimum of two sensors but the location error is expected to be dependent on the distance growth. Hence, it is necessary to decrease the site error by adding extra measuring stations. MDF triangulation and TDOA have been implemented in the current study. Due to the limitations and distance errors associated with each method based on the topology of the stations, it is proposed in this study to combine the implemented algorithms to optimize the results and achieve the localization of the lightning point with the lowest possible error. The azimuth error of each crossed-loop antenna should be taken into consideration. If a three-station MDF system is used, each antenna provides a lightning direction with bearing errors,

resulting in intersecting lines at each station that determine a location area, as shown in Figure 8. The site error is caused by the nearby structures, such as buildings, power lines, and cables. On the contrary, the total random error arisen from non-vertical channels, the noise of background, and MDF electronics is usually 1–2°.
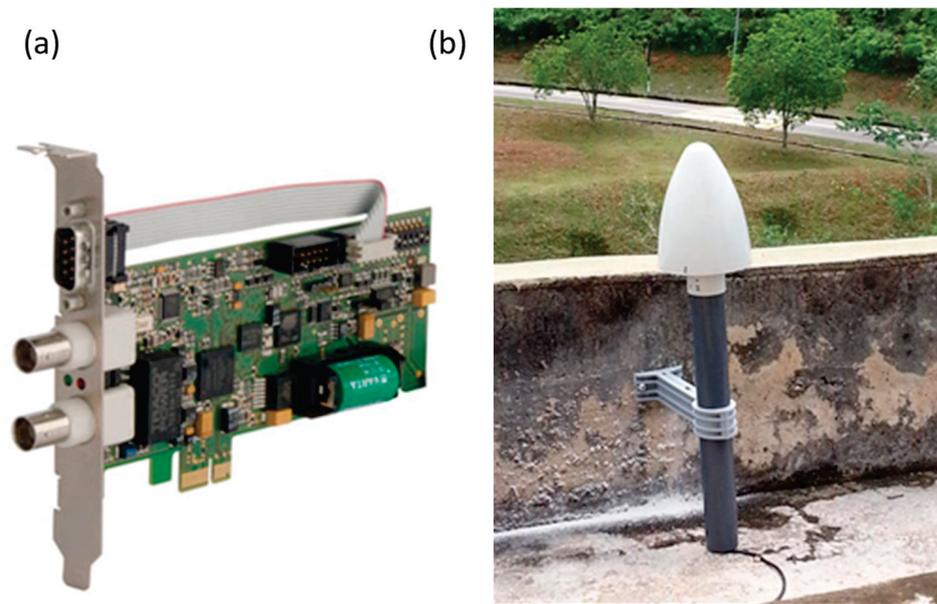
(a)   (b)

Figure 7. GPS180PEX: (**a**) Low Profile GPS Clock (PCI Express); (**b**) GPS antenna installed on the roof.
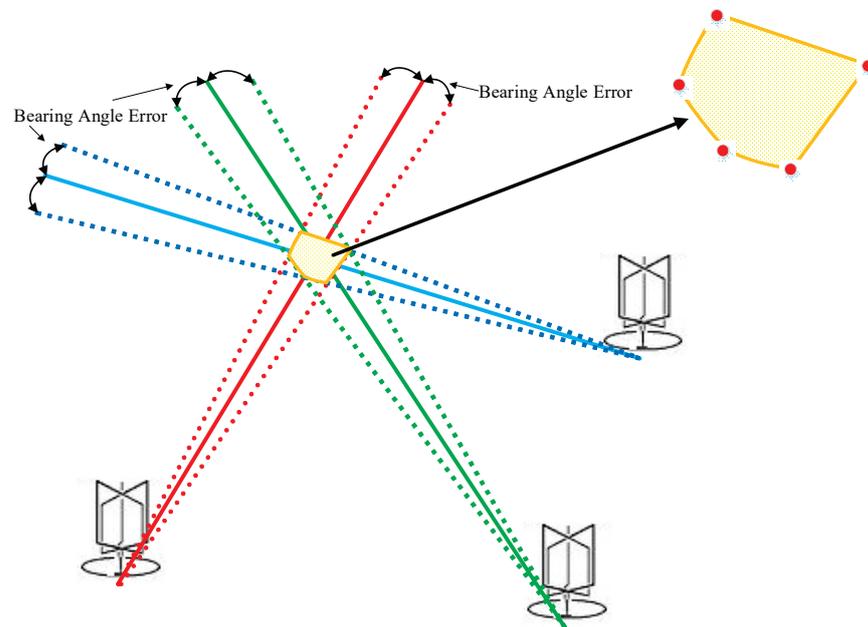
Figure 8. Location errors of three magnetic loop antennas with bearing errors of the incoming waveforms.

The TDOA based on three stations may produce more than one result (one real point and one or more ambiguous points). Regarding the localization methods, to make use of the intersection of hyperbolas in the TDOA method, it is required to install at least four stations to measure lightning flash to provide one unambiguous solution. These results provide other candidates for the PSO algorithm.

3.2. Particle Swarm Optimization (PSO)

According to the references [51–53], PSO as an artificial intelligence (AI) technique is employed to find the approximate point of the lightning strike using available particles. The PSO algorithm employs a population of individuals to explore promising regions of the search space in a synchronous manner. In this algorithm, the population is referred to as a swarm, and the individuals or search points are called particles. Each particle in the swarm represents a candidate solution in the optimization problem. The PSO algorithm enables particles to explore the search space by adjusting their position through an adaptable velocity. During this process, particles take into account their own experience as well as that of their neighboring particles. They also retain the memory of the best position they have encountered. This way, each particle utilizes its own and its neighbors' best positions to position itself towards the global minimum. The effect is that particles "fly" towards the global minimum, while still searching a wide area around the best solution [51–53]. The performance of each particle (closeness of a particle to the global minimum) is measured according to a predefined fitness function, which is related to the problem being solved.

Eberhart and Kennedy in 1995 developed PSO as a global optimization technique [54–56]. PSO exhibited good performance in finding solutions to static optimization problems. A particle swarm optimization algorithm was applied in this study as the combination mediator to find the optimum point of the lightning strike. PSO is a search engine whose working principle is based on the social information sharing of a swarm [57–59].

PSO models the swarm as particles existing in multi-dimensional space. These particles maintain a record of their personal best position and have knowledge of the global best position. Communication among particles in the swarm occurs by modifying their position and velocity [60]. In order to give a better understanding of developing the PSO algorithm, a flowchart is presented in Figure 9 as the basic or standard PSO.
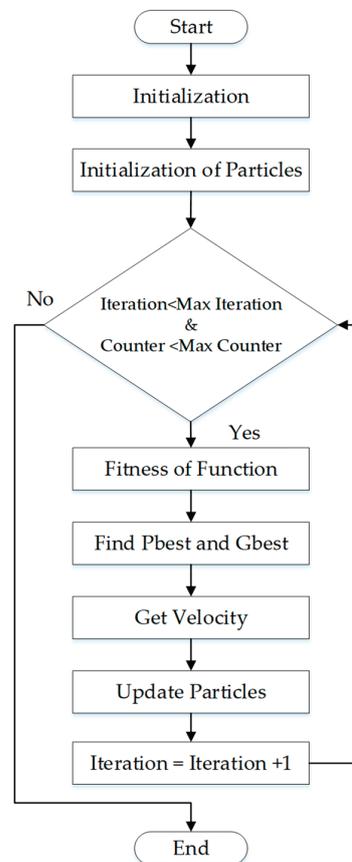
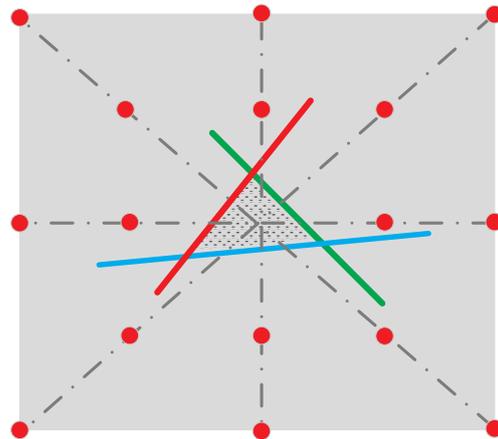**Figure 9.** Particle Swarm Optimization search engine flowchart.

First, the algorithm starts by initializing all the related parameters that will be used in the algorithm. Depending on the number of tuning parameters, a set of particles will be initialized. The initialization will be based on the calculated numbers and random values. The random values are obtained using the defined range of search, which is set in the earlier stage (see Equation (1)). During particle initialization, the fitness of each parameter will also be evaluated.

$$\text{initialization} = \text{range min} + (\text{range max} - \text{range min}) \times \text{random number} \tag{1}$$

After the initializations, the PSO will start its search for finding the parameter that gives optimum performance.

3.3. Distance and Azimuth

Azimuth in the context of a navigation system refers to the horizontal angle measured in a clockwise direction from a north reference line. The azimuth of a point can be more generally defined as an angle from a fixed reference point due to north. The geographical distance is the distance measured along the surface of the earth. Knowing the distance and azimuth to a reference point, it is possible to calculate the location (longitude and latitude) of a new point relative to that reference point. We suggest the application of the available azimuth and distance formulas be used in this research to provide auxiliary points for the PSO algorithm (see Figure 10).

Figure 10. Assumed red circle points (longitudes, latitudes) around the MDF triangle with certain distances (based on the settings) as inputs for the particle swarm optimization algorithm.

3.3.1. Distance and Azimuth Calculations between Two Coordinates on Map

Determining the distance and azimuth between two locations is a crucial aspect of spatial analysis in various fields such as industry and research. This section outlines the basic concepts and methods involved in calculating the geographic distance and azimuth. The mathematical calculations in this study were according to coordinates (longitude and latitude), which were used in GPS devices and maps. There were two essential formulas for finding the distance and azimuth between two points.

3.3.2. Azimuth Calculation

The reference plane for an azimuth is typically through the north, measured as a $0°$ azimuth, though other angular units (grad, mil) can be used. Moving clockwise on a 360-degree circle, the east has an azimuth of $90°$, the south $180°$, and the west $270°$. There is a command in MATLAB that calculates the azimuth between two coordinates as presented in Equation (2) [61]. Consider two points, X and Y. Point (X) includes (X_lat,X_lon) and Point (Y) contains (Y_lat,Y_lon). X_lat is the latitude of point X (for example $1.560973°$) and

X_lon is the longitude of point X (for example 103.643110°). Y_lat is the latitude of point Y and Y_lon is the longitude of point Y,

$$\text{Angle } (\theta \text{ or "Y regards to X"}) = \text{Azimuth } (X_{lat}, X_{lon}, Y_{lat}, Y_{lon}) \tag{2}$$

Angle is the calculated angle θ with reference to the North at point (X).

3.3.3. Distance Calculation

The formulas in this section calculate the distance between two points, which are defined by the geographical coordinates in terms of latitude and longitude.

Consider two points X and Y:

Point (X) is (X_{lat}, X_{lon}), point (Y) is (Y_{lat}, Y_{lon}). X_{lat} is the latitude of point X (for example 1.560973°) and X_{lon} is the longitude of point X (for example 103.643110°). While Y_{lat} is the latitude of point Y and Y_{lon} is the longitude of point Y, and "Π" is 3.1415.

Convert those to radians,

$$X_{lat}(r) = X_{lat} \times (\pi/180) \tag{3}$$

$$X_{lon}(r) = X_{lon} \times (\pi/180) \tag{4}$$

$$Y_{lat}(r) = Y_{lat} \times (\pi/180) \tag{5}$$

$$Y_{lon}(r) = Y_{lon} \times (\pi/180) \tag{6}$$

abs is the absolute value of the number inside the parenthesis.

$$\text{Londif\_XY} = \text{abs } (X_{lon}(r) - Y_{lon}(r) \tag{7}$$

acos stands for Arccosine of the value inside the parenthesis.

$$\text{Radian\_distance} = \text{acos } (\sin (X_{lat}(r)) \times \sin (Y_{lat}(r))) + \cos (X_{lat}(r)) \times \\ \cos (Y_{lat}(r) \times \cos(\text{Londif\_XY})) \tag{8}$$

$$\text{Naut\_distance} = \text{Radian\_distance} \times (3437.74677) \tag{9}$$

$$\text{Distance\_meter} = \text{Naut\_distance} \times (1852) \tag{10}$$

A distance meter is a unit of measurement used to quantify the distance between two points in meters. By utilizing the equations presented in Equations (3) to (10) for distance and Equation (2) for azimuth, the distance and azimuth between two coordinates on a map can be calculated.

3.3.4. Creating New Points from a Reference Point (Center of MDF Triangle) Based on Known Distance and Azimuth

It is possible to find the longitude and latitude of a new point referring to a reference point if the distance and azimuth to that reference point are known. The reference point in this study is the center point of the MDF triangle. Equations (11) to (16) show the mathematical calculations for finding the second point with an imagined distance and azimuth with regards to point X (considered as the reference point) [61].

Consider two points X and Y. It is required to find the longitude and latitude of a new point (Y), which has a distance of Distance$_{\text{meter}}$ in meter and Angle (θ) in degree with reference to point X. Sind and cosd are Sine and Cosine of the values inside the parenthesis.

$$D(X) = \text{Distance}_{\text{meter}} \times \sin d(\theta) \tag{11}$$

$$D(Y) = \text{Distance}_{\text{meter}} \times \cos d(\theta) \tag{12}$$

$$\Delta(\text{Longitude}(Y)) = D(X)/(111320 \times \cos d(X_{lat})) \tag{13}$$

$$\Delta(\text{Latitude}(Y)) = D(Y)/(110540) \tag{14}$$

$$\text{Longitude}(Y) = X_{lon} + \Delta(\text{Longitude}(Y)) \tag{15}$$

$$\text{Latitude}(Y) = X_{lat} + \Delta(\text{Latitude}(Y)) \tag{16}$$

Y_{lat} is the latitude of point Y and Y_{lon} is the longitude of point Y. The new calculated point is Point $(Y) = (Y_{lat}, Y_{lon})$.

3.4. PSO-Based Lightning Locating System

Optimization of the findings and results from the MDF method, TDOA method, and auxiliary points requires a mathematical technique. The LLS utilizes a combination of two or three detection methods to solve optimization problems. Many optimization problems have used the particle swarm optimization algorithm as a relatively new, modern, and powerful method for optimization issues. It is widely used to find the global optimum solution in a complex search space. The PSO algorithm was employed to optimize and suggest the accurate location of the lightning strikes based on the available results. The performance of the PSO-based algorithm is known to be affected by the arrangement of the searching process [62,63]. Figure 11 illustrates the flowchart of a combined LLS, which is employed in the current study. Much efforts have been made to design a suitable algorithm to filter the cloud activities and find the CG discharges. The detected waveforms are classified using discrimination methods and techniques as in Ref. [38].

Table 1 shows the results of the MDF and TDOA methods, which were used to form the input particles for the PSO initialization. The position of each particle is represented by a vector of parameters to be optimized. These parameters could represent values for variables in a mathematical function or parameters in a machine learning model. The particles are usually randomly generated within the search space, and their initial velocities can also be set randomly or uniformly. It is important to ensure that the initial particles are diverse enough to explore different regions of the search space, while also being sufficiently close to promising regions to converge quickly. Additionally, the number of particles is another parameter that needs to be carefully chosen to balance between exploration and exploitation. A minimum of 30 particles were determined based on the selected study area of the LLS. The possible input particles were added to the PSO algorithm to improve accuracy and reduce computation time.

The PSO algorithm starts with a set of particles or solutions and iteratively searches for the optimal solution by updating the generations. In each iteration, the particles are updated by considering two "best" values. The first one is the best solution (fitness). It has been achieved so far (the fitness value is also stored). This value is called pbest. Another "best" value that is tracked by the particle swarm optimizer is the best value, obtained so far by any particle in the population. This best value is a global best and called gbest. The particles are updated at each iteration and move to the result as presented in Figure 12. The calculated point would then be selected as the winner of PSO, if the answer point was better than the best answer of the last loop. The loop would continue until reaching the target or passes the maximum epoch. Max epoch refers to the maximum number of iterations or generations that the algorithm will run. It is a parameter that is set before running the PSO algorithm and controls how long the particles will search for the optimal solution. The number of epochs is typically set based on the complexity of the problem being solved and desired level of accuracy.
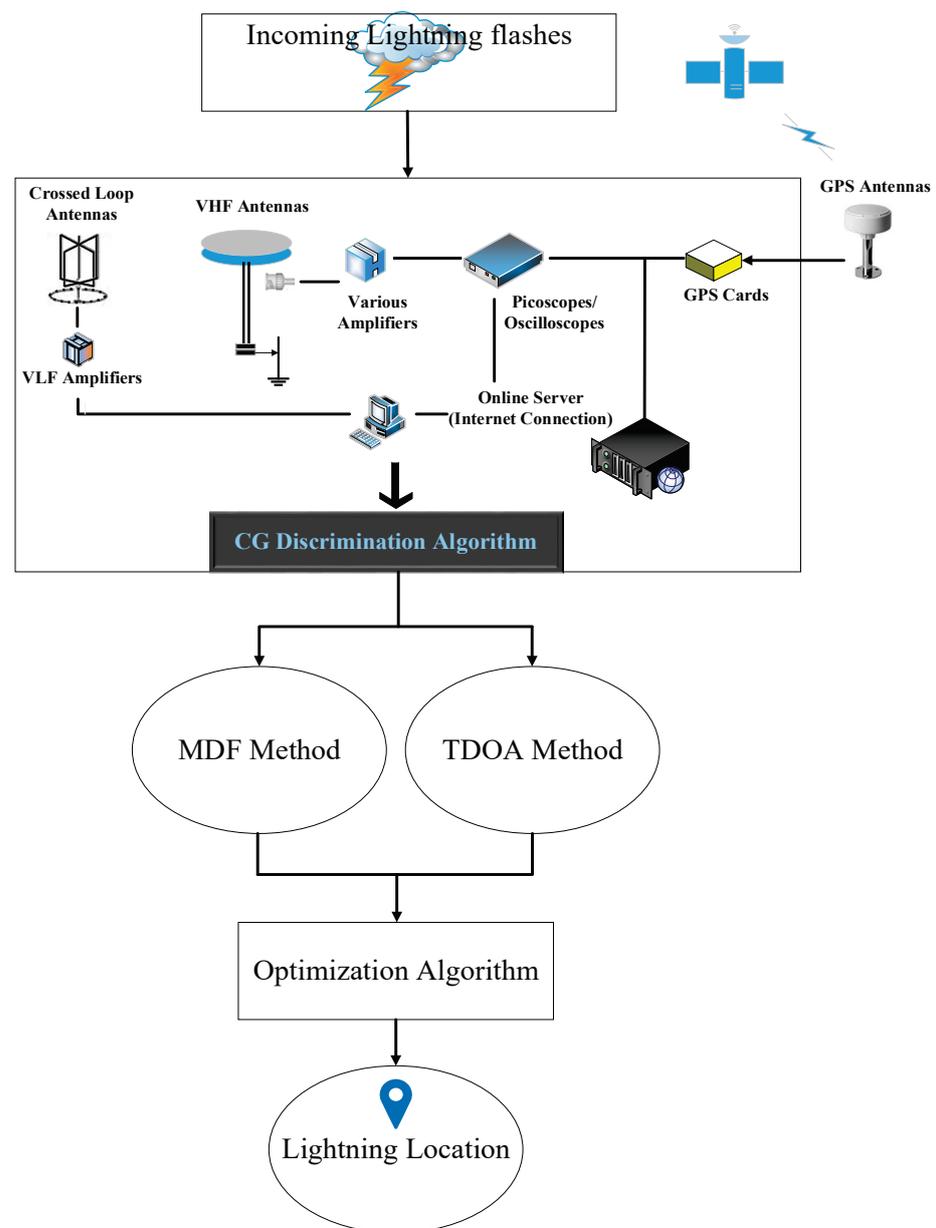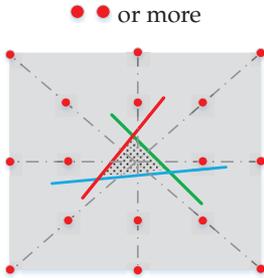
Figure 11. An overview of a methodological flowchart of the combined lightning locating system.

Table 1. Initializing the PSO algorithm.

Initial Points (Particles)	Points
MDF Method (5 points)	
MDF Method (7 points)	

Table 1. *Cont.*

Initial Points (Particles)	Points
TDOA Method (2 or more points)	
Using Azimuth and Distance (16 points)	
All input points (latitudes and longitudes)	
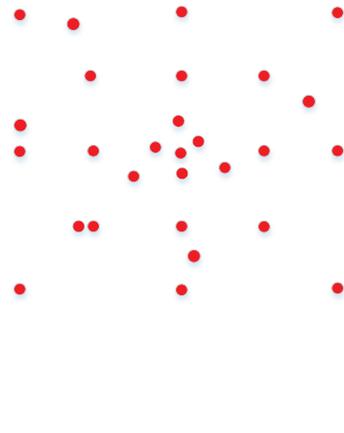
In the current research, the max epoch of 100 was selected based on several experiments. Setting the coverage range for a PSO algorithm depends on the specific problem being solved and objectives of the optimization. Generally, the coverage range refers to the range of values that the particles are allowed to explore during the optimization process. This range can be defined by setting appropriate boundaries for the particle positions, velocities, and acceleration coefficients. As mentioned in Figure 3, the case study is the coverage range for the PSO algorithm in this research. The next step in the flowchart of PSO in Figure 12 is assigning matrices of data. The positions and velocities of particles are represented as vectors of numerical values, which can be thought of as matrices. These vectors are updated in each iteration of the algorithm, based on the particle's own best position and the global best position in the swarm. The particle's position and velocity vectors are used to evaluate the objective function of the optimization problem, and to determine the new position and velocity in the next iteration. Therefore, the use of vectors is essential to the implementation of PSO. Then, the targets are set in the flowchart in Figure 12. The target is to optimize a given objective function. The particles in the swarm adjust their position and velocity based on their personal best and global best values in order to converge towards the optimal solution. The target of PSO is to find the global optimum solution by iteratively updating the positions and velocities of particles in the swarm. Evaluating the particles is the next step in Figure 12. Each candidate solution in PSO is represented by a particle that undergoes an iterative evaluation process based on a fitness function tailored to the optimization problem. The fitness function is used to assess the quality of the solution proposed by each particle. The iteration continues until a stopping criterion is met, such as the convergence of the fitness function or reaching the maximum number of iterations. Subsequently, the particles adjust their position and velocity based on their personal and global best to seek a more optimal solution. The

winner is the optimal solution to the given problem in this algorithm. In PSO, each particle represents a potential solution to the problem and is evaluated using a fitness function that measures its quality. The "gbest" refers to the global best position or solution found by the swarm of particles. This represents the best solution found by any of the particles in the swarm so far. The particles use the gbest as a guide to adjust their position and velocity towards a better solution. The gbest is updated as new, better solutions are found by the particles in the swarm. The particles adjust their position and velocity based on their individual experience (personal best) and collective experience of the swarm (global best) to move towards the target solution as shown in Figure 13.
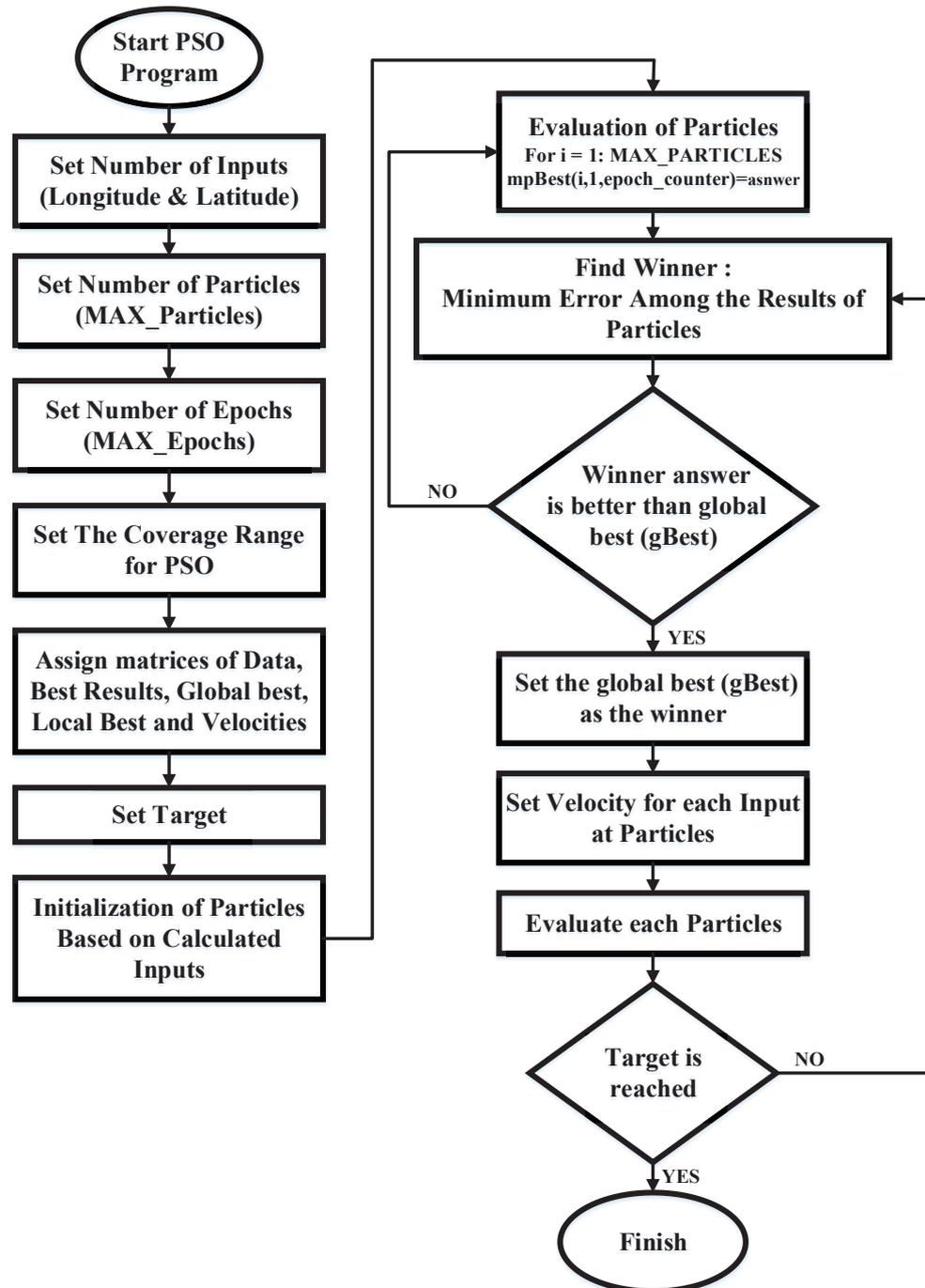
Figure 12. Flowchart of PSO settings and calculating process to find lightning strike point.
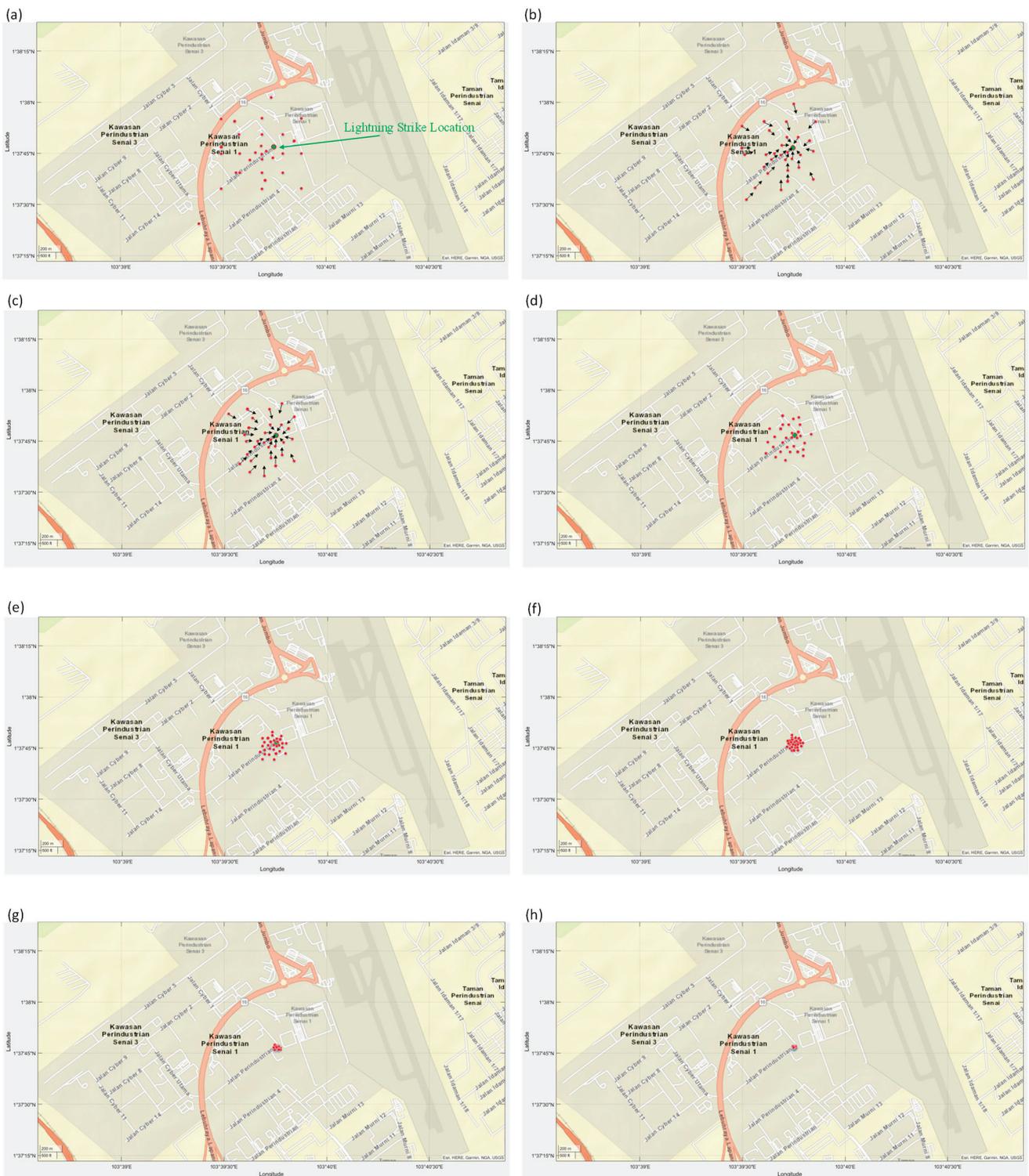
Figure 13. Sample of moving particles in PSO algorithm (**a**–**h**).

4. Results

In the preceding section, the approach for the lightning locating system was outlined. This section, however, focuses on the outcomes obtained from the proposed method. The experimental results are based on the measurements made using the proposed methods. If the waveform were recognized as a CG signal, then the signal would be packed with its corresponding timestamp and sent to the server station; otherwise, the signal would be rejected, and the system would be ready to sense another lightning flash. Figure 13

presents a sample of moving particles in the PSO algorithm. Note that only several steps of the algorithm are illustrated in Figure 13. All 30 input points for initializing the PSO algorithm were extracted based on MDF triangle, TDOA methods, and Azimuth and Distance. These points are initial points for the PSO algorithm. The PSO controlling parameters are tabulated in Table 2.

Table 2. Controlling Parameters of the PSO Algorithm.

Name	Number
Inputs (latitude, longitude)	2
Particles	30 ($\pm$1)
Max Epoch	100

PSO algorithm's loop starts by checking the number of counters and iterations. Then, the fitness function will be evaluated to find the best particle (P_{Pbest}) in that iteration. The process involves evaluating the fitness function of each particle during iterations and selecting the one with the best performance index as P best. Then, the P best is compared with the previous global best (g best) to find the new g best. It will set as a new g best if it is better, if not, the previous g best will remain. After that, a new velocity and position will be calculated. The new velocity is obtained using Equations (17) and (18) with an inertia weight (W). The new particle position is obtained by summing the current position with the new velocity value as expressed in Equation (19) (particle update rule). These steps are repeated until the stopping criteria are achieved.

$$W_k = W_{max} - (\text{iteration} \times (W_{max} - W_{min})/\text{max iteration}) \tag{17}$$

$$V_{i,k+1} = W_k V_{i,k} + C_1 \times r_1 \times (P_{Pbest,i,k} - X_{i,k}) + C_2 \times r_2 \times (P_{gbest,i,k} - X_{i,k})$$
$$\text{Intertia} : W_k V_{i,k}$$
$$\text{PersonalInfluence} : C_1 \times r_1 \times (P_{Pbest,i,k} - X_{i,k}) \tag{18}$$
$$\text{SocialInfluence} : C_2 \times r_2 \times (P_{gbest,i,k} - X_{i,k})$$

$$X_{i,k+1} = X_{i,k} + V_{i,k+1} \tag{19}$$

where,

$V_{i,k}$ = Velocity of the *ith* individual at iteration k.
W_k = Inertia weight at iteration k.
r_1 and r_2 = uniform random numbers of [0, 1].
C_1 and C_2 = acceleration factor between 0 and 2.
$X_{i,k}$ = Position of the ith individual at iteration k.
$P_{Pbest, i,k}$ = Best position of the ith individual at iteration k.
$P_{gbest, i,k}$ = Best position of the group until iteration k.
The PSO is initialized and then searches for optima by updating generations, as shown in Figure 13a. Updated candidate points in the PSO algorithm are presented in Figure 13b–h. The green circle is the real answer of the system.

4.1. Lightning Data

Lightning flash activities were monitored continuously in the study area, and the lightning locating systems recorded information such as location and time of discharge. This study focuses only on the cloud-to-ground discharges using the discriminating algorithm, which finds the CG discharges among all the cloud activities. The monthly CG lightning flashes for two months (November and December) are illustrated in Figure 14. All-day and night CG lightning discharges were collected to map the trend of the waveforms.
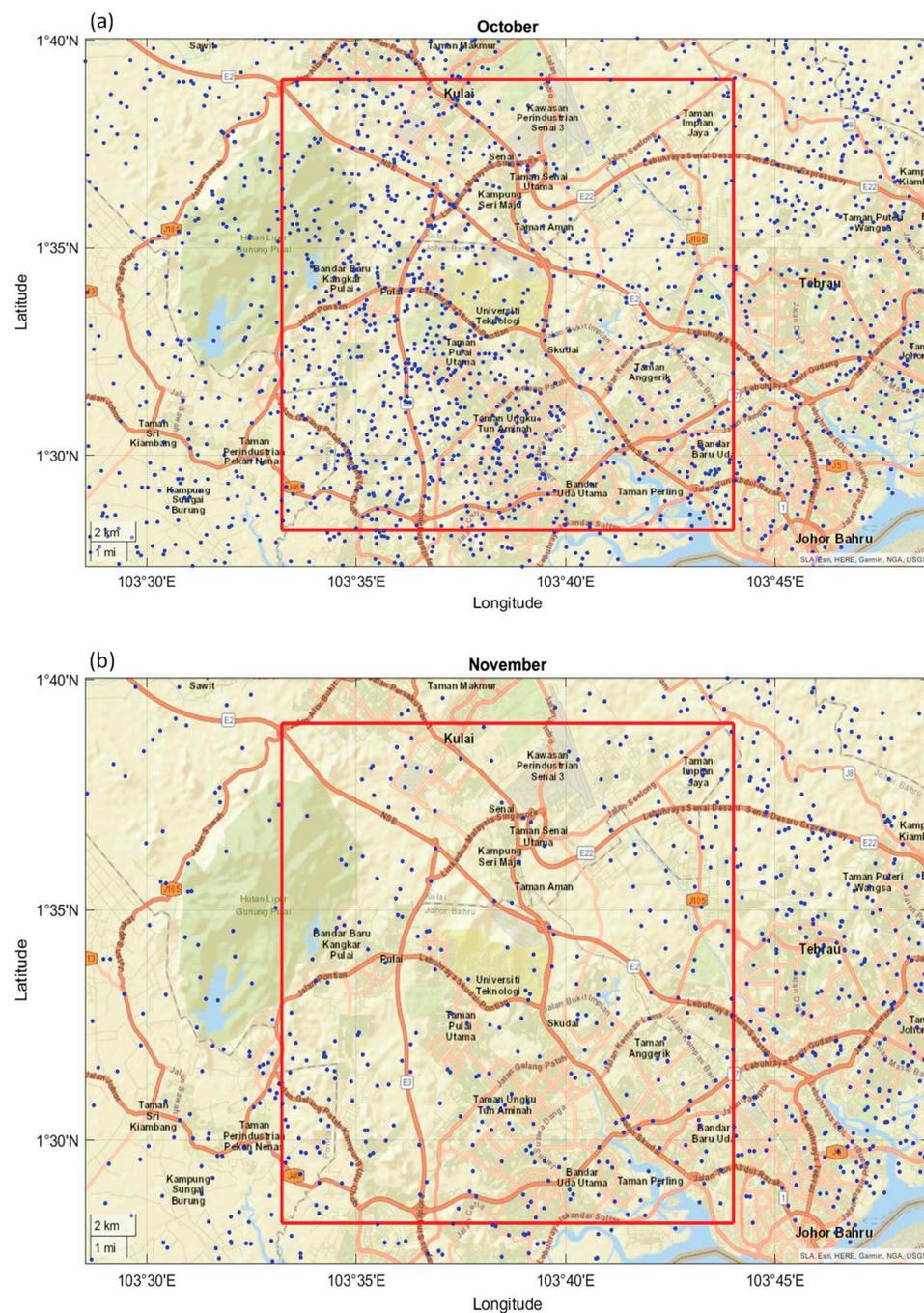
Figure 14. Lightning discharges around and within the study area; (**a**) October, (**b**) November.

During the two months under study, a high number of lightning discharges were detected. In order to implement the MDF-based LLS for a local area, three MDF sensors were installed at Universiti Teknologi Malaysia in Johor. The lightning discharges were sensed by crossed-loop antennas at the stations. The installed crossed-loop antennas in three stations provided three available known parameters for the LLS system at each incoming waveform. The detected waveforms were then saved continuously at each station. The MDF method was employed to find the direction of incoming waveforms. Three lines were then drawn based on calculated angles in the MDF method, which are a line with angle (θ_{IVAT} or $\theta_{IVAT} + 180$), a line with angle (θ_{B11} or $\theta_{B11} + 180$), and a line with angle ($\theta_{\_VAN}$ or $\theta_{\_VAN} + 180$). The intersections of the lines formed a triangle where the lightning strike would be inside or around this triangle.

The MDF method can have bearing errors that can cause angle deviations in lightning locating systems. This can result in situations where the lines drawn at each station do not intersect, especially when the distance between the stations is relatively short. To solve this problem, in this research, up to $\pm 5°$ of error was automatically added to the MDF calculation for getting the intersections of three drawn lines at the stations. The installed high-resolution GPS antennas provided three known parameters for the LLS system at each incoming waveform. These available parameters are t_{IVAT}, $t_{B11,}$ and t_{VAN}. The TDOA method used the time difference of arrivals in three stations ($\Delta t_{IVAT-B11}$, $\Delta t_{IVAT-VAN}$ and $\Delta t_{B11-VAN}$) to form the hyperbola equations to find the lightning strike point. In some conditions, hyperbola equations produced by only three sensors will result in two or more solutions or points, thus leading to ambiguous locations. The longitudes and latitudes of the points in MDF and TDOA methods had been extracted for the PSO algorithm. To calculate the average distance error of this system, only the captured lightning discharges inside the study area were considered.

4.2. Inter-Comparison Analysis of the PSO-Based LLS with an Industrial LLS in Malaysia

Comparing different lightning locating systems (LLSs) can be challenging since these systems operate at different frequency ranges, and therefore, may detect different aspects of lightning processes or flashes [64]. Different LLSs employ various technologies and detection methods such as optical methods or radiofrequency signals. Hence, careful consideration is required to compare the performance of the lightning locating systems. The lightning strikes that were detected by the PSO-based LLS during the months of December and March were selected for comparison with another industrial lightning detection system during the same time periods. The results of the comparison were presented in Table 3, which provides statistical analysis of the lightning strikes detected by both the industrial LLS and the PSO-based LLS proposed in this study. This section focuses on the number and abundance of detected lightning flashes by both systems. The last column contains the distances between two points (PSO-based LLS and Industrial LLS). The tabulated lightning strikes in Tables 3 and 4 are illustrated in Figures 15 and 16. The accuracy of a lightning detection system depends on several factors. Among them, the type and number of installed sensors, location of the sensors in relation to each other, sites' topography, surrounding obstructions, structures, and environment. Three stations are the minimum requirement for an LLS to compute data. Therefore, the sensors should not be placed apart more than their nominal range (200 km for Very High Frequency (VHF) sensors), and it would be a bad strategy to place them all in a straight line.

Table 3. Comparison of detected lightning flashes by PSO-based LLS and an industrial LLS in December.
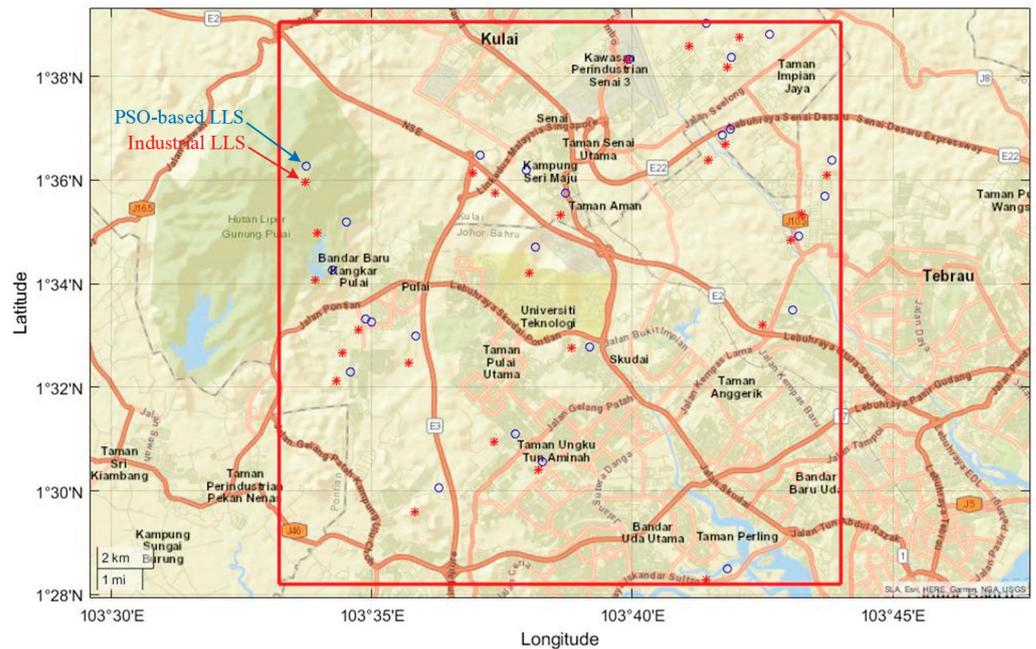
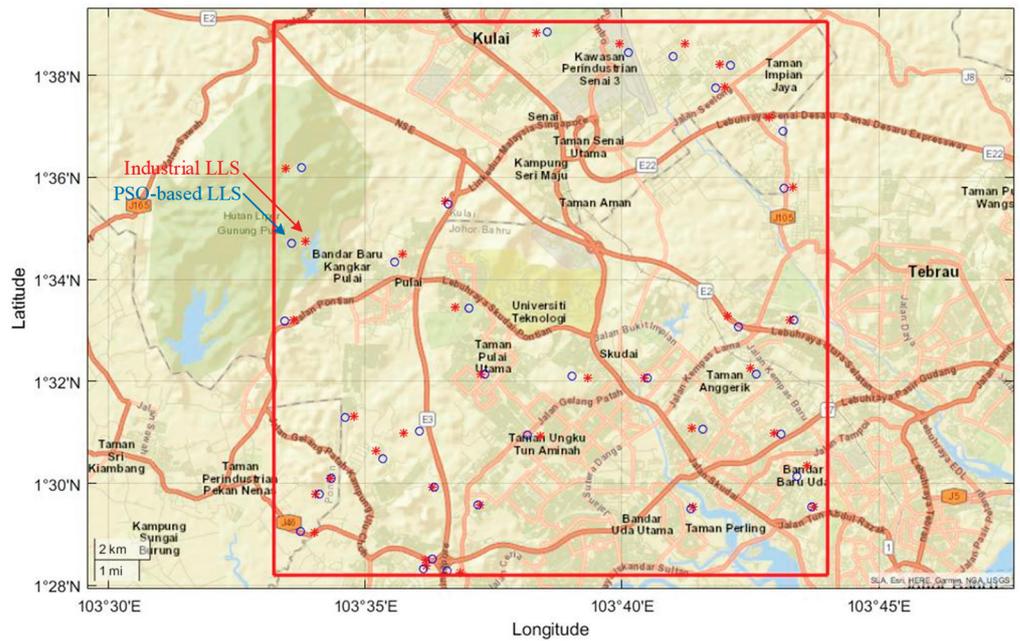| | PSO-Based LLS | | Industrial LSS | | |
Date/Time	Latitude	Longitude	Latitude	Longitude	Distance (m)
12-03/15:29:14.163	1.5093591	103.6379967	1.506859064	103.6365967	318.4
12-03/15:42:34.914	1.5183886	103.6292662	1.515788555	103.6225662	798.4
12-03/15:45:57.534	1.5497226	103.5972683	1.541122556	103.5952683	981.1
12-03/15:48:48.119	1.5553253	103.5813473	1.551825285	103.5791473	459.3
12-03/17:11:06.575	1.5463129	103.6529084	1.545912862	103.6473084	623.6
12-05/15:29:54.140	1.5382352	103.5765613	1.535335183	103.5718613	613.5
12-05/15:51:48.575	1.5011071	103.6047925	1.493307114	103.5969925	1225.5
12-05/15:53:31.207	1.5542215	103.5831449	1.544521451	103.5738449	1493.0
12-05/15:54:35.355	1.5710994	103.5705550	1.567799449	103.5651550	703.0
12-05/16:18:16.281	1.5948595	103.7283786	1.588959455	103.7208786	1060.1
12-05/16:35:36.495	1.6393911	103.6984139	1.636291146	103.6970139	377.9
12-05/17:06:06.618	1.6164738	103.6979654	1.611473799	103.6963654	583.3
12-06/12:26:13.807	1.4748742	103.6970687	1.471374154	103.6903687	839.7
12-06/12:41:42.274	1.6144218	103.6956172	1.606521845	103.6910172	1015.7

Table 3. *Cont.*

Date/Time	PSO-Based LLS		Industrial LSS		Distance (m)
	Latitude	Longitude	Latitude	Longitude	
12-06/13:01:06.473	1.6389945	103.6655620	1.638494492	103.6650620	78.6
12-06/13:17:41.495	1.6043931	103.5622202	1.599493146	103.5619202	545.5
12-10/15:12:29.891	1.6063280	103.7305589	1.601528049	103.7288589	565.8
12-10/15:21:25.714	1.6504658	103.6904365	1.642965794	103.6849365	1033.3
12-10/15:24:04.761	1.6468106	103.7104904	1.646010637	103.7007904	1081.1
12-10/15:26:28.987	1.5820068	103.7200615	1.580706835	103.7173615	332.9
12-15/16:54:06.940	1.5959161	103.6452386	1.588716149	103.6436386	819.6
12-18/18:27:07.548	1.5785293	103.6357051	1.570029259	103.6337051	970.3
12-18/18:30:28.760	1.6080253	103.6178829	1.602325320	103.6158829	671.2
12-18/18:32:41.326	1.5582600	103.7179885	1.553359985	103.7084885	1187.4
12-18/18:46:44.190	1.6031632	103.6327409	1.595763206	103.6228409	1373.1
12-18/18:49:20.753	1.5866248	103.5751790	1.582824826	103.5659790	1105.7
				Average	802

Table 4. Comparison of detected lightning flashes by PSO-based LLS and an industrial LLS in March.

Date/Time	PSO-Based LLS		Industrial LSS		Distance (m)
	Latitude	Longitude	Latitude	Longitude	
03-03/16:17:47.408	1.498726555	103.6058353	1.498801555	103.6050853	83.7
03-03/16:18:57.490	1.475235751	103.6050851	1.475010751	103.6028351	251.2
03-03/16:25:25.253	1.472001384	103.6021685	1.473001384	103.6031685	157.1
03-08/18:13:21.474	1.553102872	103.5571763	1.553382872	103.5599763	312.6
03-16/12:14:08.584	1.517029717	103.60079	1.516539717	103.59589	547.0
03-16/13:09:58.838	1.551242461	103.7044218	1.554592461	103.7010718	526.3
03-16/13:23:57.110	1.578575418	103.5594308	1.579030418	103.5639808	507.9
03-17/12:08:16.645	1.515875604	103.6359874	1.515470604	103.6400374	452.1
03-17/12:09:54.275	1.535540296	103.7103542	1.537540296	103.7083542	314.2
03-17/12:24:10.487	1.557225164	103.616924	1.557665164	103.612524	491.2
03-17/12:31:21.232	1.535005266	103.6506135	1.534510266	103.6555635	552.6
03-18/14:52:22.154	1.491793328	103.6890671	1.492443328	103.6897171	102.1
03-18/14:55:29.019	1.534398082	103.6747867	1.534473082	103.6740367	83.7
03-18/15:05:11.365	1.501574321	103.5724458	1.501534321	103.5720458	44.7
03-18/16:11:18.764	1.639561402	103.683245	1.643561402	103.687245	628.5
03-19/16:08:30.028	1.492144746	103.7281663	1.492234746	103.7290663	100.5
03-19/16:20:39.706	1.535748372	103.6222458	1.535608372	103.6208458	156.3
03-19/16:22:25.204	1.508182058	103.5889977	1.510482058	103.5866977	361.4
03-19/16:24:00.767	1.52143029	103.5768302	1.52173029	103.5798302	334.9
03-19/16:31:41.387	1.484294484	103.5625046	1.483854484	103.5669046	491.2
03-19/16:32:18.064	1.640890988	103.6688524	1.643740988	103.6660024	447.8
03-19/16:32:45.647	1.63659912	103.701985	1.63696412	103.698335	407.4
03-19/16:33:31.099	1.471336762	103.6097917	1.470891762	103.6142417	496.8
03-19/16:51:00.005	1.572182043	103.5927705	1.574932043	103.5955205	432.1
03-19/16:54:43.111	1.517641053	103.6929289	1.518006053	103.6892789	407.5
03-19/16:54:43.530	1.647467983	103.6425267	1.647112983	103.6389767	396.3
03-20/15:00:09.489	1.516248946	103.7181735	1.516473946	103.7159235	251.2
03-20/16:40:04.496	1.603316676	103.5625474	1.602821676	103.5575974	552.6
03-21/16:58:17.307	1.502410723	103.7232993	1.505910723	103.7267993	549.9
03-21/17:00:05.290	1.596392045	103.719162	1.596692045	103.722162	334.9
03-21/17:03:25.651	1.553462548	103.7225846	1.553327548	103.7212346	150.7
03-21/17:23:06.339	1.614956335	103.7190187	1.619556335	103.7144187	722.7
03-21/17:24:43.791	1.629298612	103.6969697	1.629588612	103.6998697	323.7
03-28/16:03:29.746	1.492877417	103.6197889	1.492792417	103.6206389	94.9
03-28/16:08:32.648	1.591213546	103.6104063	1.592263546	103.6093563	165.0
03-28/16:17:38.017	1.496498617	103.5685041	1.496638617	103.5671041	156.3
				Average	344

Figure 15. Locations of lightning events detected by PSO-based LLS and another industrial lightning detection system in December.

Figure 16. Locations of lightning events detected by PSO-based LLS and an industrial lightning detection system in March.

Based on the analysis, the average distance differences between PSO-based LLS in this research and an industrial LLS is around 573 m, which is acceptable in a region with an area of 400 km^2. This average is not a fixed number throughout the year and locations as the PSO-based LLS decides according to the circumstances. The national lightning detection systems cover nearly the whole of the county but cannot be totally relied on due to the large coverage area. On the other hand, the unexpected suspension of sensors operation in these systems may lead to loss of lightning data in a large area.

5. Discussion

The lightning discharges propagate electromagnetic fields in all directions. The installed measuring sensors are triggered by a radiated lightning flash, which immediately record the time and the waveforms. An integrated lightning locating system using three measurement stations was designed and installed at Universiti Teknologi Malaysia for detecting lightning strikes covering 400 km^2. The parallel plate and crossed loop antennas have been installed at each station to detect electric and magnetic fields of the lightning discharges respectively. In order to ensure high precision and accuracy, the clock of the stations in this study were synchronized with the UTC time reference to record the timestamp of the incoming waveforms. The timestamp was saved with a resolution of at least 100 ns to achieve high precision in the detection of lightning events. The detected waveforms were classified using the proposed algorithm based on wavelet analysis in Ref. [38] to extract and discriminate the ground flashes (CGs). The employed MDF method based on the magnetic field in this study capture, measure, and locate the lightning emissions.

The azimuth angle between the north and the direction of the detected return stroke in CG discharges was determined using a crossed loop antenna. To locate the lightning strike with some error, at least two sensors were needed. In contrast, the TDOA method required a minimum of three synchronized measuring stations. This approach determined the position of a CG discharge based on the time differences of a detected signal at different stations. The hyperbolics which are drawn between every two stations are the host of the lightning discharges as the time differences between the captured times are constant. Hence, at least two pairs of these hyperbolics are needed to find the possible locations of the CG discharge. The measuring stations in this study were calibrated and the time clocks were synchronized since this method depends on the accurate captured timestamps.

In this study, both the MDF and TDOA methods were utilized, and the main contribution was the optimization of the combined LLS. The first step in achieving this was to establish an appropriate solution space while considering the constraints of the system. After conducting investigations, the particle swarm optimization algorithm was chosen as the search algorithm for the optimization process. Several parameters in the PSO algorithm affect the performance of the algorithm. These values have significant impacts on the accuracy and efficiency of the PSO method. Several particles or swarm size and number of iterations are the fundamental parameters. The number of particles is the population size, and a big swarm generates larger search space in each iteration that help to obtain better optimization result. On the other hand, huge amounts of particles increase the complexity of the system and consumed time. The stopping criteria in this study were set to terminate the iterative search process.

In this study, a maximum of 100 iterations were chosen for the PSO-based LLS algorithm. However, it should be noted that setting a small maximum number of iterations could result in an inaccurate solution. This research has practical implications for various applications, including lightning tracking and mapping, which have numerous benefits for both academic and industrial purposes. The implemented PSO-based lightning locating system can provide the characteristics of lightning strikes that can be valuable to the governmental agencies as well as private sectors to design their protection systems. There are two main lightning detecting networks in Malaysia, which belong to Tenaga Nasional Berhad (TNB) and the Malaysian Meteorological Department (MMD). These systems cover the whole country, with detection efficiency and mean distance error of 90% and 500 m. TNB and MMD cover nearly the whole country but cannot be relied on due to the large coverage area. Therefore, a local lightning locating system will be helpful as a backup system of national LLSs for small regions. An automatic detecting and saving program was written in MATLAB software and installed at each station. The GPS timestamp was recorded when the incoming signal hit the threshold.

High-speed cameras can autonomously gather data regarding the location and characteristics of lightning strikes, making them an essential tool for validating the results produced by lightning locating systems [65]. These cameras possess the capability to cap-

ture and record lightning discharges at a rate of thousands of frames per second, which can identify the location of lightning strikes. Lightning locating systems rely on radiofrequency signals to detect and locate lightning strikes; however, various factors such as atmospheric conditions, the distance and geometry of the lightning channel, and the sensitivity of the system can influence the precision of the data produced [66]. In this study, the camera recordings were not accessible during the designated measurement period. Therefore, inter-comparison analysis was employed. It was conducted between the proposed particle swarm optimization (PSO)-based lightning location system (LLS) and an industrial LLS in Malaysia. To this end, a circular region with a 500-m radius was delineated around the lightning discharge captured by the industrial LLS and depicted in Figure 17. Given that industrial LLSs typically exhibit a location error of approximately 500 m, it is logical to postulate that the lightning strike may have occurred anywhere within this circular boundary. Consequently, the discrepancy in the estimated lightning location between the PSO-based LLS and industrial LLS could range from 50 m to approximately 1050 m.
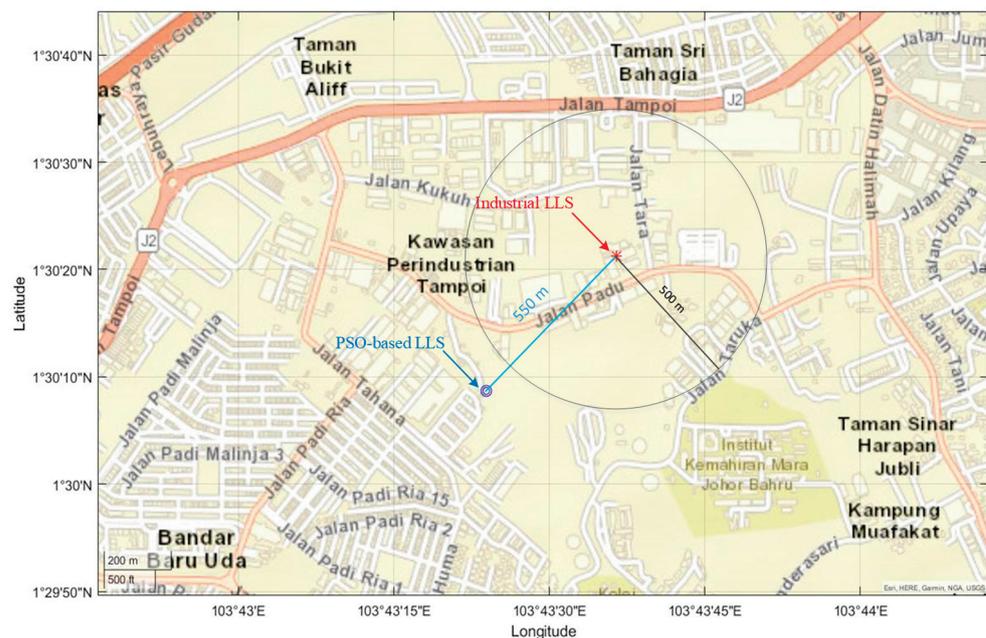
Figure 17. Locations of lightning events detected by PSO-based LLS (1.502410723, 103.7232993) and an industrial LLS (1.505910723, 103.7267993) on 21 March at 16:58:17.307.

In order to promote public awareness, it is important to educate children and the public about safety measures such as lightning protection systems for buildings, as these can significantly reduce the number of injuries caused by lightning strikes. It should be noted, however, that accurately locating the exact location of lightning strikes and predicting their behavior prior to occurrence is currently not possible. The PSO-based lightning locating system is a safety and warning system for detecting and alarming the public as soon as lightning discharge occurs. The results can be broadcasted to the local users in real time for lightning disaster early warning. In addition, a statistical data analysis is a tool for power engineers to protect their equipment and power stations from lightning strikes by improving the protection systems.

6. Conclusions

Malaysia is known to be one of the countries with a high frequency of lightning strikes in the world. According to the Malaysian Meteorological Services, the country experiences an average of 200 thunder days per year, indicating the high frequency of lightning and thunderstorm activity in the region. Knowing the total number of lightning flashes occurring worldwide is crucial in weather and meteorological research. The stations

installed in this study are capable of detecting and recording lightning discharges in both VHF and VLF frequency ranges. The collected data can be used to enhance lightning characterization studies in Malaysia, which is a tropical region known for its high lightning and thunderstorm activity.

Cloud-to-ground lightning is the most significant cause of faults and outages in power systems compared to other types of cloud discharges such as inter-cloud and cloud-to-cloud activities. Therefore, it is essential to monitor cloud activities worldwide, especially in tropical regions, which are the main hotspots of lightning strikes. The primary objective of designing lightning protection systems is to safeguard human lives and properties from lightning strikes. Moreover, national regulations in the field of lightning protection are issued to ensure compliance with safety standards.

A ground-based lightning locating system (LLS) comprises a minimum of three sensors and a central server, which work together to detect and locate electromagnetic waveforms generated by a lightning discharge. The sensors capture the signals and transmit them to the server for analysis. Once all sensors detect a cloud-to-ground (CG) lightning event, the central server uses both MDF and TDOA techniques to pinpoint the location of the CG discharge. However, in the MDF method, certain scenarios may arise where the discharge hits a point between two stations, posing a significant challenge for the LLS to accurately locate the discharge.

To address the issue of difficult cases where the hitting point is on the line between stations in the MDF method, at least three stations are required. The azimuth angle of a return stroke can be determined by analyzing the magnetic fields captured by crossed-loop antennas. The implemented PSO-based lightning locating system can provide information on lightning strikes. This study will be significant for public users to be aware of detected storms and estimation of imminent rainfalls.

The integration of MDF and TDOA techniques in the PSO-based LLS has resulted in an accurate lightning detection system with a mean location error of 573 m (up to 573 m) for a specific local region. This development has enhanced the location accuracy of lightning strikes in the region.

The study involved a comparative analysis between a system under investigation and an industrial network located in the southern region of Malaysia, which spanned a predetermined timeframe. Based on the results of the comparison, it was found that the particle swarm optimization (PSO) algorithm-based lightning location system (LLS) exhibited an average distance of 802 m and 344 m in December and March, respectively, when contrasted against the industrial network for detecting multiple lightning flashes. Based on the results obtained, it can be inferred that the LLS using PSO is proficient in precisely identifying and charting lightning discharges in the designated coverage region, thereby qualifying as an effective lightning detection system. The configuration and quantity of stations directly impacted the precision and accuracy of the LLS. The majority of distance deviations and errors were observed along lines passing through two stations. The efficacy of LLS is measured by its capability to accurately geolocate lightning events with high detection efficiency and low false detection rates, as well as correctly report various lightning features.

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

References

1. Mehranzamir, K. *Lightning Ground Flash Locating System Based on Combined Sensing Method Using Artificial Neural Network and Particle Swarm Optimization*; Universiti Teknologi Malaysia: Skudai, Malaysia, 2015.
2. Rakov, V.A. Electromagnetic methods of lightning location. In *Fundamentals of Lightning*; Cambridge University Press: Cambridge, UK, 2016; pp. 161–177.
3. Mehranzamir, K.; Salimi, B.; Abdul-Malek, Z. Comparative study of lightning models with lightning discharges in Malaysia. In Proceedings of the 2013 Annual Report Conference on Electrical Insulation and Dielectric Phenomena, Chenzhen, China, 20–23 October 2013; pp. 1085–1088. [CrossRef]
4. Rakov, V.A.; Martin, A.U. *Lightning: Physics and Effects*; Cambridge University Press: Cambridge, UK, 2007.
5. Wooi, C.-L.; Abdul-Malek, Z.; Salimi, B.; Ahmad, N.A.; Mehranzamir, K.; Vahabi-Mashak, S. A Comparative Study on the Positive Lightning Return Stroke Electric Fields in Different Meteorological Conditions. *Adv. Meteorol.* **2015**, *2015*, 1–12. [CrossRef]
6. Mills, B.; Unrau, D.; Parkinson, C.; Jones, B.; Yessis, J.; Spring, K.; Pentelow, L. Assessment of lightning-related fatality and injury risk in Canada. *Nat. Hazards* **2008**, *47*, 157–183. [CrossRef]
7. Cherington, M.; Yarnell, P.R.; London, S.F. Neurologic complications of lightning injuries. *West. J. Med.* **1995**, *162*, 413–417. [PubMed]
8. Cooper, M.A.; Holle, R.L.; Tushemereirwe, R. Mitigating the Hazard of Lightning Injury and Death across Africa, Public Health in Developing Countries—Challenges and Opportunities. In *Public Health in Developing Countries—Challenges and Opportunities*; IntechOpen: London, UK, 2019.
9. Rojas-Cubides, H.E.; Cruz-Bernal, A.S.; Cortes-Guerrero, C.A. Characteristics of electric field waveforms produced by negative return strokes in Colombia and their comparison with other regions. *J. Atmos. Sol. Terr. Phys.* **2021**, *227*, 105809. [CrossRef]
10. Koshak, W.J.; Solakiewicz, R.J. A Method for Retrieving the Ground Flash Fraction and Flash Type from Satellite Lightning Mapper Observations. *J. Atmos. Ocean. Technol.* **2015**, *32*, 79–96. [CrossRef]
11. Yaniv, R.; Yair, Y.; Price, C.; Mkrtchyan, H.; Lynn, B.; Reymers, A. Ground-based measurements of the vertical E-field in mountainous regions and the "Austausch" effect. *Atmos. Res.* **2017**, *189*, 127–133. [CrossRef]
12. Sokol, Z.; Popová, J. Differences in Cloud Radar Phase and Power in Co- and Cross-Channel—Indicator of Lightning. *Remote. Sens.* **2021**, *13*, 503. [CrossRef]
13. Norinder, H. Long-distance location of thunderstorms. In *Thunderstorm Electricity*; Univ. Chicago Press: Chicago, IL, USA, 1953; pp. 276–327.
14. Rakov, V.A. Electromagnetic Methods of Lightning Detection. *Surv. Geophys.* **2013**, *34*, 731–753. [CrossRef]
15. Wang, J.; Ma, Q.; Zhou, X.; Xiao, F.; Yuan, S.; Chang, S.; He, J.; Wang, H.; Huang, Q. Asia-Pacific Lightning Location Network (APLLN) and Preliminary Performance Assessment. *Remote. Sens.* **2020**, *12*, 1537. [CrossRef]
16. Beirle, S.; Spichtinger, N.; Stohl, A.; Cummins, K.L.; Turner, T.; Boccippio, D.; Cooper, O.R.; Wenig, M.; Grzegorski, M.; Platt, U.; et al. Estimating the NO(x) produced by lightning from GOME and NLDN data: A case study in the Gulf of Mexico. *Atmos. Chem. Phys.* **2006**, *6*, 1075–1089. [CrossRef]
17. Boersma, K.F.; Eskes, H.J.; Meijer, E.W.; Kelder, H.M. Estimates of lightning NO(x) production from GOME satellite observations. *Atmos. Chem. Phys.* **2005**, *5*, 2311–2331. (In English) [CrossRef]
18. Bond, D.W.; Zhang, R.; Tie, X.; Brasseur, G.; Huffines, G.; Orville, R.E.; Boccippio, D.J. NOx production by lightning over the continental United States. *J. Geophys. Res. Atmos.* **2001**, *106*, 27701–27710. (In English) [CrossRef]
19. Van der A, R.J.; Eskes, H.J.; Boersma, K.F.; van Noije, T.P.C.; Van Roozendael, M.; De Smedt, I.; Peters, D.H.M.U.; Meijer, E.W. Trends, seasonal variability and dominant NOx source derived from a ten year record of NO_2 measured from space. *J. Geophys. Res. Atmos.* **2008**, 113. (In English) [CrossRef]
20. Krider, E.P.; Noggle, R.C.; Pifer, A.E.; Vance, D.L. Lightning Direction-Finding Systems for Forest Fire Detection. *Bull. Am. Meteorol. Soc.* **1980**, *61*, 980–986. [CrossRef]
21. Lyons, W.A.; Nelson, T.E.; Williams, E.R.; Cramer, J.A.; Turner, T.R. Enhanced Positive Cloud-to-Ground Lightning in Thunderstorms Ingesting Smoke from Fires. *Science* **1998**, *282*, 77–80. [CrossRef]
22. Arevalo, L.; Cooray, V. On the interception of lightning flashes by power transmission lines. *J. Electrost.* **2011**, *69*, 220–227. [CrossRef]
23. Ekonomou, L. High voltage transmission lines studies with the use of artificial intelligence. *Electr. Power Syst. Res.* **2009**, *79*, 1655–1660. [CrossRef]
24. Hashim, M.M.I.; Ping, H.W.; Ramachandaramurthy, V.K. Impedance-based fault location techniques for transmission lines. In Proceedings of the TENCON 2009—2009 IEEE Region 10 Conference, Singapore, 23–26 January 2009; pp. 1–6. [CrossRef]
25. Mehranzamir, K.; Salimi, B.; Abdul-Malek, Z. Investigation of preliminary breakdown pulses in lightning waveforms. In Proceedings of the Progress in Electromagnetics Research Symposium, Stockholm, Sweden, 20 April 2013; pp. 1542–1546.
26. Nag, A.; Murphy, M.J.; Schulz, W.; Cummins, K.L. Lightning locating systems: Insights on characteristics and validation techniques. *Earth Space Sci.* **2015**, *2*, 65–93. [CrossRef]

27. Mehranzamir, K.; Afrouzi, H.N.; Abdul-Malek, Z.; Nawawi, Z.; Sidik, M.A.B.; Jambak, M.I. Hardware Installation of Lightning Locating System Using Time Difference of Arrival Method. In Proceedings of the 2019 International Conference on Electrical Engineering and Computer Science (ICECOS), Batam, Indonesia, 2–3 October 2019; pp. 29–34. [CrossRef]

28. Salimi, B.; Abdul-Malek, Z.; Mirazimi, S.J.; Mehranzamir, K. Investigation of Short Base Line Lightning Detection System by Using Time of Arrival Method. In *Intelligent Informatics: Proceedings of the International Symposium on Intelligent Informatics ISI'12 Held at Chennai, India, 4–5 August 2012*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 182, pp. 141–147. [CrossRef]

29. Mehranzamir, K.; Afrouzi, H.N.; Abdul-Malek, Z.; Nafea, M.; Rufus, S.A. Detecting Sensor Coordination in a Calibrated Lightning Locating System. In Proceedings of the 2019 International Conference on Electrical Engineering and Computer Science (ICECOS), Batam, Indonesia, 2–3 October 2019; pp. 35–40.

30. Salimi, B.; Mehranzamir, K.; Abdul-Malek, Z. Statistical analysis of lightning electric field measured under Malaysian condition. *Asia Pac. J. Atmos. Sci.* **2013**, *50*, 133–137. [CrossRef]

31. Mehranzamir, K.; Afrouzi, H.N.; Abdul-Malek, Z.; Nawawi, Z.; Sidik, M.A.B.; Jambak, M.I. Hardware and Software Implementation of Magnetic Direction Finding Sensors. In Proceedings of the 2019 International Conference on Electrical Engineering and Computer Science (ICECOS), Batam, Indonesia, 2–3 October 2019; pp. 23–28.

32. Cummins, K.; Murphy, M.J.; Bardo, E.A.; Hiscox, W.L.; Pyle, R.B.; Pifer, A.E. A Combined TOA/MDF Technology Upgrade of the U.S. National Lightning Detection Network. *J. Geophys. Res. Atmos.* **1998**, *103*, 9035–9044. [CrossRef]

33. Wood, T.G.; Inan, U.S. Localization of individual lightning discharges via directional and temporal triangulation of sferic measurements at two distant sites. *J. Geophys. Res. Atmos.* **2004**, 109. [CrossRef]

34. López, R.E.; Passi, R.M. Simulations in site error estimation for direction finders. *J. Geophys. Res. Atmos.* **1991**, *96*, 15287. [CrossRef]

35. Ortéga, P. A three magnetic direction finder network for a local warning device. *J. Light. Res.* **2007**, *2*, 18–27.

36. Chen, M.; Lu, T.; Du, Y. Properties of "site error" of lightning direction-finder (DF) and its modelling. *Atmos. Res.* **2013**, *129–130*, 97–109. [CrossRef]

37. Mehranzamir, K.; Abdul-Malek, Z.; Salimi, B.; Ahmad, N.A. Observation of Isolated Breakdown Lightning Flashes in a Tropical Region. *Appl. Mech. Mater.* **2014**, *554*, 583–587. [CrossRef]

38. Mehranzamir, K.; Davarpanah, M.; Abdul-Malek, Z.; Afrouzi, H.N. Discriminating cloud to ground lightning flashes based on wavelet analysis of electric field signals. *J. Atmos. Sol. Terr. Phys.* **2018**, *181*, 127–140. [CrossRef]

39. Salimi, B.; Abdul-Malek, Z.; Mehranzamir, K.; Mashak, S.V.; Afrouzi, H.N. Localised Single-Station Lightning Detection by Using TOA Method. *J. Teknol.* **2013**, *64*, 73–77. [CrossRef]

40. Hwang, Y.J.; Wooi, C.L.; Rohani, M.N.K.; Mehranzamir, K.; Arshad, S.N.M.; A Ahmad, N. Prototyping a RF signal-based lightning warning device using with Internet of Things (IOT) integration. *J. Phys. Conf. Ser.* **2020**, *1432*, 012078. [CrossRef]

41. Lojou, J.Y.; Cummins, K.L. Total Lightning Mapping using both VHF interferometry and Time of Arrival Technique. In Proceedings of the International Conference on Lightning Protection, Kanazawa, Japan, 18–22 September 2006; pp. 391–396.

42. Vahabi-Mashak, S.; Abdul-Malek, Z.; Mehranzamir, K.; Nabipour-Afrouzi, H.; Salimi, B.; Wooi, C.-L. Modeling of Time of Arrival Method for Lightning Locating Systems. *Adv. Meteorol.* **2015**, *2015*, 1–12. [CrossRef]

43. Mehranzamir, K.; Abdul-Malek, Z.; Afrouzi, H.N.; Mashak, S.V.; Wooi, C.-L.; Zarei, R. Artificial neural network application in an implemented lightning locating system. *J. Atmos. Sol. Terr. Phys.* **2020**, *210*, 105437. [CrossRef]

44. Boltek. Available online: http://www.boltek.com/ (accessed on 20 April 2023).

45. Mehranzamir, K.; Abdul-Malek, Z.; Salimi, B.; Ahmad, N.A. Return Strokes Measurements of Electric Field Produced by Lightning Discharges in Malaysia. *Appl. Mech. Mater.* **2014**, *554*, 618–622. [CrossRef]

46. Salimi, B.; Abdul-Malek, Z.; Mehranzamir, K.; Ahmad, N.A. Study on the Vertical Component of Lightning Electric Field during Monsoon Period in Malaysia. *Appl. Mech. Mater.* **2014**, *554*, 623–627. [CrossRef]

47. World Lightning Map. Available online: https://earthobservatory.nasa.gov/images/6679/patterns-of-lightning-activity (accessed on 20 April 2023).

48. Kilinc, M.; Beringer, J. The Spatial and Temporal Distribution of Lightning Strikes and Their Relationship with Vegetation Type, Elevation, and Fire Scars in the Northern Territory. *J. Clim.* **2007**, *20*, 1161–1173. (In English) [CrossRef]

49. Ramlee, N.A.; Ahmad, N.; Baharudin, Z.; Esa, M. Temporal analysis on pulse train of lightning discharge observed in Malacca, Malaysia. *Energy Rep.* **2023**, *9*, 618–625. [CrossRef]

50. T. Maps. 2021. Available online: https://en-gb.topographic-map.com/maps/zrb/Malaysia/ (accessed on 20 April 2023).

51. Kiran, R.; Jetti, S.R.; Venayagamoorthy, G.K. Online Training of a Generalized Neuron with Particle Swarm Optimization. In Proceedings of the 2006 IEEE International Joint Conference on Neural Network Proceedings, Vancouver, BC, Canada, 16–21 July 2006. [CrossRef]

52. Kwok, N.; Liu, D.; Tan, K.; Ha, Q. An Empirical Study on the Settings of Control Coefficients in Particle Swarm Optimization. In Proceedings of the 2006 IEEE International Joint Conference on Neural Network Proceedings, Vancouver, BC, Canada, 16–21 July 2006. [CrossRef]

53. Richer, T.; Blackwell, T. When is a Swarm Necessary? In Proceedings of the 2006 IEEE International Joint Conference on Neural Network Proceedings, Vancouver, BC, Canada, 16–21 July 2006. [CrossRef]

54. Eberhart, R.; Kennedy, J. A new optimizer using particle swarm theory. In Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, 4–6 October 1995. [CrossRef]

55. Kennedy, J. The particle swarm: Social adaptation of knowledge. In Proceedings of the 1997 IEEE International Conference on Evolutionary Computation (ICEC '97), Indianapolis, IN, USA, 13–16 April 1997; pp. 303–308. [CrossRef]
56. Kennedy, J.; Kennedy, J.F.; Eberhart, R.C. *Swarm Intelligence*; Morgan Kaufmann: Burlington, MA, USA, 2001.
57. Eberhart, R.C.; Shi, Y. Particle swarm optimization: Developments, applications and resources. In Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546), Seoul, Republic of Korea, 27–30 May 2001; Volume 1, pp. 81–86.
58. Robinson, J.; Rahmat-Samii, Y. Particle swarm optimization in electromagnetics. *IEEE Trans. Antennas Propag.* **2004**, *52*, 397–407. [CrossRef]
59. Shi, Y.; Eberhart, R.C. Empirical study of particle swarm optimization. In Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406), Washington, DC, USA, 6–9 July 1999; pp. 1945–1950. [CrossRef]
60. Bergh, F.V.D.; Engelbrecht, A. A Cooperative Approach to Particle Swarm Optimization. *IEEE Trans. Evol. Comput.* **2004**, *8*, 225–239. [CrossRef]
61. Matlab. Available online: www.mathworks.com (accessed on 20 April 2023).
62. Salimi, B.; Mehranzamir, K.; Abdul-Malek, Z. Statistical Analysis of Lightning Electric Field Measured Under Equatorial Region Condition. *Procedia Technol.* **2013**, *11*, 525–531. [CrossRef]
63. Johari, D.; Amir, M.F.A.M.; Hashim, N.; Baharom, R.; Haris, F.A. Positive Cloud-to-Ground Lightning Observed in Shah Alam, Malaysia based on SAFIR 3000 Lightning Location System. In Proceedings of the 2021 IEEE International Conference in Power Engineering Application (ICPEA), Shah Alam, Malaysia, 8–9 March 2021; pp. 178–182. [CrossRef]
64. Zhang, D. Inter-Comparison of Space- and Ground-Based Observations of Lightning. Doctoral Dissertation, University of Arizona, Tucson, AZ, USA, 2019.
65. Poelman, D.R.; Schulz, W.; Pedeboy, S.; Hill, D.; Saba, M.; Hunt, H.; Schwalt, L.; Vergeiner, C.; Mata, C.T.; Schumann, C.; et al. Global ground strike point characteristics in negative downward lightning flashes—Part 1: Observations. *Nat. Hazards Earth Syst. Sci.* **2021**, *21*, 1909–1919. [CrossRef]
66. Poelman, D.R.; Schulz, W.; Pedeboy, S.; Campos, L.Z.S.; Matsui, M.; Hill, D.; Saba, M.; Hunt, H. Global ground strike point characteristics in negative downward lightning flashes—Part 2: Algorithm validation. *Nat. Hazards Earth Syst. Sci.* **2021**, *21*, 1921–1933. [CrossRef]

MDPI

Article

Research on Scale Improvement of Geochemical Exploration Based on Remote Sensing Image Fusion

Haifeng Ding [1,2], Linhai Jing [1,2], Mingjie Xi [3,*], Shi Bai [4], Chunyan Yao [5] and Lu Li [6]

1 International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China
2 Key Laboratory of Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
3 Institute of Geophysical and Geochemical Exploration, Chinese Academy of Geological Sciences, Langfang 065000, China
4 Nanhu Laboratory, Jiaxing 314001, China
5 Nanjing Center of China Geological Survey, Nanjing 210016, China
6 National Geological Library of China, Geoscience Documentation Center of China Geological Survey, Beijing 100083, China
* Correspondence: xmingjie@mail.cgs.gov.cn; Tel.: +86-10-82178079

Abstract: Both remote sensing and geochemical exploration technologies are effective tools for detecting target objects. Although information on anomalous geochemical elemental abundances differs in terms of professional attributes from remote sensing data, both are based on geological bodies or phenomena on the Earth's surface. Therefore, exploring the use of remote sensing data with high spatial resolution to improve the accuracy of small-scale geochemical data, and fusing them to obtain large-scale geochemical layers could provide new data for geological and mineral exploration through inversion. This study provides a method of fusing remote sensing images with small-scale geochemical data based on a linear regression model that improves the resolution of geochemical elemental layers and provides reference data for mineral exploration in areas lacking large-scale geochemical data. In the Xianshuigou area of Northwest China, a fusion study was conducted using 200,000 geochemical and remote sensing data. The method provides fused large-scale regional chemical data in well-exposed areas where large-scale geochemical data are lacking and could provide potential data sources for regional mineral exploration.

Keywords: geochemical exploration; remote sensing; image fusion; mineral exploration

1. Introduction

Geochemical data play one of the most important roles in mineral exploration and environmental pollution studies [1–4]. Geochemical data play a significant role in mineral exploration. They provide information about the chemical composition of rocks, soils, and sediments which can be used to identify mineral deposits that may not be visible on the Earth's surface. Geochemical surveys can help identify areas with anomalous concentrations of metals and minerals, which can be indicative of mineral deposits. These surveys involve collecting and analyzing samples of rocks, soils, and water to determine the presence and concentration of certain elements. Its application in mineral exploration can lead to the discovery of economically viable mineral deposits and can contribute to the growth of the mining industry. The United Nations Educational, Scientific, and Cultural Organization (UNESCO) International Centre on Global-Scale Geochemistry was established in 2016 in China. The global geochemistry network, benchmark chemical elements, and China's geochemistry observation network have been gradually established with the support of this center. Due to the limitations of terrain and the cost of manual acquisition, the sampling points for geochemical elemental abundances are usually insufficient. According to the current situation of geochemical exploration in China, the scale is generally

1:500,000, 1:250,000, or 1:200,000. The sampling distance resolution is at least 1 km × 1 km. The low sampling point density leads to the low resolution of the geochemical elemental abundances, and due to the massive workload of 1:50,000 geochemical explorations, this density is far from sufficient to cover most of China. In many works, only small-scale and low-resolution geochemical data can be used for reference. With existing technology, if the resolution of geochemical elemental layers is to be improved, the sampling density of geochemical elemental abundance sampling points must increase, which will not only greatly increase the cost of manual sampling but also increase the difficulty of sampling due to the complex terrain in some areas.

Based on the spectral reflectance features from the visible (VIS) to visible and near-infrared (VNIR), shortwave infrared (SWIR), and thermal infrared (TIR) wavelengths, many alteration minerals, such as carbonates, sulfates, hydroxides, oxides, quartz, olivine, feldspar, and silicate minerals, have been well identified and mapped [5]. Synthetic aperture radar (SAR) can be used in vegetation cover areas because its microwave radiation penetrates clouds and vegetation [6]. For these features, the VIS, VNIR, SWIR, and TIR images are generally more useful for lithological mapping, alteration information, and mineral deposits. However, radar images are more useful for structural mapping, surface and subsurface geomorphological features, and roughness levels of different rocks.

Both remote sensing and geochemical exploration technologies are very effective methods for detecting target objects [7,8]. Although information on anomalous geochemical elemental abundances and that which is provided by remote sensing data are different in terms of their professional attributes, they are both obtained based on geological bodies or geological phenomena on the Earth's surface, and they are both geological information that include different forms of physical and chemical properties related to mineralization; therefore, these two sources of information must be correlated, which is confirmed by many studies. High anomalies in geochemical elemental abundance data indicate the enrichment of geochemical elements, while remote sensing alteration information indicates the type and intensity of hydrothermal alteration at the surface, which is directly related to the enrichment of elements [9–11].

To solve the problem of large-scale geochemical data using remote sensing images, this paper provides an image fusion method based on a regression model to improve the resolution of geochemical elemental layers by fusing small-scale geochemical data and remote sensing images. The large-scale geochemical data that were obtained from this fusion process may not be comparable to the real geochemical mapping data, but they can at least provide finer and richer information on the details of geochemical anomalies. Thus, the purpose of improving the resolution of geochemical elemental layers without increasing the cost and difficulty of manual sampling is achieved.

2. Geological Setting

The Xianshuigou area (Figure 1), which is located in the Beishan orogenic belt, is part of the Ejina Banner, Inner Mongolia Autonomous Region, China, and is at the intersection of the Kazakhstan, Tarim, and North China plates along the southern margin of the Central Asian Orogenic Belt (CAOB). The orogenic belt is composed of several arc belts that are separated by ophiolite belts, arcs, and blocks formed by subduction–accretion between the Tarim craton and the Kazakhstan plate [12–18]. The Beishan orogenic belt has several complete ophiolites that probably formed during the early Paleozoic in southern China. There are three major ophiolitic belts that are distributed from north to south: the Hongshishan-Baiheshan, Shibanjing-Xiaohuangshan, Yueyashan-Xichangjing, and Liuyuan-Zhangfangshan ophiolitic belts [17,19,20]. According to the spatial and temporal distributions and rock assemblage characteristics of the ophiolite belts, the Xianshuigou area can be divided into the Huaniushan magmatic arc, Gongpoquan-Qiyishan magmatic arc, Heiyingshan magmatic arc, Hanshan microcontinent, and Queershan magmatic arc from south to north [20].
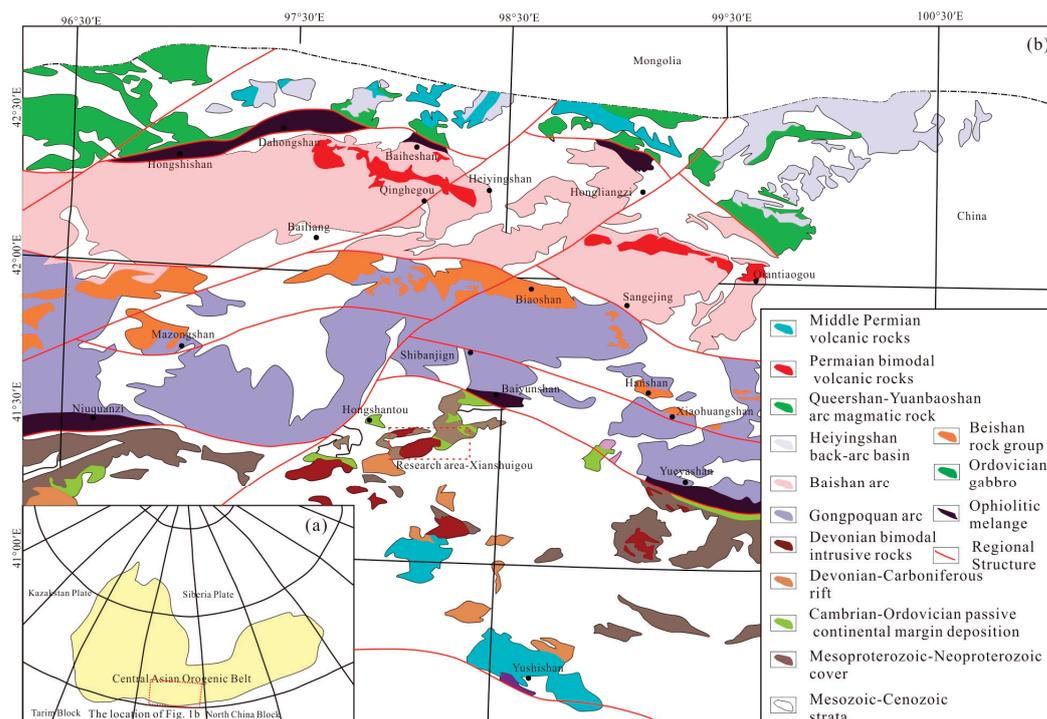
Figure 1. The regional geological map (**a**,**b**).

Paleozoic intrusive rocks are widely developed in the Xianshuigou area and mainly include early Silurian quartz amphibolite, tonalite, granite amphibolite, and diorite formed in an island arc environment; Early Devonian medium-grained porphyritic orthogranite, diorite granite, and quartz diorite formed in a post-collisional orogenic environment; and Late Devonian gabbro, quartz amphibolite, tonalite, and granite amphibolite. Stratigraphically, the Paleozoic Beishan Group is mainly exposed in the northern part of the Shibanjing-Xiaohuangshan tectonic zone; the Ordovician–Silurian Gongpoquan Group, with an island arc volcanic nature, is mainly distributed in the southern part of the tectonic zone; and the Early–Middle Triassic Erduanjing Group and the Early Cretaceous Chijinbao Group are mainly distributed in the northwestern part of the study area and are in angular unconformity contact with the underlying geological bodies.

The study area is also located in the Beishan metallogenic belt, which is an important part of the CAOB metallogenic belt and has an excellent geological background for mineralization. Several mineralization zones are developed in the Beishan metallogenic belt from north to south, namely the Hongshishan-Baiheshan-Pengboshan, Jijitaizi, Hongliuhe-Baiyunshan-Yueyashan, and Huitongshan mineralization zones. In particular, the Beishan metallogenic belt is a favorable area for finding copper polymetallic deposits. A series of copper polymetallic deposits, such as the Sandaominshui, Gaoshishan, Elegenwula, and Liushashan deposits, have been found in the Beishan area of Inner Mongolia. The geochemical investigation results show that the geochemical anomalies of copper-based metallogenic elements in the area are characterized by large scales, high intensities, and obvious zoning. Therefore, there are not only copper polymetallic deposits in the area but also a large amount of valuable information for mineral searches. However, the number of copper polymetallic deposits found in the region is small compared to those in the North Mountain mineralized belt in Xinjiang and Gansu Provinces.

3. Data

3.1. Geochemical Data

The Regional Geochemistry–National Reconnaissance (RGNR) has been carried out since 1978 and covers more than seven million square kilometers of China's territory at scales of 1:500,000, 1:250,000, or 1:200,000. In the program, the work method specification,

sampling methods of stream sediments and rocks, multimethod analytical schemes, and certification of standard reference materials were developed at additional scales (1:100,000, 1:50,000, etc.) for global standard works (http://www.cgs.gov.cn).

In this study, 1:200,000 geochemical data were used as low-scale layers, which compose the database of the China Geological Survey. Geochemical data at a scale of 1:50,000 were used as high-scale layers to verify the fusion results which were collected by the Institute of Geophysical and Geochemical Exploration, Chinese Academy of Geological Sciences, in the Heiyingshan Geological and Mineral Survey Program supported by the China Geological Survey.

All the samples were sieved, and the <80 μm fraction was subjected to inductively coupled plasma–optical emission spectrometry (ICP–OES) for the determination of 44 elements, except Au.

3.2. Remote Sensing Data

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

In this study, the Advanced Spaceborne Thermal Emission and Reflectance Radiometer (ASTER) and Sentinel-2 remote sensing images were used for fusion. ASTER data were downloaded from the United States Geological Survey (USGS) website (https://glovis.usgs.gov/), and Sentinel-2 data were downloaded from the open hub of the European Space Agency (ESA) website (https://scihub.copernicus.eu).

The ASTER on the Terra platform, launched on December 18 1999, provides enhanced mineral mapping capabilities for remote sensing geology because of its additional SWIR band settings [21–25]. The ASTER data consist of three separate subsystems with a total of 14 spectral bands, including three VIS-VNIR bands (0.52–0.86 μm) with a 15 m spatial resolution, six SWIR bands (1.60–2.43 μm) with a 30 m spatial resolution, and five TIR bands (8.12–11.65 μm) with a 90 m spatial resolution [26,27]. There are several levels of ASTER data, and the ASTER Level 1 Precision Terrain Corrected Registered At-Sensor Radiance (AST_L1T) data can be freely downloaded from the USGS Global Visualization Viewer (GloVis), which also contains calibrated at-sensor radiance (https://lpdaac.usgs.gov/). The Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercube (FLAASH) algorithm was used in this study for the atmospheric correction of ASTER data using ENVI 5.3 software.

To address issues related to environmental monitoring, the European Commission (EC) and European Space Agency (ESA) have established the Global Monitoring for Environment and Security (GMES) and the European Earth Observation program Copernicus [28]. The space component of the program consists of satellite missions labeled Sentinel-1 to 5 [29,30]. The Sentinel-2 Multispectral Instrument (MSI) has 13 spectral bands from the VNIR (ten bands) to the SWIR bands (three bands), four bands with 10 m spatial resolution, six bands with 20 m spatial resolution, and three bands with 60 m spatial resolution. Sentinel-2 provides high-resolution and wide-range multispectral data with a revisit time of 5 days [31].

4. Methods

4.1. Remote Sensing

The ASTER scene includes Level-1T format data that have already been used for radiometric calibration, geometric calibration, and orthorectified calibration. The 30 m SWIR bands were resampled to a 15 m pixel size to generate 9-band stacked datasets of the VNIR and SWIR bands. In addition, crosstalk correction was applied to bands 4, 5, and 9 for energy spillover errors [32]. For atmospheric correction, the model fast line-of-sight atmospheric analysis of spectral hypercubes (FLAASH) module was employed to eliminate atmospheric transmittance, solar irradiance topographic effects, and albedo effects by ENVI version 5.3 [33–36]. ASTER data have a rich design of shortwave infrared bands, which can be used for a variety of remotely sensed alteration information extraction. Figure 1 shows

the ASTER band 741 pseudo-color composite map (Figure 2), from which it can be clearly seen that different geological units are displayed as different colors.
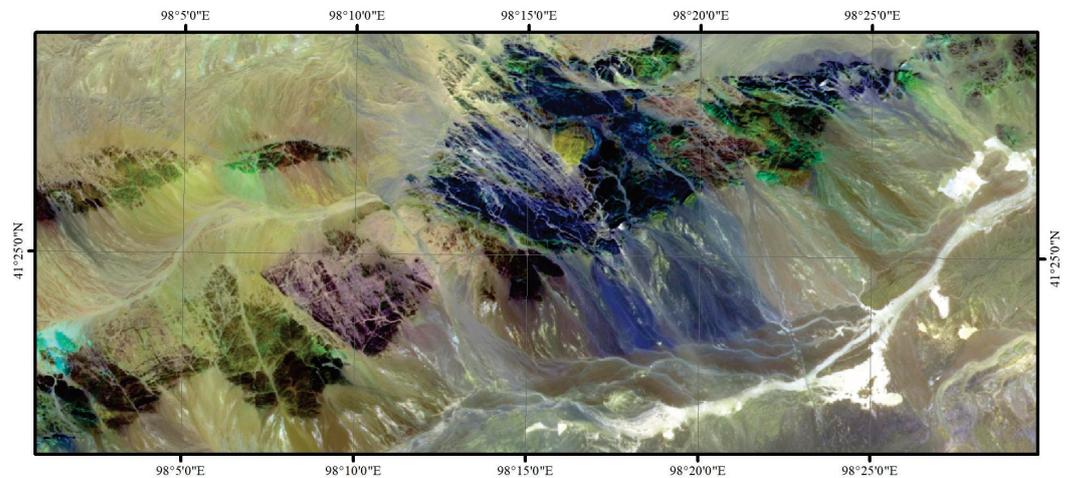
Figure 2. The ASTER image (ASTER red, green, and blue: bands 7, 4, and 1, respectively).

The Level-1C data of Sentinel-2 were freely downloaded from the ESA website, which provides orthorectified top-of-atmosphere (TOA) reflectance with subpixel multispectral registration. The Sen2Cor is a processor for Sentinel-2 Level 2A product generation and formatting, which creates bottom-of-atmosphere, optional terrain- and cirrus-corrected reflectance images, as well as aerosol optical thickness, water vapor, scene classification maps, and quality indicators for cloud and snow probabilities [37]. With a spatial resolution of 10 m, Sentinel-2 has the highest spatial resolution of any data that can be acquired free and unrestricted. Using Sentinel-2 band 842 pseudo-color composite maps (Figure 3), it is possible to see finer delineations and distinctions in the interior of geological bodies that cannot be shown in ASTER.
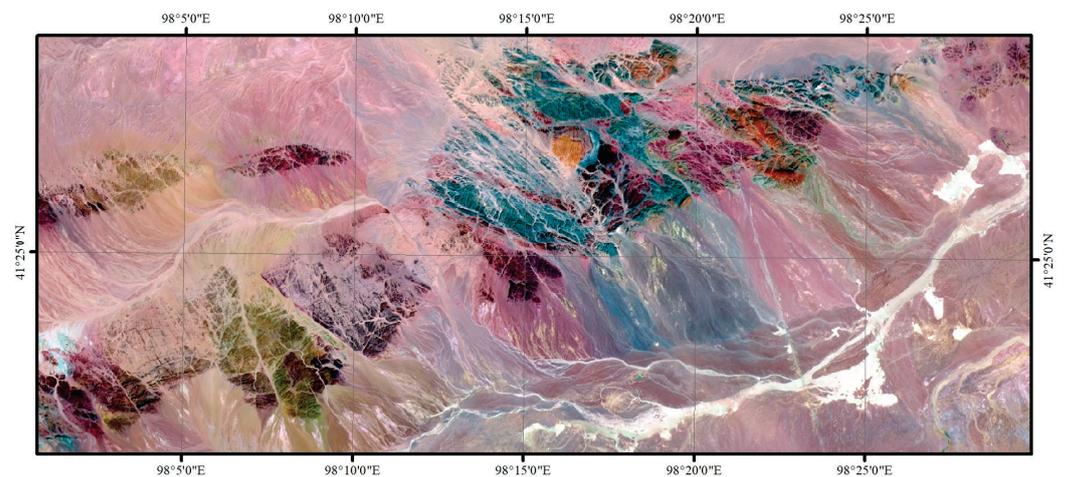
Figure 3. The Sentinel-2 image (Sentinel-2 MSI red, green, and blue: bands 8, 4, and 2, respectively).

4.2. Geochemical Data

The fusion geochemical data used to extract geochemical anomalies in this study are geochemical survey data at a scale of 1:200,000. A scale of 1:50,000 for geochemical data was used to verify the accuracy of the fusion results. The samples were 0–20 cm weathered surfaces that overlie the stratigraphic units and exposed range of intrusive rocks. These samples were collected within 100 m/25 m of the sampling point, with one sample consisting of three to five points, which were then naturally air-dried and screened indoors for measurement. The distribution of sampling points is shown in Figures 4 and 5.

According to the major deposit types, we selected Cu for data processing and analysis. The spatial distribution of Cu with the Kriging interpolation method is shown in Figure 6 (200,000) and Figure 7 (50,000). The 50,000 geochemical sampling was divided into debris and soil. For better integration studies, only the debris sampling area on the west side of the study area was considered when referring to the 50,000 geochemical data.
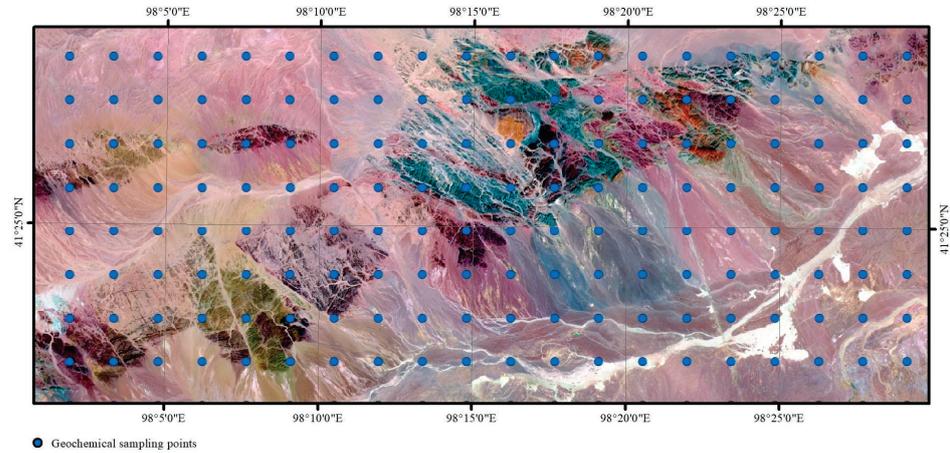
Figure 4. The 200,000 geochemical sampling location map (Sentinel-2 MSI red, green, and blue: bands 8, 4, and 2, respectively).
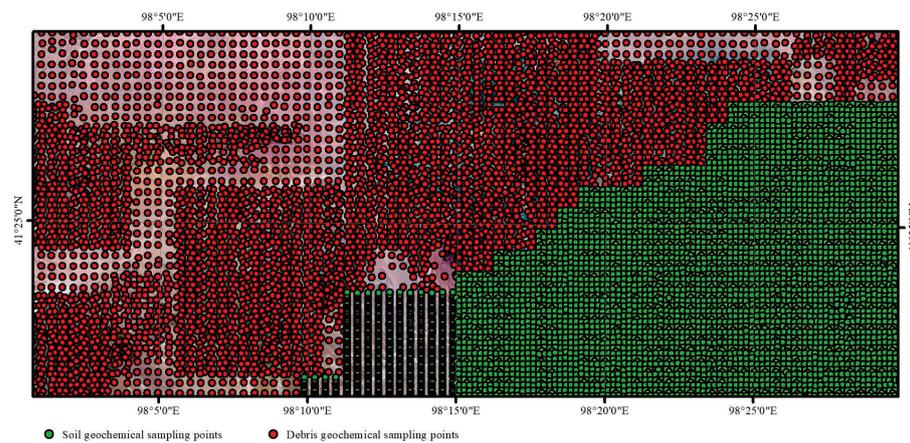
Figure 5. The 50,000 geochemical sampling location map (Sentinel-2 MSI red, green, and blue: bands 8, 4, and 2, respectively).
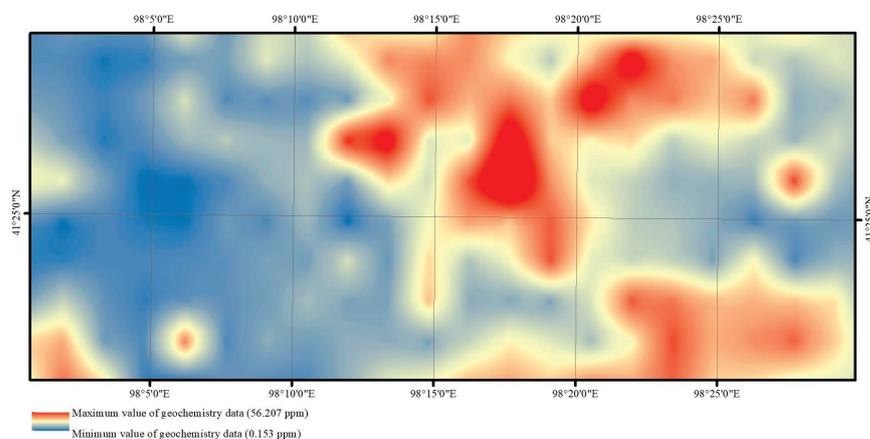
Figure 6. Map showing the spatial distribution of Cu modeled using the Kriging interpolation method (200,000).
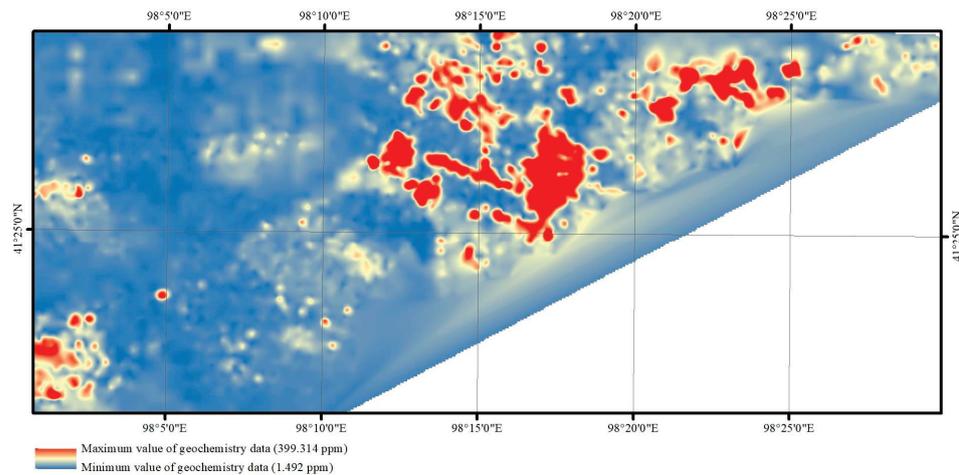
Figure 7. Map showing the spatial distribution of debris Cu modeled using the Kriging interpolation method (50,000).

4.3. Regression Method

As mentioned previously, remote sensing data were obtained by detecting the reflection of objects on the Earth's surface from solar radiation. These data can provide information on lithology, geological structure, and alteration through various wave bands, with different remote sensing sources offering complementary advantages and correlations between features. Remote sensing images offer a wide coverage area, intuitive visuals, and rich comprehensive data. These images allow for direct extraction of geological features, including strata, lithology, structure, and quaternary distribution, as well as bare rocks, topographic and geomorphic features, and water system distribution. However, geochemical data represent the spatial distribution of chemical elements, with geochemical indexes indicating the main chemical elements present on the Earth. Geochemical exploration methods rely on chemical elemental anomaly information as a crucial indicator for direct prospecting.

In this study, the method used for image fusion is pixel-level fusion. This method directly processes the image to obtain detailed information by fusing the original layers. Pixel-level fusion can be divided into two categories, including spatial and transform domain-based fusion methods. Spatial domain-based fusion directly uses pixel values for fusion, while the transform domain-based fusion transforms the image into coefficients and combines the obtained individual coefficients to achieve fusion. The fusion method used in this paper is based on a linear regression model, which incorporates the spectral and spatial characteristics of the remote sensing and geochemical data and is a spatial domain-based fusion method. The proposed method produces fused geochemical layers with higher spatial resolution, which can potentially provide a comprehensive understanding of the distribution of geochemical elements in geographic areas, making it a useful tool for geological and mineral exploration.

To improve the resolution of the geochemical elemental layer, the method used in this study involves several steps. Firstly, the geochemical elemental layer and corresponding areas of remote sensing images are obtained and assessed. Secondly, preprocessing of the remote sensing image is done, followed by clustering and alteration information acquisition. Thirdly, alteration information is acquired for image decomposition into low-frequency and high-frequency images. Fourthly, within the scope of ground objects in each cluster, the pending geochemical element layer and the first graphic first function relation are obtained. Finally, the first function relation and the secondary graphic spatial details and the pending geochemical elemental layer are used to obtain a high-resolution layer. The purpose is to achieve artificial sampling without any increase in the cost and difficulty of sampling, based on the improved resolution of the geochemical element layer.

According to the above characteristics, this paper selects logistic regression methods for fusion research, including global and local logistic regression methods. The overall technical process is shown in Figure 8.
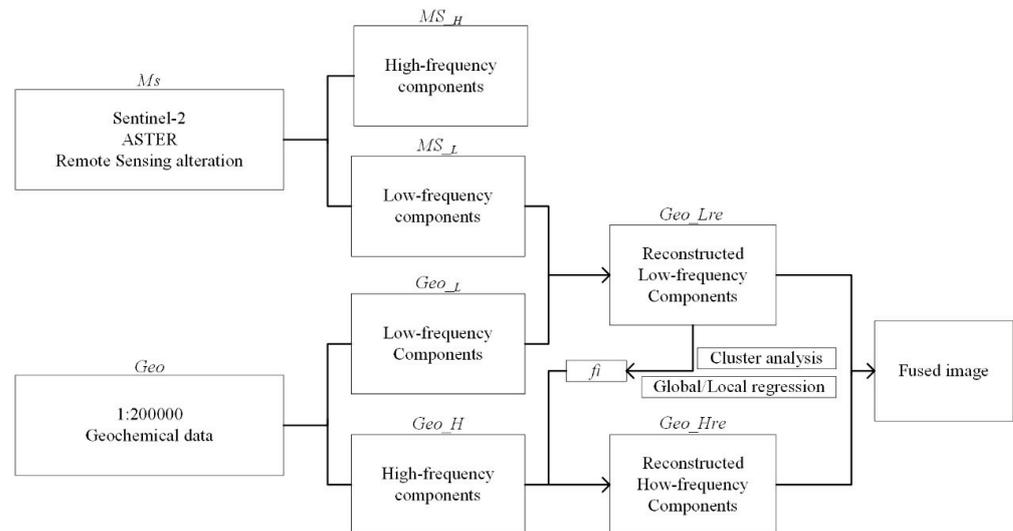
Figure 8. Methodological flowchart.

1. Image decomposition

The remote sensing images and alteration information images (MS) are used for data fusion. The image Laplacian pyramid algorithm is used to decompose the remote sensing images and alteration information images into a low-frequency image (MS_L) and a high-frequency image (MS_H) with respect to the spatial details and background, respectively.

2. Geochemistry data scaling transformation

As the geochemical elemental layers (Geo) are the result of interpolated data at interval sampling points, the geochemical element layers need to be resampled by cubic convolution to the same resolution as the remote sensing or alteration information images. The resampled geochemical data are similarly decomposed into two parts: a low-frequency image (Geo_L) and a high-frequency image (Geo_H).

3. Cluster analysis

Cluster analysis is a process that divides a concrete or abstract data set into several groups or classes. This technique is widely used in data mining, image segmentation, pattern recognition, spatial remote sensing technology, feature extraction, signal compression, and many other fields and has obtained many satisfactory results [38,39]. Cluster analysis identifies clustering structures in a dataset that are characterized by the maximum similarity within the same cluster and the maximum dissimilarity between different clusters [40,41]. In this case, the similarity between similar datasets is as large as possible, and the similarity between dissimilar data sets is as small as possible. Cluster analysis is required for local regression fusion. In this paper, a K-means clustering algorithm is used for the unsupervised classification of all the geological bodies contained in remote sensing images and alteration information images. N classes of clustering centers, each of which represents a feature range of a geological body, where $N \geq 2$, are obtained.

The K-means algorithm is generally the most well-known and widely used clustering method. Various extensions of K-means have been proposed in the literature [38]. The K-means clustering algorithm is a hard clustering algorithm that is representative of the prototype-based clustering method of objective functions. The algorithm involves the distance between data points and uses the prototype as the optimized objective function. The optimization rules of iterative operation can be obtained by using the method of finding

the extremum of the function. The K-means clustering algorithm takes the Euclidean distance as a similarity measure to find the optimal classification of the center vector V of an initial cluster to minimize the evaluation index J. The algorithm uses the error sum of squares criterion function as the clustering criterion function [42,43].

The process of the K-means clustering algorithm is as follows. First, the initial clustering center is determined, i.e., K points are randomly selected from the dataset. Second, the distance from each sample point to the selected clustering center is calculated according to a distance formula, and according to the minimum distance criterion, each sample point is classified into the class represented by the clustering center with the shortest distance from the point. Then, the average value of the objects in each class is calculated and compared with the last calculation. If there is no change in the clustering center between these two calculations or the change does not exceed the given threshold, it means that the clustering criterion function converges and the clustering algorithm ends. Otherwise, we continue to adjust the clustering centers and restart the next iteration. In one iteration of the algorithm, the clustering center does not change, and the class to which each sample point belongs no longer changes, which means that all samples are correctly classified and the algorithm ends [44,45].

4. Establishment of the correlation function

Third, a multiple linear regression function is established independently for the global region or every cluster by considering the linear relationship between MS_L and Geo_L. According to the correlation between the concentrations of Geo_L and MS_L, reconstructed fused low-frequency geochemical data (Geo_Lre) are obtained.

The expression of the linear function relation is as follows:

$$Geo\_Lre = f_i(MS\_L_1, MS\_L_2, \ldots\ldots, MS\_L_n), \tag{1}$$

where i is the total number of bands in the MS image or alteration information image and n (n = 1, 2, . . . , i) is the band number. The correlation coefficients and constants of the function can be calculated using the least squares approach.

According to the linear function, the spatial details of MS_H are injected into Geo_H to be processed within the feature range of each geological body to obtain the target geochemical elemental layer. The correlation coefficients and constants of the linear function are used to obtain the reconstructed fused high-frequency geochemical data (Geo_Hre) as follows:

$$Geo\_Hre = Geo\_H + f_i(MS\_H_1, MS\_H_2, \ldots\ldots, MS\_H_n), \tag{2}$$

On the basis of the above example, in the application of an implemented case as shown in Figures 2 and 3, each category of different geological body units within the scope of the first function is described based on the relationship. The categories are described in the second image space based on the details mentioned above, including the geochemical elemental layer, to achieve the target geochemical elements. It is a linear model within different geological body units, and the detailed information of high frequency images is fully considered to further improve the accuracy of subsequent fusion.

5. Image fusion

The high-resolution fusion image Geo_f can be obtained by merging the resampled geochemical elemental layer Geo_Lre with the fused high-frequency image Geo_Lre, as follows:

$$Geo\_f = Geo\_Lre + Geo\_Hre, \tag{3}$$

4.4. Image Fusion Quality Evaluation

The evaluation of the quality of remote sensing image fusion results includes two main aspects: subjective evaluation and objective evaluation. The subjective evaluation method involves researchers directly observing the fused images with the naked eye and assessing the quality of the fusion based on their subjective experience. In this study, researchers can assess the quality of the fusion by examining whether mineralized points are found during

field surveys. The objective evaluation method involves using quantitative formulas such as information entropy, spectral information, and correlation coefficients to evaluate the fused image information. In this study, the evaluated 50,000 geochemical mapping data points serve as the basis for both the subjective and objective evaluation of the quality of image fusion.

5. Research and Discussion

Sentinel-2 data are used as the original images for fusion because of the high signal-to-noise ratio among the free remote sensing data. ASTER data are used as another remote sensing data source for fusion because of their rich shortwave infrared band settings that can provide remote sensing alteration information. The ASTER Mineral Index Processing Manual by Aleks Kalinowski and Simon Oliver (http://www.ga.gov.au/image_cache/GA7 833.pdf, accessed on 26 February 2023) provides a comprehensive overview of the mineral indices derived from ASTER data. These authors suggest a range of band combinations and ratios for mapping various mineral assemblages in relation to different styles of alteration (Table 1).

Table 1. ASTER band ratios for enhancing mineral features.

Mineral Feature	ASTER Band Combination(s)
Ferric iron	2/1
Ferrous iron	5/3 and 1/2
Ferric oxide	4/3
Gossan	4/2
Carbonate/chlorite/epidote	$(7 + 9)/8$
Epidote/Chlorite/Amphibole	$(6 + 9)/(7 + 8)$
Amphibole	$(6 + 9)/8$ and 6/8
Dolomite	$(6 + 8)/7$
Carbonate	13/14
Sericite/Muscovite/Illite/Smectite	$(5 + 7)/6$
Alunite/Kaolinite/Pyrophyllite	$(4 + 6)/5$
Phengite	5/6
Kaolinite	7/5
Silica	11/10, 11/12, 13/10
SiO_2	13/12, 12/13
Siliceous rocks	$(11 \times 11)/(10 \times 12)$

Several band ratios have also been proposed to map mineral indices (Table 2) [21,46].

Table 2. ASTER false color composites for enhancing mineral features.

Mineral Feature	ASTER Band Combination(s)
Silica index	band 11/band 10, band 11/band 12, band 13/band 10
Biotite-epidote–chlorite–amphibole index	(band 6 + band 9)/(band 7 + band 8)
Skarn carbonates–epidote index	(band 6 + band 9)/(band 7 + band 8), band 13/band 14
Garnets-pyroxenes index	band 12/band 13
Iron oxide index	band 2/band 1
White micas Al-OH depth	(band 5 + band 7)/band 6
Carbonates Mg-OH depth	(band 6 + band 9)/(band 7 + band 8)
Carbonate abundance	band13/band14

To better obtain the fusion relationship between alteration information and geochemistry, a principal component transformation was performed on all the extracted alteration information that was obtained from ASTER data. This was done because ASTER data provide diverse alteration information. PC2 was selected from the principal component analysis as it concentrates the information of the maximum relevant variability outside the

two factors and more reflects the alteration information. The same approach was applied to the Sentinel-2 data which has rich multispectral bands. In the principal component analysis, PC1 reflects the weighted information of all the bands, which includes the background values and topographic factors. Therefore, PC2 was chosen as the data source for fusion because it concentrates on the most relevant information regarding alteration information and geochemistry.

Figures 9–12 demonstrate the results of Sentinel-2 data PC2 global regression fusion, Sentinel-2 data PC2 local regression fusion, ASTER alteration information PC2 global regression fusion, and ASTER alteration information PC2 local regression fusion. Overall, global regression is better at providing the range of anomalies in small-scale geochemical data, while local regression cannot show relevant information. The global regression fusion results give a more precise prediction of a relatively concentrated and small number of new anomalies with a wider area of individual anomalies. Contrarily, the local regression fusion results provide a larger and more dispersed number of new anomalies with a smaller area of individual anomalies, giving more detailed information on new anomalies for mineral exploration prediction. Visually inspecting the results, both features have their advantages. The global regression fusion results display new anomalies within the original small-scale geochemical anomalies, whereas the local fusion results provide richer and finer information on new anomalies.
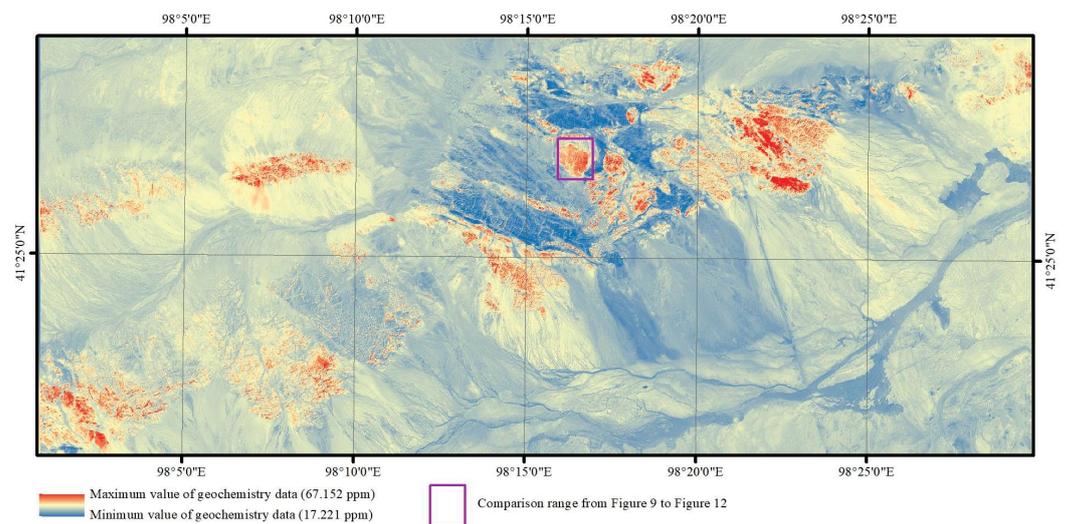
Figure 9. Fusion image of Sentinel-2 data PC2 local regression.
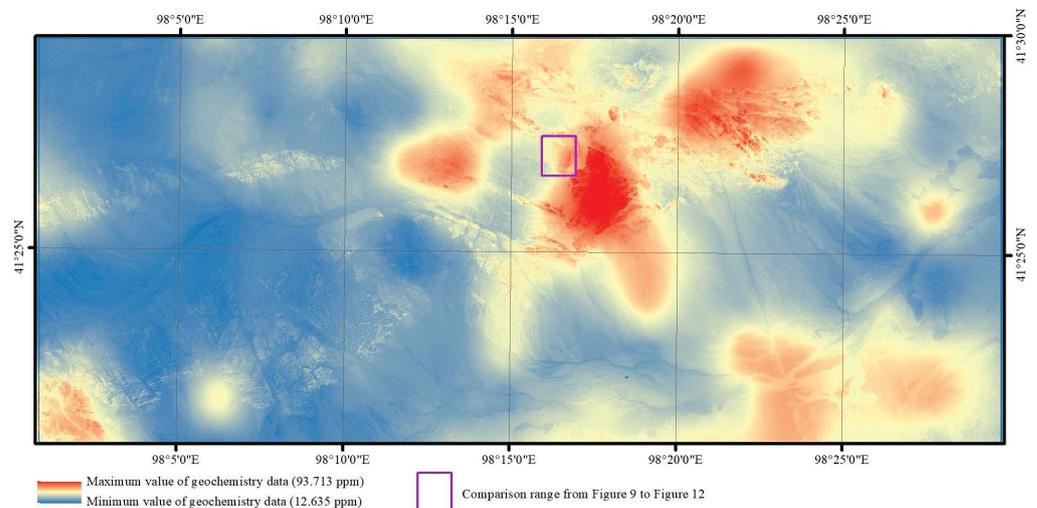
Figure 10. Fusion image of Sentinel-2 data PC2 global regression.
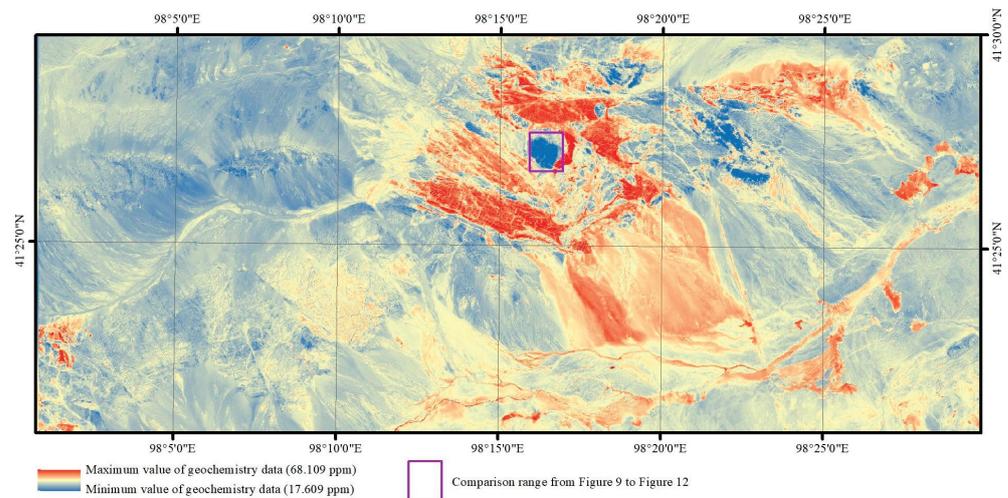
Figure 11. Fusion image of ASTER alteration information PC2 local regression.
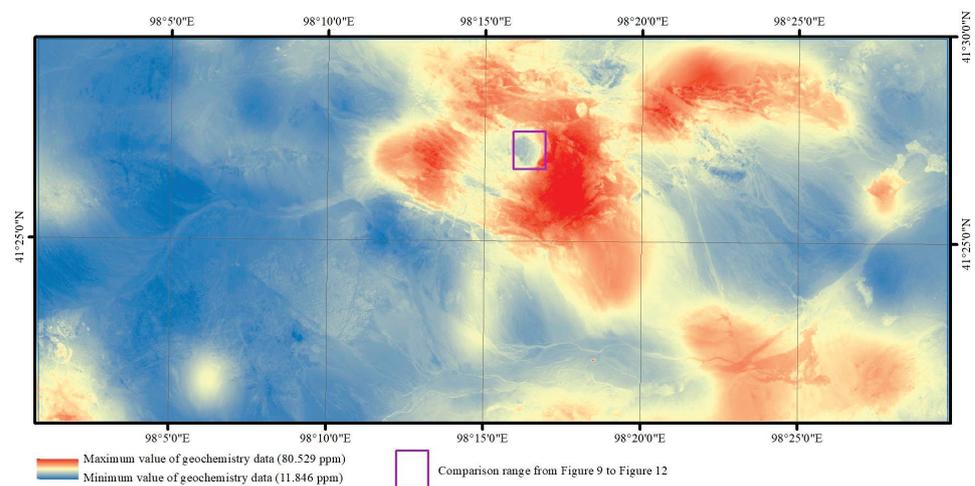
Figure 12. Fusion image of ASTER alteration information PC2 global regression.

A comparison of the fusion results from the original image and the alteration information, the fusion results of band PC2 of the original image and band PC2 of the alteration information reveals a certain negative correlation between the two fusion results, such as the areas shown in the purple boxes in Figures 9–12. Compared with the 1:50,000-scale geochemical data, the fusion results of the original image band PC2 show a somewhat negative correlation, while the fusion results of alteration information PC2 show a somewhat positive correlation. The results of the original image fusion are shown as medium-high values in yellow and red colors, while the results of the alternation fusion are shown as low values in blue color.

Magnifying the distribution of the largest alteration anomalies in the middle position of the 1:200,000-scale geochemical data layer can more clearly yield the above-described conclusions (Figure 13a–c). Local regression fusion can provide a range of anomalies in small-scale geochemical data, and it may be more effective to screen new anomalies in the range of 1:200,000-scale geochemical anomalies. However, we also see that the 1:200,000-scale and 1:50,000-scale geochemical anomalies are not completely consistent, and possible new anomalies that are found only in the range of 1:200,000-scale geochemical anomalies may also lead to the omission of valid information. The local images also show that the results that were obtained from the PC2 fusion of alteration information are more scattered, with more anomalies and smaller areas. In contrast, the fusion results obtained from the original image PC2 are more concentrated and have fewer anomalies and larger areas. The positive and negative correlations are also consistent with the previous analysis;

i.e., the fusion results of the original image band PC2 show a somewhat negative correlation (Figure 13c,d), while the fusion results of the PC2 etched information PC2 show a somewhat positive correlation (Figure 13e,f).
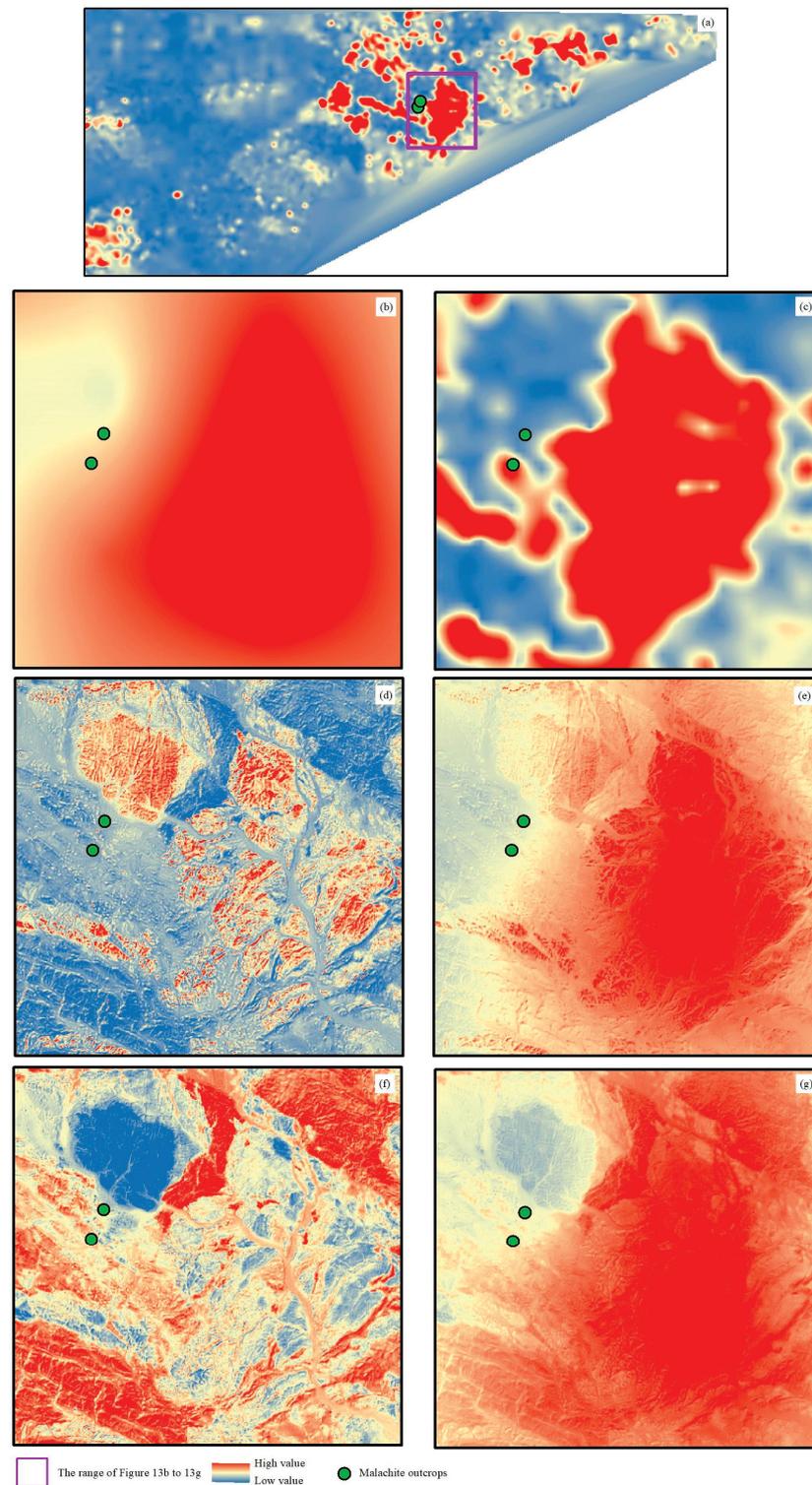
Figure 13. Local comparison image of: (**a**) the comparison range of (**a**–**g**); (**b**) 200,000-scale geochemistry; (**c**) 50,000-scale geochemistry; (**d**) Sentinel-2 data PC2 local regression fusion; (**e**) Sentinel-2 data PC2 global regression fusion; (**f**) ASTER alteration information PC2 local regression fusion; and (**g**) ASTER alteration information PC2 global regression fusion.

To verify the quantitative analysis, see Table 3, which contains the correlation between the original image fusion results, the fusion results of the alteration information, and the 1:200,000-scale geochemical data. Due to the complexity of geochemical data filling, the correlation between the 1:50,000- and 1:200,000-scale geochemical data also did not show an obvious positive correlation, where the correlation coefficient is −0.3. The highest correlation among the fusion results is the regression fusion result of local alteration information, which is higher than the correlation between the 1:50,000- and 1:200,000-scale geochemical data. The correlation results also show that original image band PC2 shows a positive correlation with the 1:200,000-scale geochemical data, while the PC2 alteration information shows a negative correlation with the 1:200,000-scale geochemical data. The negative correlation of the fusion results of the alteration information may be related to the fact that too many types of alteration information were extracted, the geological background corresponding to each type of alteration information was not analyzed, and the principal component analysis was performed directly. The subsequent detailed analysis and fusion of each type of alteration information is needed to find the optimal combination. In fact, a negative correlation is also shown between the 1:200,000 and 1:50,000 data. This indicates that geochemical anomalies are a very complex correspondence that is influenced by a variety of factors such as sampling site location, density, and sampling method.

Table 3. Correlation coefficient of different fusion methods.

	Sentinel-2 Bands PC2	**ASTER Alteration Information PC2**
Global	0.12	−0.17
Local	0.27	−0.32

In the field survey, which is the most effective means to verify the indoor interpretation results, we found several malachite outcrops (Figure 14). In the 1:200,000 geochemical layer, malachite outcrops are located at the edge of the geochemical high value (Figure 13a), while in the 1:50,000 geochemical layer, one malachite outcrop is located in the geochemical high value area and another malachite outcrop is located in the geochemical low value area (Figure 13b). In the fused image, malachite outcrops are in the high value area in the ASTER alteration information fusion results (Figure 13c,d), while they are in the low value area or the edge of high value area in the original image fusion results (Figure 13e,f). From the field survey results, the fused results, especially the large-scale geochemical layers obtained by fusing the alteration information, can provide some high value information that cannot be shown in the small-scale geochemical layers.

(**a**) (**b**)

Figure 14. Photographs of malachite outcrops and hand specimens: (**a**) malachite outcrops; (**b**) malachite hand specimens.

6. Conclusions

In this paper, we establish a linear regression equation for data fusion between remote sensing images and geochemical data. Remote sensing raw images and remote sensing alteration information are fused with small-scale geochemistry to obtain fused large-scale geochemical inversion data, which provide a source of data analysis for mineral exploration

in areas lacking large-scale geochemical data. The regression equation is established using the low-frequency images obtained from image decomposition; then, the detailed spatial information on the high-frequency images is injected for fusion and the obtained fusion results achieve better validation results in quantitative correlation coefficients, visual observation, and field sampling. The fusion results of Sentinel-2 raw images show positive correlations, while the fusion results of ASTER alteration information show negative correlations. Global regression fusion can provide a range of anomalies in small-scale geochemical data, and it may be more effective to screen new anomalies in the range of 200,000 geochemical anomalies. This equation provides a new idea for data source acquisition for regional geological and mineral exploration in areas lacking large-scale geochemical data. More fusion models should be used in the future to explore the most effective fusion method. This method will be especially beneficial to mineral exploration work in high-altitude uninhabited areas, such as Xinjiang and Tibet in western China.

References

1. Plant, J.; Smith, D.; Smith, B.; Williams, L. Environmental geochemistry at the global scale. *Appl. Geochem.* **2001**, *16*, 1291–1308. [CrossRef]
2. Xie, X.; Wang, X.; Zhang, Q.; Zhou, G.; Cheng, H.; Liu, D.; Cheng, Z.; Xu, S. Multi-scale geochemical mapping in China. *Geochem. Explor. Environ. Anal.* **2008**, *8*, 333–341. [CrossRef]
3. Zuo, R. Identifying geochemical anomalies associated with Cu and Pb–Zn skarn mineralization using principal component analysis and spectrum–area fractal modeling in the Gangdese Belt, Tibet (China). *J. Geochem. Explor.* **2011**, *111*, 13–22. [CrossRef]
4. Li, M.; Xi, X.; Xiao, G.; Cheng, H.; Yang, Z.; Zhou, G.; Ye, J.; Li, Z. National multi-purpose regional geochemical survey in China. *J. Geochem. Explor.* **2014**, *139*, 21–30. [CrossRef]
5. Wei, J.L.; Liu, X.N.; Ding, C.; Liu, M.L.; Jin, M.; Li, D.D. Developing a thermal characteristic index for lithology identification using thermal infrared remote sensing data. *Adv. Space Res.* **2017**, *59*, 74–87. [CrossRef]
6. Radford, D.D.G.; Cracknell, M.J.; Roach, M.J.; Cumming, G.V. Geological Mapping in Western Tasmania Using Radar and Random Forests. *IEEE J.-Stars* **2018**, *11*, 3075–3087. [CrossRef]
7. Wang, W.; Cheng, Q. Mapping Mineral Potential by Combining Multi-Scale and Multi-Source Geo-Information. In Proceedings of the IGARSS 2008—2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, 7–11 July 2008; pp. II-1321–II-1324.
8. He, H.; Yang, X.; Li, Y.; Li, Y.; Wang, L. Multi-source data fusion technique and its application in geological and mineral survey. *J. Earth Sci. Environ.* **2010**, *32*, 44–47.
9. Wang, Z.; Zuo, R.; Jing, L. Fusion of Geochemical and Remote-Sensing Data for Lithological Mapping Using Random Forest Metric Learning. *Math. Geosci.* **2021**, *53*, 1125–1145. [CrossRef]
10. Eldosouky, A.M.; Abdelkareem, M.; Elkhateeb, S.O. Integration of remote sensing and aeromagnetic data for mapping structural features and hydrothermal alteration zones in Wadi Allaqi area, South Eastern Desert of Egypt. *J. Afr. Earth Sci.* **2017**, *130*, 28–37. [CrossRef]
11. Kuhn, S.; Cracknell, M.J.; Reading, A.M. Lithologic mapping using Random Forests applied to geophysical and remote-sensing data: A demonstration study from the Eastern Goldfields of AustraliaLithologic mapping with RF. *Geophysics* **2018**, *83*, B183–B193. [CrossRef]
12. Windley, B.F.; Alexeiev, D.; Xiao, W.; Kröner, A.; Badarch, G. Tectonic models for accretion of the Central Asian Orogenic Belt. *J. Geol. Soc.* **2007**, *164*, 31–47. [CrossRef]
13. Fu, Y.G.; Cai, Y.W.; Ma, S.M.; Xi, M.J.; Gao, J.W.; Wang, Z.L. Discovery of similar to 338 Ma Shibanjing Plagiogranite in the Beishan Area, NW China, and its Geological Significance. *Acta Geol. Sin.-Engl. Ed.* **2020**, *94*, 577–579. [CrossRef]
14. Kroner, A.; Kovach, V.; Belousova, E.; Hegner, E.; Armstrong, R.; Dolgopolova, A.; Seltmann, R.; Alexeiev, D.V.; Hoffmann, J.E.; Wong, J.; et al. Reassessment of continental growth during the accretionary history of the Central Asian Orogenic Belt. *Gondwana Res.* **2014**, *25*, 103–125. [CrossRef]

15. Zheng, R.G.; Wu, T.R.; Zhang, W.; Xu, C.; Meng, Q.P. Late Paleozoic subduction system in the southern Central Asian Orogenic Belt: Evidences from geochronology and geochemistry of the Xiaohuangshan ophiolite in the Beishan orogenic belt. *J. Asian Earth Sci.* **2013**, *62*, 463–475. [CrossRef]

16. Niu, Y.Z.; Shi, G.R.; Ji, W.H.; Zhou, J.L.; Wang, J.Q.; Wang, K.; Bai, J.K.; Yang, B. Paleogeographic evolution of a Carboniferous-Permian sea in the southernmost part of the Central Asian Orogenic Belt, NW China: Evidence from microfacies, provenance and paleobiogeography. *Earth-Sci. Rev.* **2021**, *220*, 103738. [CrossRef]

17. Xiao, W.J.; Mao, Q.G.; Windley, B.F.; Han, C.M.; Qu, J.F.; Zhang, J.E.; Ao, S.J.; Guo, Q.Q.; Cleven, N.R.; Lin, S.F.; et al. Paleozoic multiple accretionary and collisional processes of the Beishan orogenic collage. *Am. J. Sci.* **2010**, *310*, 1553–1594. [CrossRef]

18. Li, M.; Xin, H.T.; Ren, B.F.; Ren, Y.W.; Liu, W.G. Early-Middle Permian post-collisional granitoids in the northern Beishan orogen, northwestern China: Evidence from U-Pb ages and Sr-Nd-Hf isotopes. *Can. J. Earth Sci.* **2020**, *57*, 681–697. [CrossRef]

19. Zheng, R.; Li, J.; Xiao, W. Mid-Paleozoic ridge subduction in the Central Beishan of the southern Altaids: Evidence from geochemical, Sr–Nd and zircon U–Pb–Hf–O isotopic data of Gongpoquan volcanic rocks. *J. Geol. Soc.* **2019**, *176*, 755–770. [CrossRef]

20. Li, M.; Xin, H.; Tian, J.; Meng, X.; Pang, Z.; Chen, C.; Liang, G. Composition, Age and Polarity of Gongpoquan Arc and Its Tectonic Significance in Beishan Orogen. *Earth Sci.* **2020**, *45*, 2393–2412.

21. van der Meer, F.D.; van der Werff, H.M.A.; van Ruitenbeek, F.J.A.; Hecker, C.A.; Bakker, W.H.; Noomen, M.F.; van der Meijde, M.; Carranza, E.J.M.; de Smeth, J.B.; Woldai, T. Multi- and hyperspectral geologic remote sensing: A review. *Int. J. Appl. Earth Obs.* **2012**, *14*, 112–128. [CrossRef]

22. Abrams, M.; Tsu, H.; Hulley, G.; Iwao, K.; Pieri, D.; Cudahy, T.; Kargel, J. The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) after fifteen years: Review of global products. *Int. J. Appl. Earth Obs.* **2015**, *38*, 292–301. [CrossRef]

23. Ge, W.Y.; Cheng, Q.M.; Jing, L.H.; Armenakis, C.; Ding, H.F. Lithological discrimination using ASTER and Sentinel-2A in the Shibanjing ophiolite complex of Beishan orogenic in Inner Mongolia, China. *Adv. Space Res.* **2018**, *62*, 1702–1716. [CrossRef]

24. Rowan, L.C.; Mars, J.C.; Simpson, C.J. Lithologic mapping of the Mordor, NT, Australia ultramafic complex by using the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). *Remote Sens. Environ.* **2005**, *99*, 105–126. [CrossRef]

25. Yamaguchi, Y.; Kahle, A.B.; Tsu, H.; Kawakami, T.; Pniel, M. Overview of Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). *IEEE T. Geosci. Remote* **1998**, *36*, 1062–1071. [CrossRef]

26. Liu, L.; Li, Y.; Zhou, J.; Han, L.; Xu, X.L. Gold-copper deposits in Wushitala, Southern Tianshan, Northwest China: Application of ASTER data for mineral exploration. *Geol. J.* **2018**, *53*, 362–371. [CrossRef]

27. Rowan, L.C.; Mars, J.C. Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. *Remote Sens. Environ.* **2003**, *84*, 350–366. [CrossRef]

28. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]

29. van der Werff, H.; van der Meer, F. Sentinel-2 for Mapping Iron Absorption Feature Parameters. *Remote Sens.* **2015**, *7*, 12635–12653. [CrossRef]

30. van der Meer, F.D.; van der Werff, H.M.A.; van Ruitenbeek, F.J.A. Potential of ESA's Sentinel-2 for geological applications. *Remote Sens. Environ.* **2014**, *148*, 124–133. [CrossRef]

31. Yokoya, N.; Chan, J.C.W.; Segl, K. Potential of Resolution-Enhanced Hyperspectral Data for Mineral Mapping Using Simulated EnMAP and Sentinel-2 Images. *Remote Sens.* **2016**, *8*, 172. [CrossRef]

32. Iwasaki, A.; Tonooka, H. Validation of a crosstalk correction algorithm for ASTER/SWIR. *IEEE T. Geosci. Remote* **2005**, *43*, 2747–2751. [CrossRef]

33. Module, F. Atmospheric Correction Module: QUAC and FLAASH User's Guide. Version 4, p. 44. 2009. Available online: https://scirp.org/reference/referencespapers.aspx?referenceid=2208171 (accessed on 23 February 2023).

34. Adler-Golden, S.; Berk, A.; Bernstein, L.; Richtsmeier, S.; Acharya, P.; Matthew, M.; Anderson, G.; Allred, C.; Jeong, L.; Chetwynd, J. FLAASH, a MODTRAN4 atmospheric correction package for hyperspectral data retrievals and simulations. In Proceedings of the 7th Annual JPL Airborne Earth Science Workshop, Pasadena, CA, USA, 12 December 1998; pp. 9–14.

35. Testa, F.J.; Villanueva, C.; Cooke, D.R.; Zhang, L. Lithological and Hydrothermal Alteration Mapping of Epithermal, Porphyry and Tourmaline Breccia Districts in the Argentine Andes Using ASTER Imagery. *Remote Sens.* **2018**, *10*, 203. [CrossRef]

36. Eldosouky, A.M.; El-Qassas, R.A.Y.; Pour, A.B.; Mohamed, H.; Sekandari, M. Integration of ASTER satellite imagery and 3D inversion of aeromagnetic data for deep mineral exploration. *Adv. Space Res.* **2021**, *68*, 3641–3662. [CrossRef]

37. De Keukelaere, L.; Sterckx, S.; Adriaensen, S.; Knaeps, E.; Reusen, I.; Giardino, C.; Bresciani, M.; Hunter, P.; Neil, C.; Van der Zande, D.; et al. Atmospheric correction of Landsat-8/OLI and Sentinel-2/MSI data using iCOR algorithm: Validation for coastal and inland waters. *Eur. J. Remote Sens.* **2018**, *51*, 525–542. [CrossRef]

38. Sinaga, K.P.; Yang, M.S. Unsupervised K-Means Clustering Algorithm. *IEEE Access* **2020**, *8*, 80716–80727. [CrossRef]

39. Liu, Y.; Yin, H.P.; Chai, Y. An Improved Kernel K-means Clustering Algorithm. pp. 275–280. 2016. Available online: https://link.springer.com/chapter/10.1007/978-981-10-2338-5_27 (accessed on 23 February 2023).

40. Jain, A.K. *Algorithms for Clustering Data*; Prentice-Hall, Inc.: Hoboken, NJ, USA, 1988.

41. Gentle, J.E. Finding Groups in Data: An Introduction to Cluster Analysis. *Biometrics* **1991**, *47*, 788. [CrossRef]

42. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 1–22. [CrossRef]
43. Yu, J.; Chaomurilige, C.; Yang, M.-S. On convergence and parameter selection of the EM and DA-EM algorithms for Gaussian mixtures. *Pattern Recognit.* **2018**, *77*, 188–203. [CrossRef]
44. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Los Angeles, CA, USA, 21 June–18 July 1965; pp. 281–297.
45. Alhawarat, M.; Hegazi, M. Revisiting K-means and topic modeling, a comparison study to cluster arabic documents. *IEEE Access* **2018**, *6*, 42740–42749. [CrossRef]
46. Cudahy, T.; Hewson, R. ASTER geological case histories: Porphyry-skarnepithermal, iron oxide Cu-Au and Broken hill Pb-Zn-Ag. In Proceedings of the Annual General Meeting of the Geological Remote Sensing Group 'ASTER Unveiled', Burlington House, Piccadilly, London, UK, 11–14 December 2002.

Article

Mapping of Greenhouse Gas Concentration in Peninsular Malaysia Industrial Areas Using Unmanned Aerial Vehicle-Based Sniffer Sensor

Mazlan Hashim [1,2,*], Hui Lin Ng [1], Dahiru Mohammed Zakari [1], Dalhatu Aliyu Sani [1], Musa Muhammad Chindo [1], Noordyana Hassan [1,2], Muna Maryam Azmy [3] and Amin Beiranvand Pour [2,4]

[1] Faculty of Built Environment & Surveying, Universiti Teknologi Malaysia, Johor Bahru 81310, Johor, Malaysia
[2] Geoscience and Digital Earth Centre (INSTeG), Research Institute for Sustainable Environment, Universiti Teknologi Malaysia, Johor Bahru 81310, Johor, Malaysia
[3] School of Tourism, Hospitality & Environmental Management, Universiti Utara Malaysia, Sintok 06010, Kedah Darul Aman, Malaysia
[4] Institute of Oceanography and Environment (INOS), Universiti Malaysia Terengganu (UMT), Kuala Nerus 21030, Terengganu, Malaysia
* Correspondence: mazlanhashim@utm.my; Tel.: +60-197173861

Abstract: The increasing concentration of greenhouse gas (GHG) emissions due to increased fossil fuel consumption for manufacturing activities to support population growth is worrisome. Carbon dioxide (CO_2) and methane (CH_4) remain the two GHGs that contribute to the impact of global warming, and inventorying their concentrations is important for monitoring their changes, which can be used to infer their emissions over time. Hence, this article highlights sniffer4D, an unmanned aerial vehicle (UAV)-based air pollutant mapping system that visualise and analyse three-dimensional (3D) air pollution data in real time, for mapping GHGs concentrations within industrial areas. Consequently, GHGs concentrations for two industrial and adjacent residential areas in Johor, Peninsular Malaysia were mapped. The GHGs concentrations were validated using a ground-based portable gas detector. The results revealed that CO_2 has the highest concentration mean of 625.235 mg/m^3, followed by CH_4 with a mean of 249.239 mg/m^3. The mapped UAV GHG concentration also reported good agreement with the in situ observations with an RMSE of 7 and 6 mg/m^3 for CO^2 and CH_4 concentration, respectively. Ozone and nitrogen dioxide mixture ($O_3 + NO_2$) with a mean concentration of 249 μg/m^3 and an RMSE of 9 μg/m^3 are the remaining significant concentrations reported. This approach shall assist in fast-tracking the United Nations climate change mitigation agenda.

Keywords: GHG concentration; industrial area; remote sensing sensor; UAV; mapping

1. Introduction

Although greenhouse gases (GHGs) are said to have a modulating effect on the Earth's atmosphere, which the atmosphere will be cooler by as much as 33 °C, a process known as the so-called "atmospheric greenhouse effect" [1,2]; an increase in the concentration of GHGs such as carbon dioxide (CO_2) in the atmosphere can result in temperature increases on both the Earth's surface and in the troposphere due to an increase in heat trapping, a concept known as global warming [3]. Global warming is widely acknowledged as a problem worldwide and remains a topic of discussion among scientists and policymakers [4]. Thus, an intergovernmental body, the Intergovernmental Panel on Climate Change (IPCC), has been formed since 1988 with the mandate of assessing the problem of global warming, which remained the focus of much of the ongoing assessments of climate change [3].

The concentration of GHGs in the atmosphere has increased since 1750 (preindustrial) due to rise in emissions from human activities and global energy system [5]. In fact, atmospheric measurements have shown that the concentration of CO_2 and other GHGs has increased by over 20% relative to 1958 [6]. Relating atmospheric GHGs concentration

to emission paths requires models, which account for both natural and anthropogenic sources as well as their sinks [5]. The fact that when CO_2 from either fossil or terrestrial sources is released into the atmosphere, atmospheric concentration increases [5], changes in atmospheric concentrations over time can infer emission increases or decreases. Part of the fallout from the discourse on global warming is the signing of a number of agreements like the Kyoto Protocol and the Paris Agreement by the international community with the ultimate goal of limiting GHG emissions and introducing mechanisms for quota trading, which quite a significant number of countries, including Malaysia, have ratified.

Malaysia signed the Kyoto Protocol on 4 September 2002, prompting the development of a national strategy on the Clean Development Mechanism (CDM) to consider both the long- and short-term perspectives of the country's position on climate change mitigation measures. Malaysia benefited tremendously from investments geared toward reduction in GHG emissions through the CDM under the Kyoto Protocol, with a total of 143 registered CDM projects as of April 2015, which was expected to yield a reduction of 23.95 million tCO_2eq emission by the end of the first Kyoto commitment period (2012). Also, in 2009, Malaysia announced its voluntary commitment to reduce its GHG emissions by up to 40% by 2020 relative to the 2005 levels [7]. Nevertheless, the spatial distribution and magnitude of emissions at fine resolutions are required for monitoring, reporting, and verification of emissions [8], and of course, appraisals of the level of commitments to reduce emissions levels over time.

Generally, two main data types (from point sources and non-point sources) and three approaches (top-down, bottom-up, and hybrid approaches) are used for subnational resolution emissions inventories. Point sources are single identifiable sources of emissions such as gas flaring oil fields and factories, while non-point sources are emissions not originating from discrete sources, such as transportation. Among the databases often used for point and non-point sources inventories are the Carbon Dioxide Information and Analysis Center (CDIAC), Emission Database for Global Atmospheric Research (EDGAR), International Energy Agency (IEA), Fossil Fuel Data Assimilation System (FFDAS), and Open-source Data Inventory for Anthropogenic CO_2 (ODIAC). However, while some of the databases, such as the CDIAC, do not distinguish between point source and non-point source emissions, others, such as EDGAR, do. The top-down approaches often utilize proxies such as population density and or observed nightlight satellite data to distribute emissions spatially within a country. Thus, the spatial resolution of the emission inventory will depend on the proxy data resolution. The bottom-up approaches typically involve collecting fuel consumption or emissions data at buildings or highway segments or even lower scales and summing them up to estimate emissions at local, state, or national scales. The hybrid approaches involve the use of a mix of proxies from the top-down approach and sources from the bottom-up approach [8].

In Malaysia, the industrial processes sector contributed merely 6% of the total 290.23 Mt CO_2eq GHG emissions in the year 2000, lagging behind the energy and waste sectors. However, between 2000 and 2011, industrial processes sector emissions increased by 46%, with carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N2O) emissions contributing 72%, 23% and 5% of the total GHG emissions, respectively, in 2011 [2]. As part of Malaysia's plan to live up to its commitment to reduce GHG emissions, many initiatives are being taken by the government to reduce emissions from point sources such as industrial factories and power plants. Among the initiatives is the Efficient Management of Electrical Energy Regulations 2008, which requires the disclosure of particulars of both new and existing energy consumers with total electricity consumption equal to or exceeding 3,000,000 kWh as measured at one metering point for a consecutive period not exceeding six months by the licensee or supply authority [9].

Under the upcoming Energy Efficiency Conservation Act by the Malaysian government, thermal energy will also be regulated and must be reported [10]. However, self-reported energy and emissions data, particularly thermal data lacking a specific meter such as a power meter that can be monitored in real time and verified by an electricity

utility provider, may be difficult to verify. Thus, remotely sensed (RS) data from either satellite and or lower altitude platforms such as UAV, may aid in providing report on concentration of GHGs on-site and their changes over time, which can invariably be used to infer emissions sources and rate. We embarked on a full RS study on deriving the changes in concentration of GHGs in industrial areas, focusing on both the use of: (i) UAV-sensor for large-scale near-surface mapping; and (ii) satellite image-based for larger area extent concentration mapping. This article focuses on UAV-based GHGs concentration mapping from industrial areas, and part two of the study to be published next will report the large area satellite-based GHGs concentration mapping.

Previously, studies were conducted on mapping GHG emission using UAV [11,12], pollution concentrations [13,14], GHG emissions using satellite observations [15,16], prediction of carbon dioxide (CO_2) concentration and emission inventories [17–19], and biogas potential from plants and manure [20,21]. However, most of these previous studies were mainly on GHG extracted from transportation and vegetation. Indeed, it would be a crucial and desirable goal if the concentration and changes of CO_2, CH_4, and other GHGs in the industrial areas could be systematically mapped. To achieve this, precise mapping and continuous measurement of CO_2 and CH_4 concentrations over a large area and at a periodical/temporal scale with disruptive technology, such as the use of sensors on the drone, calibrated with minimal ground source measurement, is necessary. In addition, the corresponding O_3, NO_2, and SO_2 concentrations could also be determined to complement the CO_2 and CH_4 concentration mapping.

This study, therefore, used a UAV sniffer-based sensor to map changes in GHGs concentration from industrial areas, which is a multi-step approach that can effectively map changes in GHG over an area; this established mapping procedure could be repeated to areas of various sizes permissible pending the range of the UAV-navigation radius. The specific objectives are (a) to characterize the spatial and temporal pattern of the concentration of GHGs (CO_2, CH_4, O_3, NO_2, and SO_2) with emphasis on industrial areas and adjacent residential areas; and (b) to examine and analyse the GHG concentration range based on industry sectors, namely chemical and petrochemical, chemical industry, clay products and refractory, electronics industry, engineering construction, food and beverages, furniture and related products, iron and steel, oil and gas refinery, and oil palm refinery. This spatiotemporal mapping of the concentration of CO_2, CH_4, and other GHGs could later serve for inferring GHGs emissions, assisting in low-carbon planning and speeding up the accomplishment of the 2030 agenda on issues related to sustainable development goal 13, which encourage taking urgent action to combat climate change and its impacts.

2. Materials and Methods

2.1. Study Area

The study areas are the Pasir Gudang and Tanjung Landsat industrial areas. The two industrial areas were employed in the setup for continuous CO_2 and other GHG concentration mapping and monitoring. The Pasir Gudang industrial area (PGIA) is one of the homes of heavy industries in southern Peninsular Malaysia. The PGIA is situated along the Johor State south-eastern tip, covering an area of 1568.673 ha. The area grew rapidly due to the petrochemical and palm oil industries. It is now home to about 300 local and multinational companies, which comprise large-, medium-, and light-scale industries. The main industries at the PGIA include electronics and electrical, research and development, chemicals, food products, engineering-based, and plastics industries. The area is known as the location of the foremost Johor Port as well as the most vital industrial township within the Peninsular Malaysian southern region. The Tanjung Landsat industrial area (TLIA) is located beside the PGIA and covers a total area of about 812.210 ha. TLIA is a fully equipped integrated industrial park and is among the very few industrial parks within southern Malaysia specially chosen for heavy industrial sites. The industry sectors involve chemicals, construction, palm oil, tank storage, warehousing, marine supply, and port services. TLIA is also located close to major seaports, an international airport, and

Singapore. The two industrial locations offer jobs to more than 70,000 people. Figure 1 demonstrates the extent of the study areas.
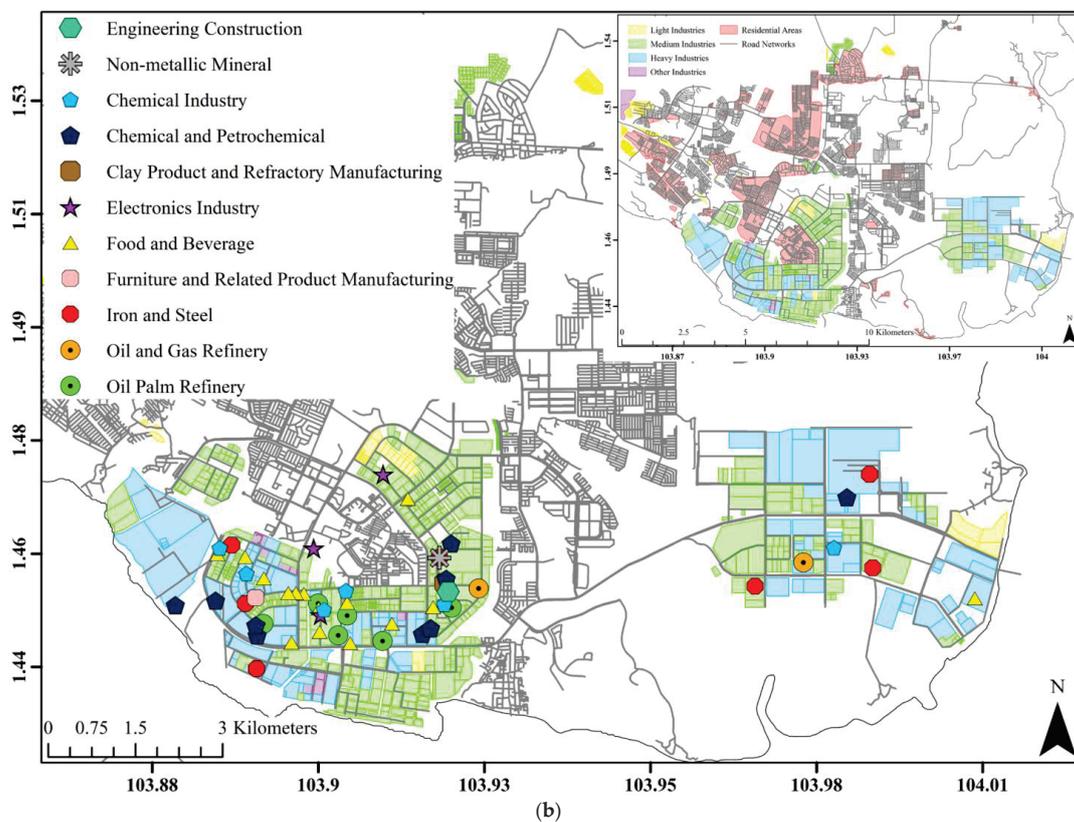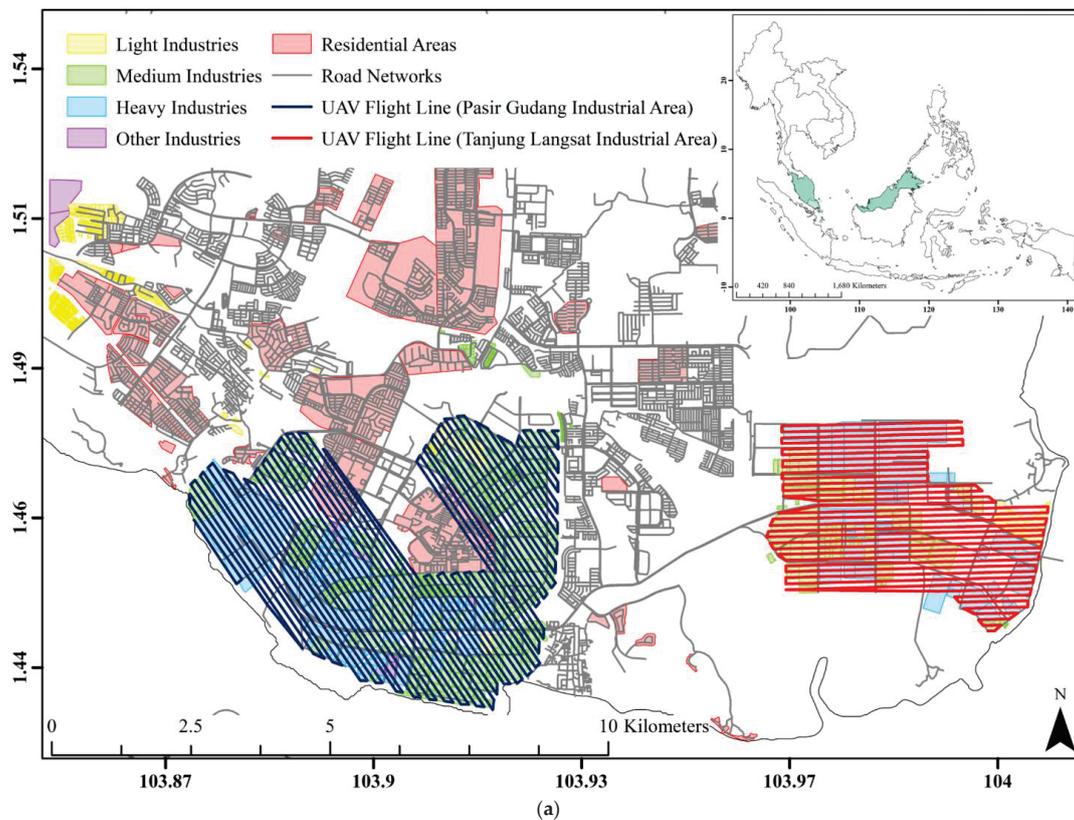

Figure 1. Study area: (**a**) Pasir Gudang and Tanjong Langsat Industrial Area; and (**b**) industrial sectors.

2.2. Materials

The study utilises two important material sets, namely images derived from (a) a UAV-based sniffer4D sensor and (b) a portable GHG detector. The sniffer4D sensor is attached to the UAV for measuring GHGs concentration over relatively large industrial areas in Pasir Gudang (PG) and Tanjung Langsat (TL), while the portable GHG detector is used to measure GHG at ground level at selected locations near real-time of the UAV flight, utilised in calibrating the UAV measurement. Figure 2 depicts the UAV-based sniffer4D sensor and portable gas detector, respectively. The technical details of the sniffer sensor and UAV platform used in this study are shown in Table 1.

(**a**)   (**b**)

Figure 2. (**a**) UAV-based sniffer4D sensor; and (**b**) portable gas detector.

Table 1. Technical specification for UAV-based CO_2 and other GHG emission detection: (**a**) Sniffer sensor; and (**b**) the platform.

(a) UAV and portable ground sensor. (Source; Soarability, (2020))				
Sensor/Brand/Model	**Size**	**Weight**	**Resolution/Response Time**	**Gas Detector**
Sniffer4D	$150 \times 148 \times 50$ mm	600 g	1 ppm/1 s	CO_2, CH_4, CO, $O_3 + NO_2$, and SO_2
Portable gas meter/Weilu/WL-3000	$157 \times 84.5 \times 59.5$ mm	365 g	0.1 ppm/3 s	CO_2, SO_2, NO_2, CH_4
(b) UAV platform (Source; DJI, 2020.)				
Brand/Model	**Drone Type**	**Weight**	**Brand/Model**	**Drone Type**
DJI Matrice 100	Quadra copter	DJI Matrice 100	Quadra copter	DJI Matrice 100

The Sniffer4D Sensor (Table 1a) is a detection and mapping system for GHG concentration data collection. The sensor can be mounted on various platforms, including UAVs, and also works at ground level. The sensor is equipped with a mapping system, able to

capture, visualise, and analyse georeferenced and time-stamped gas concentrations. Up to five GHGs can be sensed simultaneously, and the data can be viewed in real time. The five GHGs sensed are methane (CH_4), carbon monoxide (CO), carbon dioxide (CO_2), mixture of ozone and nitrogen dioxide ($O_3 + NO_2$), and sulfur dioxide (SO_2).

The UAV is a quadcopter flying platform (Table 1b) that has an expandable centre-frame that allows it to mount the sniffer sensor. An extra battery compartment was installed to carry the second intelligent flight battery for an extended flight time of up to 40 min. With the built-in API control feature, navigation for flight missions have been managed comprehensively from planning to execution.

2.3. Methods

2.3.1. UAV Data Acquisition

To obtain the CO_2, CH_4, and other three GHG elements (CO, $O_3 + NO_2$, and SO_2), the UAV-based data collection approach was used. These five gases were acquired in the study area from a field campaign between 15 and 21 October 2020. The data collections were performed using the Sniffer4D sensor, where the flight plans of the study area were designed prior to the data acquisition. The UAV was flown at an altitude of 100 m for data collection. The flight lines and size of the area for the data collection are shown in Figure 1, while the detailed executed data acquisition mission is tabulated in Table 2. The GHGs in all areas in Figure 1 were collected in near real-time and 1 h before and after UAV data acquisition using the Sniffer4D sensor and portable gas detector installed on a car-platform traversing the planned road network to cover the entire study area.

Table 2. UAV data acquisition missions.

	PGIA *	**TLIA ***
Date of data acquisition	15–21 October 2020	15–21 October 2020
Time of data acquisition	0000; 0600; 1200; 1800	0000; 0600; 1200; 1800
Average temperature Min/Max (°C)	24/33	24/33
Average wind speed (km/h)	3	3
Wind direction (bearing)	350°	350°
Period of data acquisition (mins)	245	148
Area of data acquisition (ha)	1837	1177
Total image/files	2880	1090
Front overlap (%)	30	30
Side overlap (%)	30	30
Flight speed (m/s)	15	15
Flight altitude (m)	100	100
Flight direction (bearing)	56°	182°

* Notes: PGIC—Pasir Gudang Industrial Area; TLIA—Tanjung Langsat Industrial Area.

2.3.2. In Situ Data Acquisition

Corresponding in situ surface GHG concentration observations were performed continuously during the period of UAV data acquisition and also one hour before and after UAV flight. A portable GHG detector was used to record the GHGs concentration at selected 52 points to represent the 11 industrial sectors: chemical and petrochemical; chemical industry; clay products and refractory; electronics industry; engineering construction; food and beverages; furniture and related products; iron and steel; oil and gas refinery; oil palm refinery; and non-metallic mineral.

In addition, the corresponding surface temperatures at all these 52 points were also observed using field thermal infrared camera. These temperatures were used as additional information on the surface temperature of the industrial targets where GHG concentrations were observed.

2.3.3. Data Processing of UAV Data

The data processing involved two main tasks: data preprocessing and feature extraction for the mapping of GHG concentration. Figure 3 illustrates the flowchart of the entire stages of the data processing. The data processing tasks were performed using Sniffer4D Mapper software, Arc GIS v10.1, Quantum GIS (QGIS) and Digital Image Processing System (ENVI v5.0).
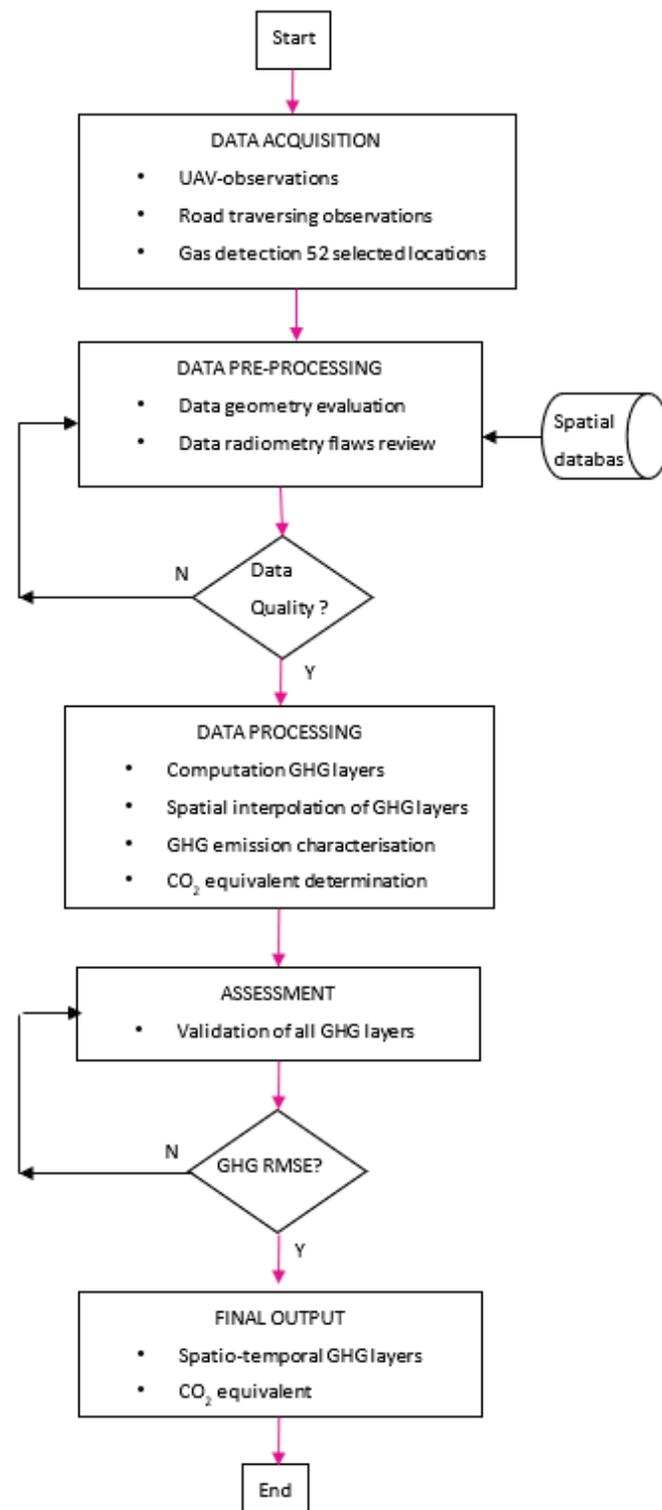
Figure 3. Flowchart of the entire stages of the data processing applied in this study.

In the data preprocessing stage, two main tasks are performed: firstly, ensuring all the UAV acquired data were geometrically corrected to the local mapping system of the area; and secondly, the radiometric data calibration to reduce the GHG concentration at the flying altitude to a level corresponding to near-surface GHG concentration. The calibration was carried out independently by employing linear regression of UAV-Sniffer observations against the corresponding in situ GHG observations. The calibration functions yielded are shown in Figure 4. Applying these calibration functions to the UAV-GHG observation yields adjusted absolute GHG concentration. Table 3 tabulates the summary of the GHG concentration derived from the calibrated concentration data with respect to the area of study.

Table 3. Calibrated GHG concentration data.

GHG	Mean * (mg m^{-3}/μg m^{-3})	Std.dev
CO_2	633.038	28.045
CH_4	303.054	10.667
CO	1.872	0.909
$O_3 + NO_2$	156.162	18.069
SO_2	23.163	2.176

* Notes: the unit for CO_2, CH_4, and CO is in mg m^{-3}, and unit for $O_3 + NO_2$ and SO_2 is in μg m^{-3}.
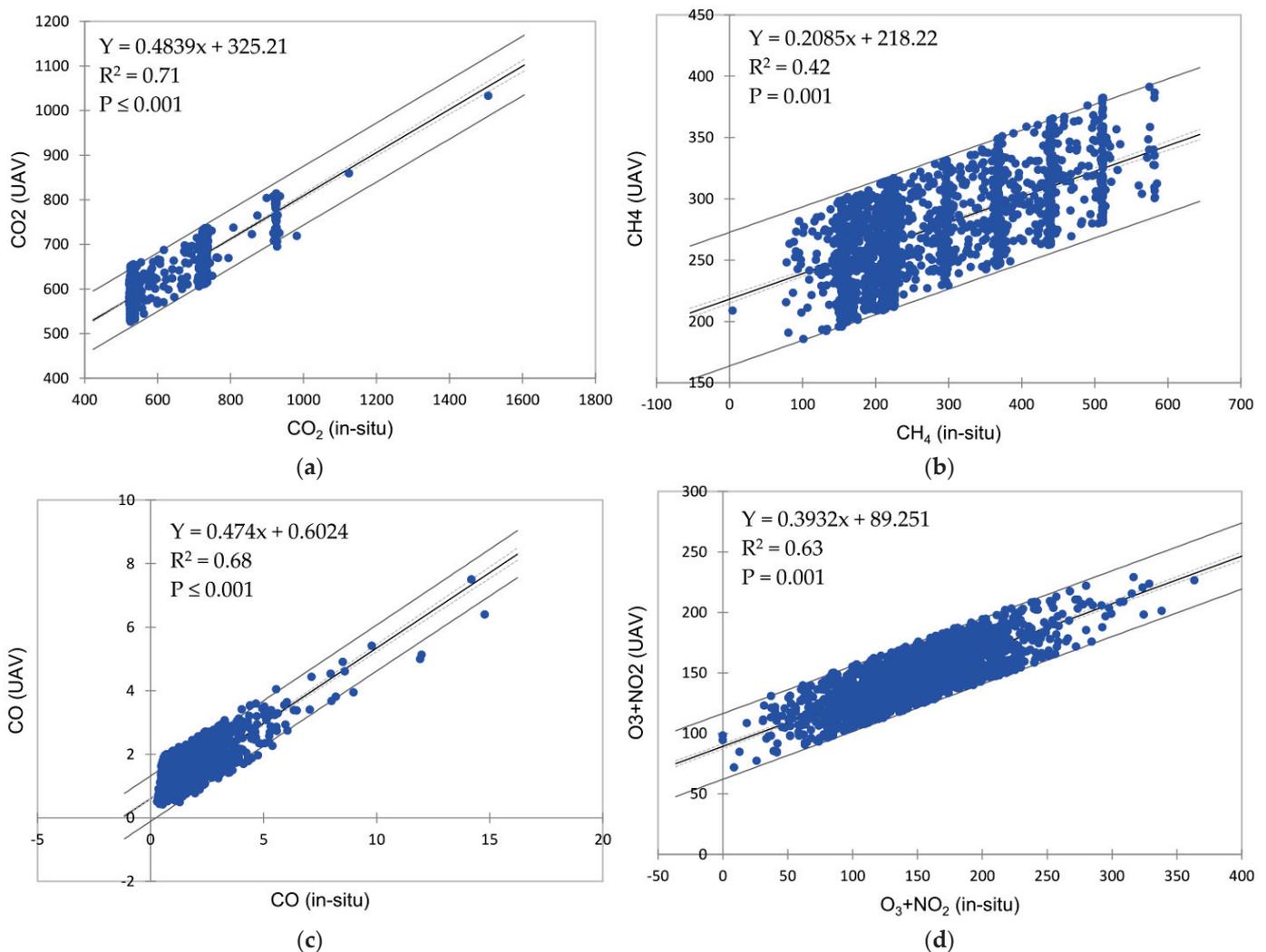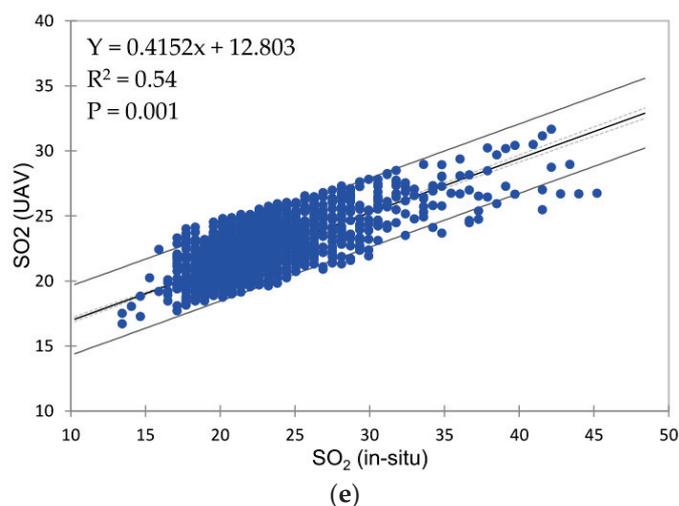
Figure 4. *Cont.*

Figure 4. Calibration functions of GHG generated using linear regression analysis of UAV observations versus in situ ground measurement: (**a**) CO_2, (**b**) CH_4, (**c**) CO, (**d**) O_3 + NO_2 mixture and (**e**) SO_2.

The data processing involves the mapping of the five GHG concentrations into separate layers, each representing CO_2 and the other four greenhouse gases—CH_4, CO, O_3 + NO_2 mixture, and SO_2. UAV data processing for mapping GHG concentration is elaborated in the following sections.

2.3.4. Mapping of CO_2 and Greenhouse Gasses

The mapping of CO_2 and other GHG concentrations starts with the densification of the gas spatial distributions and observations of the daily concentration throughout the mission period over the area of study. This is accomplished using an interpolation process where the concentration of gases at unknown points at selected regular planimetric locations were determined. The Inverse Distance Weighted (IDW) interpolation technique of the QGIS software was used on the data with the same layer extent. The interpolation parameters set for the IDW are (i) the distance coefficient (P: 2) and (ii) the output raster pixel size (X and Y: 10 m). The main input is the mean daily GHG concentration, i.e., a total of seven layers for a week of field campaign data acquisition period. The output of this interpolation is the average daily GHG concentration.

The outputs created from interpolation were clipped with shapefiles in ArcGIS software for mapping GHG concentration for residential areas and industrial zones (light, medium, and heavy industries) and GHG concentrations against the identified industrial sectors: (1) chemical and petrochemical; (2) chemical industry; (3) clay products and refractory; (4) electronics industry; (5) engineering construction; (6) food and beverages; (7) furniture and related products; (8) iron and steel; (9) oil and gas refinery; (10) oil palm refinery; and (11) non-metallic mineral.

Statistical analyses of GHG concentration were performed to characterize the pattern of GHG concentration in the industrial areas (PGIA and TLIA) and adjacent residential areas. The GHG concentration was characterized based on industrial manufacturing sectors. In this analysis, the GHG concentrations were obtained at 52 validation points, which were also earlier used as inputs in the assessment of the derived GHG concentration with the corresponding in situ observation. In the assessment of GHG concentration, a comparison of both UAV-derived concentration and the corresponding in situ observations was performed at 30 random points from the pool of the 52 known validation points, reporting the root mean square error (RMSE). The analysis of variance (ANOVA), which is a statistical test used for determining the significant difference between categorical groups of data by testing differences of means using variance, was used to report the trend of total GHG concentration with respect to all the industrial sectors. Correlation analysis among

these GHG concentrations was also examined to consolidate the ANOVA test. In addition, the surface temperature at the 52 validation points was also observed and analysed in the ANOVA and the correlation analysis to assist in inferring the outcomes of the GHG concentrations, as the rank of hotness among the industrial sectors.

3. Results

3.1. GHG Concentrations and Distribution for the Entire Study Area

The average spatial distribution of all the GHG emission concentrations in the study area is shown in Figure 5, with the magnitude ranges within the industrial zones presented in Table 4. The trend of the spatial concentration of the CO_2 is heavily located in the western-half of the study area, with PGIA as the focal source where the concentrations are dominated, ranging from mid-to-high concentration, while the eastern-half focusing on TLIA has low-to-mid concentration. The other four GHG concentrations follow the same spatial pattern but with different magnitudes of occurrence. CO_2 concentration is the highest among GHG concentration within industrial area zones, with little variation in intensity and respective means of 633, 628, and 614 g/m^3 for light, medium, and heavy industrial zones. CO_2 concentration at the residential area recorded a higher mean of 649 mg/m^3 compared to 625.235 mg/m^3 for the entire industrial area, with a similar trend in the other four GHG concentration. CH_4 is the second most prominent GHG in PGIA and TLIA, with a mean of 303, 289, and 290 mg/m^3 for light, medium, and heavy industrial zones, respectively. CH_4 mean concentration in the residential area is at 293 mg/m^3.

Table 5 presents the RMSE of in situ GHG concentrations by gas detector at 30 independent sample points chosen randomly from 52 in situ validation points collected from all industrial sectors in PGIA and TLIA industrial areas relative to the UAV concentration maps. $O_3 + NO_2$ has the highest RMSE of 8.964, followed by CO_2 and CH_4 with an RMSE of 7.418 and 6.134, respectively. CO has the lowest RMSE of 0.234.

Table 4. Average GHG concentration trend in industrial areas (PG and TL) and residential areas.

GHG	Entire Industrial Area Mean * (mg m^{-3}/μg m^{-3})	Std.dev	Light Industries Mean * (mg m^{-3}/μg m^{-3})	Std.dev	Medium Industries Mean * (mg m^{-3}/μg m^{-3})	Std.dev	Heavy Industries Mean * (mg m^{-3}/μg m^{-3})	Std.dev	Residential Area Mean * (mg m^{-3}/μg m^{-3})	Std.dev
CO_2	625.24	30.67	633.04	28.05	628.04	31.16	614.63	32.82	649.82	14.58
CH_4	294.24	8.96	303.05	10.67	289.27	8.05	290.40	8.17	293.72	9.06
CO	1.43	0.59	1.87	0.91	1.29	0.40	1.14	0.46	1.87	0.52
$O_3 + NO_2$	149.30	15.63	156.16	18.07	148.44	12.55	143.29	16.26	157.92	9.05
SO_2	22.24	1.62	23.16	2.18	22.07	1.22	21.49	1.46	23.42	1.09

* Notes: the unit for CO_2, CH_4 and CO is in mg m^{-3}, and unit for $O_3 + NO_2$ and SO_2 is in μg m^{-3}.

(**a**)

(**b**)
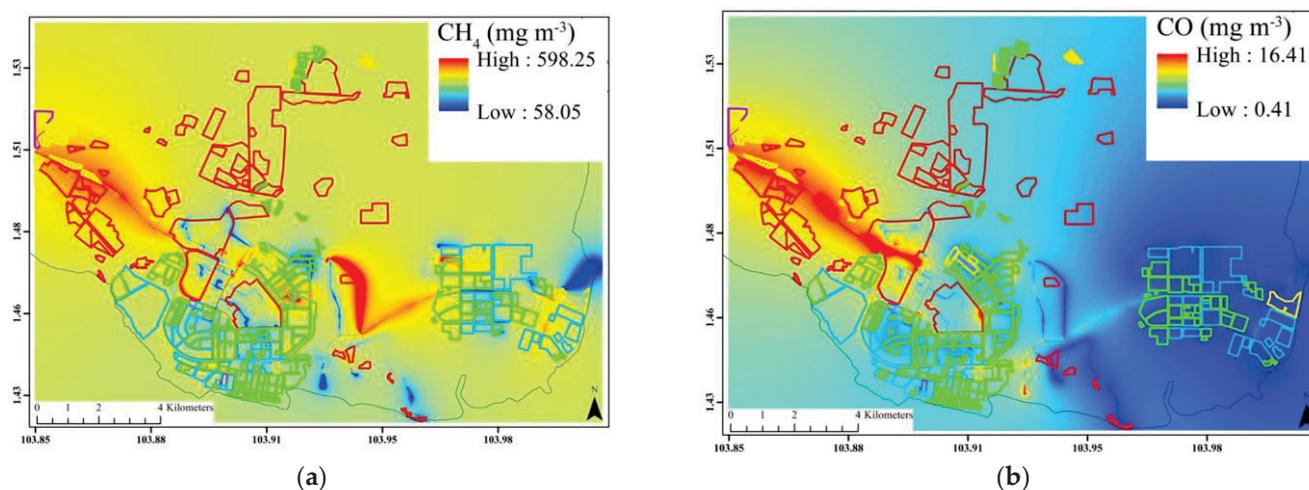
Figure 5. *Cont.*
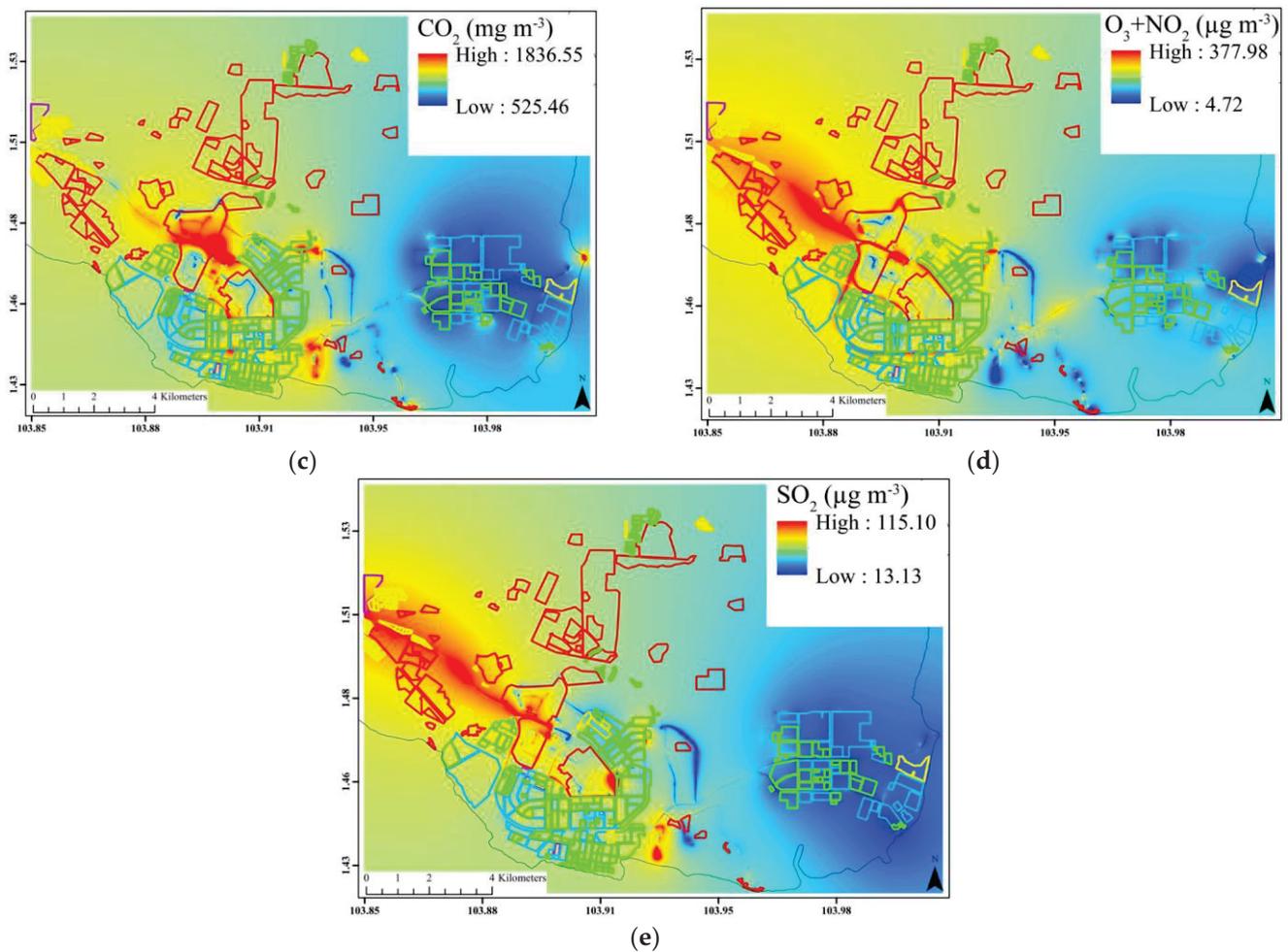
Figure 5. Average GHG concentration of entire Pasir Gudang and Tanjung Langsat area: (**a**) CH_4; (**b**) CO; (**c**) CO_2; (**d**) $O_3 + NO_2$; and (**e**) SO_2.

Table 5. Assessment of GHG concentration maps generated.

GHG Emissions	RMSE *
CO_2	7.418
CH_4	6.134
CO	0.234
$O_3 + NO_2$	8.964
SO_2	0.819

* Note: the unit for CO_2, CH_4, and CO is in mg m^{-3}, and unit for $O_3 + NO_2$ and SO_2 is in µg m^{-3}.

3.2. Characterization of GHG Pattern in the Industrial Sector

Figure 6 depicts the percentage concentration of total GHG in relation to all gases for the industry sectors. Across all the sectors, CO_2 and CH_4 are the main gases as well. However, the concentration of the two main gases varies with the industrial sector with the highest concentration of CO_2 found in the electronics industry and the lowest in engineering construction. While the highest concentration of CH_4 is found in engineering construction and the lowest is found in food and beverages. The temperature of the respective industries' activities was also added in the analysis and shown as a good indicator for GHG concentrations.
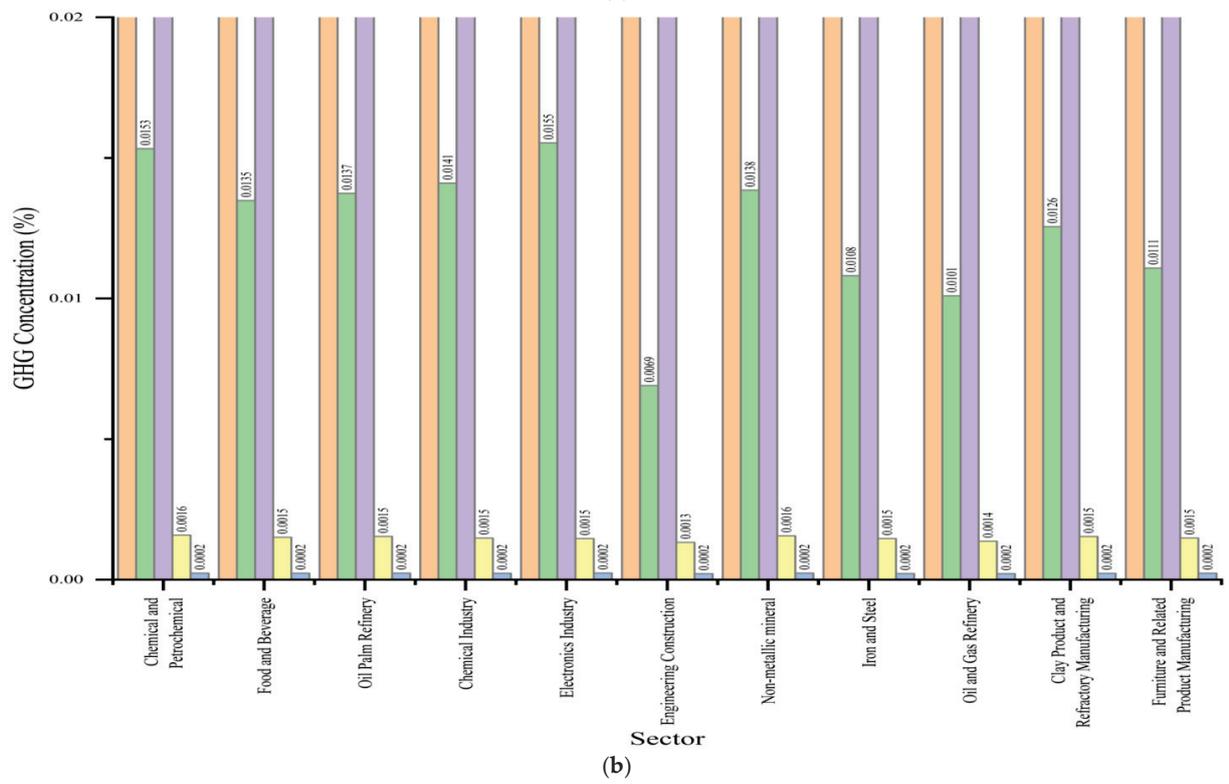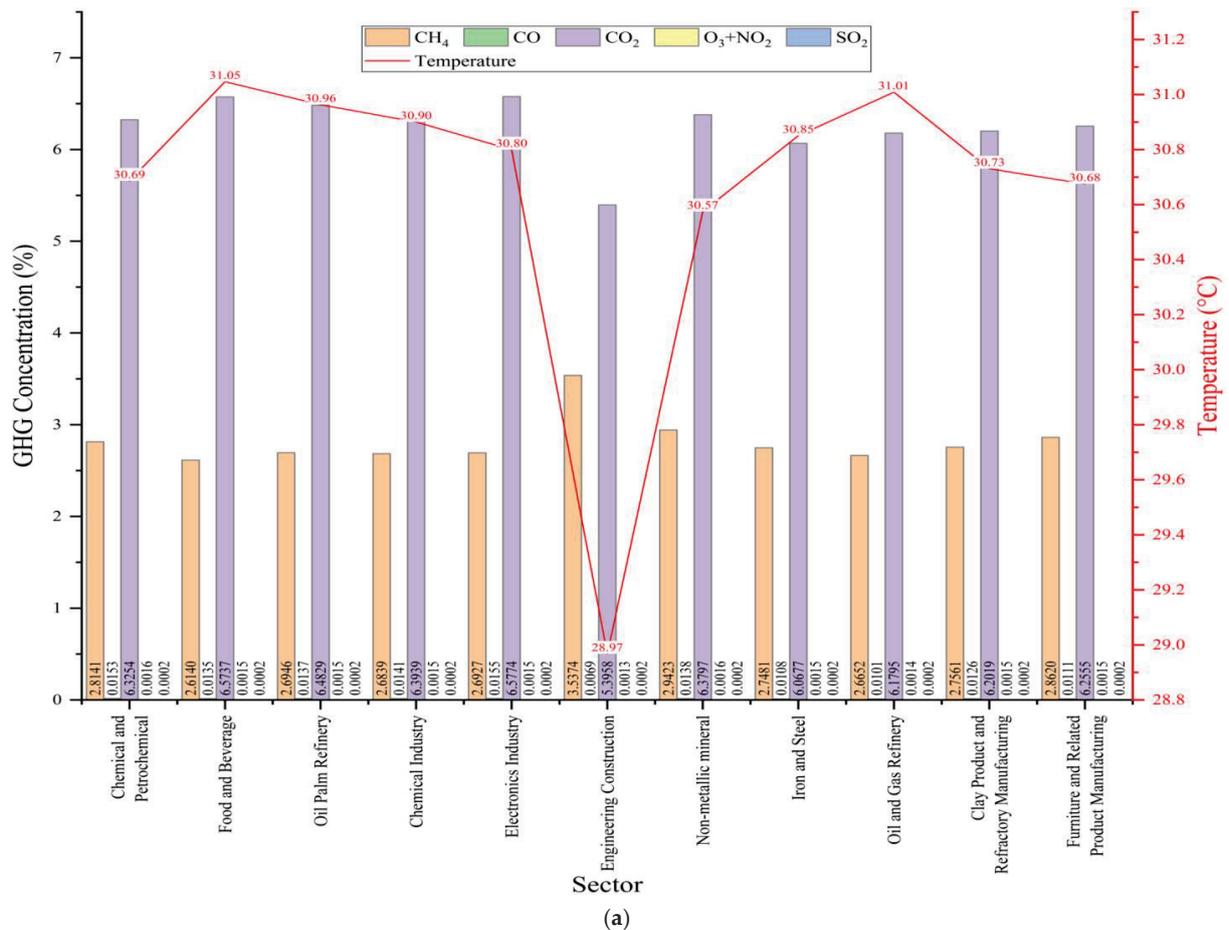
Figure 6. Average GHG concentrations and temperature of different industrial sectors (*n* = 52). Note: the colours of the other GHG bars are too small to be seen in (**a**), as such they are exaggerated in (**b**).

The ANOVA test was further conducted to confirm GHG concentration variations with respect to industrial activities. Table 6 summarises the ANOVA test results. The mean square, F-value, and Pr(>F) for the GHG gases concentration are presented. The mean square suggests the average amount of gas concentration from each industry, while the F-value suggests the amount of average difference by industry. Pr(>F) suggests the significant error. Consequently, the result reveals only three GHG concentration, namely SO_2, CO_2, and CO, differ significantly by industrial activity.

Table 6. ANOVA test for gases by industrial activities ($n = 52$).

	Mean Square	F-Value	Pr(>F)
SO_2	34.77	6.455	0.0142 *
CO_2	35.56	6.621	0.0131 *
CO	31.66	5.812	0.0196 *
$O_3 + NO_2$	13.16	2.262	0.1390
CH_4	5.48	0.917	0.3430
Temperature °C	49.93	9.824	0.0029 **

* Significant result at $p < 0.01$; ** Significant result at $p < 0.001$.

In addition, the temperature of these industrial activities also significantly differs among the industrial activities. The other two GHG concentrations, namely $O_3 + NO_2$ and CH_4, however, are not significantly different among industrial activities. Further investigation into the relationship between GHG concentration and temperature within industrial activities was also carried out, and the results of this correlation analysis revealed that CO_2 is mostly related to all gas types and temperatures (see Table 7).

Table 7. Correlation analysis of GHG concentration and temperature with the industrial sector ($n = 52$).

	Temp	SO_2	O_3NO_2	CO_2	CO	CH_4
Temp	1.00	0.25 **	−0.15	0.46 **	0.01	−0.89
SO_2	0.25 **	1.00	0.41 **	0.69 **	0.78 **	−0.25
$O_3 + NO_2$	−0.15	0.41 **	1.00	0.48 **	0.73 **	0.16
CO_2	0.46 **	0.69 **	0.48 **	1.00	0.60 **	−0.46 **
CO	0.01	0.78 **	0.73 **	0.60 **	1.00	0.00
CH_4	−0.89 **	−0.25	0.16	−0.46 **	0.00	1.00

** Correlation significant at $p < 0.001$.

3.3. Changes in GHG Concentration for PGIA and TLIA

Figure 7 depicts the average hourly GHG concentration for the seven-day period of data collection. While Table 8 presents the daily average concentration for the five respective GHG changes in GHG concentration for the entire industrial and residential area through the seven-day measurement period.

As earlier noted, CH_4 and CO_2 are the main GHGs with significant concentrations, followed by the $O_3 + NO_2$ mixture, while CO and SO_2 have relatively insignificant concentrations in both the industrial and residential areas. The lowest concentration of CO_2 and CH_4 were recorded on the same day, i.e., the second over both the industrial and residential areas, while the highest concentration for CO_2 and CH_4 were respectively recorded on the sixth and seventh day. However, for the $O_3 + NO_2$ mixture, the lowest concentrations over both the industrial and residential areas were recorded on the fourth day (the same day that the lowest CO_2 was recorded), while the highest were recorded on the third day.
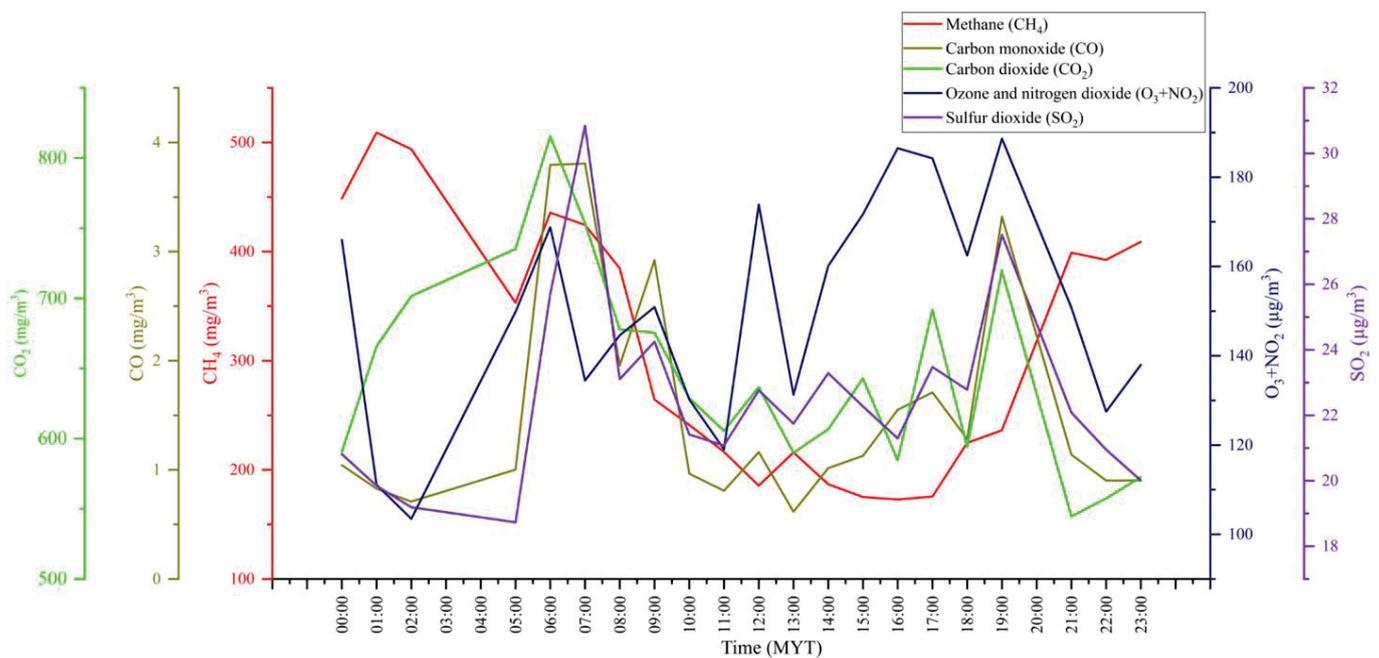
Figure 7. Hourly GHG concentration graph (based on average of 7 days of data). Note: MYT—Malaysia Standard Time.

Table 8. Mean concentration of GHGs for entire industrial and residential area through the study period after calibration.

| GHG | CH_4 (mg m^{-3}) | | CO (mg m^{-3}) | | CO_2 (mg m^{-3}) | | $O_3 + NO_2$ (µg m^{-3}) | | SO_2 (µg m^{-3}) | |
Date	IA *	RA *	IA *	RA *	IA *	RA *	IA *	RA *	IA *	RA *
15 October 2020	198.79	204.80	1.50	1.87	604.59	632.32	166.83	178.42	22.54	24.09
16 October 2020	176.36	174.77	1.53	2.02	612.61	639.25	140.13	146.42	23.10	24.02
17 October 2020	212.24	208.00	1.23	1.51	633.77	665.11	167.23	181.48	23.08	23.90
18 October 2020	399.66	413.96	0.93	1.13	544.15	547.40	134.14	139.30	20.58	21.33
19 October 2020	294.24	293.72	1.43	1.87	625.24	649.82	149.30	157.92	22.24	23.42
20 October 2020	333.22	332.07	1.34	2.02	689.50	710.31	142.72	144.14	21.63	23.69
21 October 2020	445.16	428.74	2.06	2.68	666.80	704.52	144.74	157.73	22.51	23.51

* Note: IA—Industrial Area; RA—Residential Area.

4. Discussion

The industrial GHG concentration mapping in this article relies on UAV sniffer4D sensor-based data acquired during a field campaign between 15 October and 21 October 2020, at PG and TL industrial areas. Adjacent residential area GHG concentrations were also mapped accordingly to see the pattern and compare with those of industrial area. It is, however, important to note that the campaign period was at the end of the inter-monsoon two, when wind speed is almost nil. Notwithstanding, the average windspeed in the area then was 3 km/h as presented in Table 2, and the wind is moving in the northerly direction as it is the transition period from the southwest to northeast monsoons.

However, this article reveals similar patterns of CO_2, CH_4, CO, $O_3 + NO_2$, and SO_2 concentration exist between industrial and neighbouring residential areas and among the various industrial zones (light, medium, and heavy) within the industrial areas. Overall, CO_2, followed by CH_4, constitutes the largest GHG concentration among the two industrial areas and within the various zones and sectors. Nevertheless, the GHG concentration does not vary much with industrial activities. The large variability is very prominent in the engineering construction industry, and the least variability is found within oil palm refineries. An intriguing fact is the inversion of the GHG concentration among the three industrial zones, with a slightly higher mean in the lighter industrial zone compared to the medium and heavy industrial zones. However, upon cursory investigations at selected locations, it is noted that most of the factories in the heavy zone, such as refineries, including

the oil palm refineries, are equipped with indoor air filtration before releasing it. This, of course, highlights the importance of not overlooking emissions from specific industrial zones or sectors solely based on their kind of activities in low-carbon planning.

Furthermore, this study reveals a higher concentration of GHG over neighbouring residential areas. However, this will not be unrelated to pollution from the industrial areas due to wind transport, as prevailing wind direction during the field campaign shows a northerly flow from the south. Nonetheless, a similar report of higher concentration of GHGs over neighbouring residential areas of industrial areas has long been made about the city of Toronto [22] and 21 countries in Europe between 2007 and 2017 [23]. This calls for a comprehensive look at all environmental components in sectoral land allocations during land-use planning and highlights the need for a cursory look at the concentration of GHGs over residential areas neighbouring industrial areas. Similarly, this study reveals a diurnal temporal variability of GHGs concentrations as seen in Figure 7. This may not be unrelated to the change in the pattern of activities in the respective factories that run 24 h, but characterised by a slowdown of activities during certain times of the day, particularly during labour shift hours. Additionally, our one-week campaign also showed a gradual increase in the GHGs concentrations, though with slight variability in some days, particularly on the fourth day of the campaign as presented on Table 8. However, we believe a longer campaign period can reveal a better temporal pattern of GHGs concentration over any industrial area. Finally, the established relationship between GHG concentration and temperature within industrial activities in this study indicates that CO_2 concentration is the best indicator of GHG emissions from fuel burning in industrial areas.

The deployment of UAV for mapping GHG concentration in industrial areas is indeed a step-up initiative towards effective monitoring of GHG emissions and their global warming potential, as well as the realisation of a low-carbon economy (LCE). Aside from the obvious near-ground synoptic measurement, this approach can better detect concentration changes that can be used to infer emission hotspots compared to the use of proxy data. Relative to other approaches for measuring GHG emissions, such as the ground-based level measurements at selected points, the UAV sniffer4D sensor-based approach provides rapid, comprehensive, and yet very cost-effective wider area coverage of GHGs concentration mapping. Thus, this work is very crucial and timely, particularly to Malaysia's plan to reduce carbon emissions by 45 percent by 2030. With this study's approach, the baseline and regular status of CO_2 concentration for any specific industrial area with respect to time and date are indeed timely. Similarly, the mapping of GHG concentration in industrial and residential areas may be useful to several private and public sectors concerned about GHG inventory and reduction initiatives, as it could also help individuals or industries gather information that can be used as a reference to prepare a more detailed GHG inventory, which can be used to make inferences about emissions. Nonetheless, industrial areas will continue generating CO_2, CH_4, and NO_2. Once the concentration of these GHGs can be effectively mapped, strategies for their capture can more easily be developed. If this is done, they can be used to make biogas, which will always be a useful source of energy.

5. Conclusions

Industrial manufacturing processes produce a large proportion of GHG emissions due to fossil fuel burning. The Pasir Gudang industrial area, as the focal industrial area, has been a source of pollution that affects the neighbouring areas, including residential areas, resulting in several reported cases of mortality and health-related problems. This study explores the capability of the UAV-based Sniffer4D sensor as a rapid mapping system to characterise the spatial and temporal pattern of the concentration of GHGs (CO_2, CH_4, O_3, NO_2, and SO_2) with emphasis on industrial areas and adjacent residential areas, and examine and analyse the GHG concentration range based on industry sectors, namely chemical and petrochemical, chemical industry, clay products and refractory, electronics industry, engineering construction, food and beverages, furniture and related products, iron and steel, oil and gas refinery, and oil palm refinery. However, this initiative can speed

up the accomplishment of the 2030 agenda on issues related to sustainable development goal 13, which encourages taking urgent action to combat climate change and its impacts.

References

1. Manabe, S. Role of greenhouse gas in climate change. *Tellus A Dyn. Meteorol. Oceanogr.* **2019**, *71*, 1–13. [CrossRef]
2. Smith, A.P. Proof of the Atmospheric Greenhouse Effect. *arXiv* **2008**, arXiv:0802.4324.
3. Houghton, J. *Global Warming: The Complete Briefing*; Cambridge University Press: Cambridge, UK, 2004.
4. Halushchak, M.; Bun, R.; Shpak, N.; Valakh, M. Modeling and spatial analysis of greenhouse gas emissions from fuel combustion in the industry sector in Poland. *Int. Q. J.* **2016**, *5*, 19–26.
5. Clarke, L.; Edmonds, J.; Jacoby, H.; Pitcher, H.; Reilly, J.; Richels, R. *CCSP Synthesis and Assessment Product 2.1, Part A: Scenarios of Greenhouse Gas Emissions and Atmospheric Concentrations*; US Government Printing Office: Washington, DC, USA, 2006.
6. IPCC. Climate change 2013: The physical science basis. In *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2013. Available online: http://www.ipcc.ch/report/ar5/wg2/ (accessed on 5 November 2022).
7. Ministry of Natural Resources and Environment Malaysia. Malaysia: Biennial Update Report to the United Nations Framework Convention on Climate Change. 2015. Available online: https://unfccc.int/sites/default/files/resource/MALBUR1.pdf (accessed on 23 October 2022).
8. Hutchins, M.G.; Colby, J.D.; Marland, G.; Marland, E. A comparison of five high-resolution spatially-explicit, fossil-fuel, carbon dioxide emission inventories for the United States. *Mitig. Adapt. Strateg. Glob. Change* **2017**, *22*, 947–972. [CrossRef]
9. Malaysia Energy Commission. Efficient Management of Electrical Energy Regulations 2008. Available online: https://www.st.gov.my/images/article/polisi/regulation_suruhanjaya/bi-efficient_management_of__electrical_energy_regulations_2008.pdf (accessed on 20 October 2022).
10. Ashrae Malaysia Chapter. Energy Efficiency and Conservation Act Final Drafting. 2017. Available online: https://www.ashrae.org.my/newsinfo/energy-efficiency-and-conservation-act-final-drafting-1 (accessed on 23 October 2022).
11. Gålfalk, M.; Nilsson Påledal, S.; Bastviken, D. Sensitive Drone Mapping of Methane Emissions without the Need for Supplementary Ground-Based Measurements. *ACS Earth Space Chem.* **2021**, *5*, 2668–2676. [CrossRef] [PubMed]
12. Mlambo, R.; Woodhouse, I.H.; Gerard, F.; Anderson, K. Structure from Motion (SfM) Photogrammetry with Drone Data: A Low Cost Method for Monitoring Greenhouse Gas Emissions from Forests in Developing Countries. *Forests* **2017**, *8*, 68. [CrossRef]
13. Afzali, A.; Rashid, M.; Afzali, M.; Younesi, V. Prediction of air pollutants concentrations from multiple sources using AERMOD coupled with WRF prognostic model. *J. Clean. Prod.* **2017**, *166*, 1216–1225. [CrossRef]
14. Spezzano, P. Mapping the susceptibility of UNESCO World Cultural Heritage sites in Europe to ambient (outdoor) air pollution. *Sci. Total Environ.* **2021**, *754*, 142345. [CrossRef] [PubMed]
15. Buchwitz, M.; Reuter, M.; Schneising, O.; Hewson, W.; Detmers, R.G.; Boesch, H.; Hasekamp, O.P.; Aben, I.; Bovensmann, H.; Burrows, J.P. Global satellite observations of column-averaged carbon dioxide and methane: The GHG-CCI XCO2 and XCH4 CRDP3 data set. *Remote Sens. Environ.* **2017**, *203*, 276–295. [CrossRef]
16. Van Beijma, S.; Chatterton, J.; Page, S.; Rawlings, C.; Tiffin, R.; King, H. The challenges of using satellite data sets to assess historical land use change and associated greenhouse gas emissions: A case study of three Indonesian provinces. *Carbon Manag.* **2018**, *9*, 399–413. [CrossRef]
17. He, Z.; Lei, L.; Zhang, Y.; Sheng, M.; Wu, C.; Li, L.; Zeng, Z.C.; Welp, L.R. Spatio-temporal mapping of multi-satellite observed column atmospheric CO_2 using precision-weighted kriging method. *Remote Sens.* **2020**, *12*, 576. [CrossRef]
18. Wang, K.Y.; Wang, J.L.; Liu, W.T. Ambient carbon dioxide concentrations in industrial park areas: A monitoring and modeling study. *Atmos. Pollut. Res.* **2014**, *5*, 179–188. [CrossRef]

19. Zeng, Z.C.; Lei, L.; Strong, K.; Jones, D.B.; Guo, L.; Liu, M.; Deng, F.; Deutscher, N.M.; Dubey, M.K.; Griffith, D.W. Global land mapping of satellite-observed CO_2 total columns using spatio-temporal geostatistics. *Int. J. Digit. Earth* **2017**, *10*, 426–456. [CrossRef]
20. Scarlat, N.; Fahl, F.; Dallemand, J.F.; Monforti, F.; Motola, V. A spatial analysis of biogas potential from manure in Europe. *Renew. Sustain. Energy Rev.* **2018**, *94*, 915–930. [CrossRef]
21. Spyridonidis, A.; Vasiliadou, I.A.; Akratos, C.S.; Stamatelatou, K. Performance of a Full-Scale Biogas Plant Operation in Greece and Its Impact on the Circular Economy. *Water* **2020**, *12*, 3074. [CrossRef]
22. VandeWeghe, J.R.; Kennedy, C. A Spatial Analysis of Residential Greenhouse Gas Emissions in the Toronto Census Metropolitan Area. *J. Ind. Ecol.* **2007**, *11*, 133–144. [CrossRef]
23. Zhang, Y.; Abbas, M.; Koura, Y.H.; Su, Y.; Iqbal, W. The impact trilemma of energy prices, taxation, and population on industrial and residential greenhouse gas emissions in Europe. *Environ. Sci. Pollut. Res.* **2021**, *28*, 6913–6928. [CrossRef] [PubMed]

MDPI

Article

Neuro-Fuzzy-AHP (NFAHP) Technique for Copper Exploration Using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and Geological Datasets in the Sahlabad Mining Area, East Iran

Aref Shirazi [1], Ardeshir Hezarkhani [1,*], Amin Beiranvand Pour [2,3,*], Adel Shirazy [1] and Mazlan Hashim [3]

[1] Department of Mining Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran 1591634311, Iran

[2] Institute of Oceanography and Environment (INOS), University Malaysia Terengganu (UMT), Kuala Nerus 21030, Malaysia

[3] Geoscience and Digital Earth Centre (INSTeG), Research Institute for Sustainable Environment, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia

* Correspondence: ardehez@aut.ac.ir (A.H.); beiranvand.pour@umt.edu.my (A.B.P.); Tel.: +98-21-6454-2968 (A.H.); +60-9-668-3824 (A.B.P.); Fax: +98-21-6640-5846 (A.H.); +60-9-669-2166 (A.B.P.)

Abstract: Fusion and analysis of thematic information layers using machine learning algorithms provide an important step toward achieving accurate mineral potential maps in the reconnaissance stage of mineral exploration. This study developed the Neuro-Fuzzy-AHP (NFAHP) technique for fusing remote sensing (i.e., ASTER alteration mineral image-maps) and geological datasets (i.e., lithological map, geochronological map, structural map, and geochemical map) to identify high potential zones of volcanic massive sulfide (VMS) copper mineralization in the Sahlabad mining area, east Iran. Argillic, phyllic, propylitic and gossan alteration zones were identified in the study area using band ratio and Selective Principal Components Analysis (SPCA) methods implemented to ASTER VNIR and SWIR bands. For each of the copper deposits, old mines and mineralization indices in the study area, information related to exploration factors such as ore mineralization, host-rock lithology, alterations, geochronological, geochemistry, and distance from high intensity lineament factor communities were investigated. Subsequently, the predictive power of these factors in identifying copper occurrences was evaluated using Back Propagation Neural Network (BPNN) technique. The BPNN results demonstrated that using the exploration factors, copper mineralizations in Sahlabad mining area could be identified with high accuracy. Lastly, using the Fuzzy-Analytic Hierarchy Process (Fuzzy-AHP) method, information layers were weighted and fused. As a result, a potential map of copper mineralization was generated, which pinpointed several high potential zones in the study area. For verification of the results, the documented copper deposits, old mines, and mineralization indices in the study area were plotted on the potential map, which is particularly appearing in high favorability parts of the potential map. In conclusion, the Neuro-Fuzzy-AHP (NFAHP) technique shows great reliability for copper exploration in the Sahlabad mining area, and it can be extrapolated to other metallogenic provinces in Iran and other regions for the reconnaissance stage of mineral exploration.

Keywords: copper exploration; machine learning; BPNN; NFAHP; ASTER; geological data; mineral potential map

1. Introduction

Machine learning approaches are steadfast tools of mineral exploration due to their capability for precise processing of remote sensing data and fusing various and high dimensional data to detect features with problematic attributes [1–4]. The utilization of

machine learning algorithms is an inexpensive and automatic approach for accurate mineral potential mapping, fusing the data derived from remote sensing, geology, geophysics, and geochemistry [1,3–6]. Machine learning algorithms are typified to (i) dimensionality reduction methods such as Principal Component Analysis, Independent Component Analysis and Minimum Noise Fraction; (ii) classification methods such as Minimum Distance, Support Vector Machine, Artificial Neural Networks and Random Forest; (iii) regression methods such as Multi-Linear Regression, Multivariate Regression, Logistic Regression; and (iv) clustering methods such as K-means and ISODATA [7].

The application of machine learning methods, particularly Artificial Neural Networks (ANN), has great potential in processing of various data for accurate mineral potential mapping [8]. Appropriate training of a Neural Network (NN) is a significant characteristic of producing a consistent model. This training is generally named "Back-propagation", which is the principle of neural net training. Back-propagation is the exercise of fine-tuning the weights of a neural net built on the error rate acquired in the preceding epoch. Appropriate tuning of the weights guarantees lower error rates, generating the more steadfast model by improving its generalization [9–11]. The Back Propagation Neural Network (BPNN) algorithm is a practical approach for improving the accuracy of predictions in data mining [12,13]. For mineral exploration, one of the most significant concerns is the identification of potential zones based on the characteristics of ore deposits, mining area, and mineral occurrences (indices) in a study area [14]. One of the most important benefits of BPNN is the ability to estimate the predictive power and accuracy of factors related to mineralization, which is accomplished through training and testing. In fact, the power of network estimation is evaluated using available data.

In this study, the BPNN was used to evaluate the accuracy of exploration factors related to volcanic massive sulfide (VMS) copper mineralization in the Sahlabad mining area, South Khorasan province, east Iran (Figure 1A,B). Consequently, the predictive power and accuracy of information layers such as host-rock lithology, alteration minerals, geological age of the host-rock, ore mineralization and distance from the community of important fault systems were evaluated to determine the location of copper occurrences in the study area. As a result, the predictive power of the data was accomplished with a lower percentage of error. In fact, implementation of the BPNN algorithm to the exploration factors of copper mineralization has verified the accuracy of inputs layers/information for the next stage [15,16]. Subsequently, the information layers were weighted and fused according to the rules of the Fuzzy Analytic Hierarchy Process (AHP) [17–19]. In this study, the combination of BPNN and Fuzzy-AHP methods in the field of mineral exploration is executed and introduced as Neuro-Fuzzy-AHP (NFAHP).

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) remote sensing sensor has great capabilities to map hydrothermal alteration zones associated with a variety of ore mineralization particularly massive sulfide copper mineralization [20–22]. The hydrothermal alteration zones (e.g., gossan, argillic, phyllic and propylitic) associated with volcanic massive sulfide (VMS) copper mineralization in the Sahlabad mining area have been reported and documented [23]. Consequently, remote sensing and geological information layers, including ASTER alteration maps, host-rock lithology, geochronological, structural data and geochemistry of copper mineralization in the study area were evaluated and verified by the BPNN, then appraised and appropriate layers were subsequently fused by the Fuzzy-AHP method for generating a mineral potential map. In view of that, the main objectives of this study are: (1) to map hydrothermal alteration zones (e.g., gossan, argillic, phyllic and propylitic) associated with copper mineralization using the visible and near infrared (VNIR) and shortwave infrared (SWIR) bands of ASTER data; (2) to estimate the predictive power and accuracy of information layers (remote sensing and geological data) for copper mineralization using the BPNN algorithm; (3) to fuse remote sensing layers (i.e., hydrothermal alteration zones) with geological information layers (i.e., lithology, structural geology, geochemistry and geochronology) using Fuzzy-AHP

method; (4) to generate an accurate potential mineral map of VMS copper mineralization for the Sahlabad mining area.

2. Geological Setting

The Sistan Suture Zone (SSZ) is situated in the eastern part of Iran, 800 km in length (N-S) and 200 km in width (E-W) (Figure 1A); it was formed during the Paleogene collision of the Central Iran Block (CIB) with the Afghan Block (AB) [24–26]. It splits the continental Lut sub-block of CIB to the west from the AB to the east [27]. This belt consisting of peridotites, serpentinites, gabbros, and leucogabbros, dolerites, basalts, and radiolarites represents remnants of the lithosphere of the Sistan oceanic basin and its pelagic sedimentary cover [28–30]. The Sahlabad mining area is situated in the SSZ and is bounded between longitudes 59°30′ to 60° and 32° to 32°30′ (Figure 1A,B). It is located in the flysch and colored mélange belt of the SSZ and consists of igneous, metamorphic and sedimentary lithological units that are shown in Figure 1B, comprehensively. In the Sahlabad mining area, there are three main VMS copper deposits, namely (1) Mesgaran, (2) Chah-Rasteh, and (3) Zahri [23]. Moreover, there are also some copper indices and two abandoned old copper mines in this area. Information about the geographical location, alterations, host lithology and Cu minerals for the deposits, old mines, and indices are represented in Table 1.
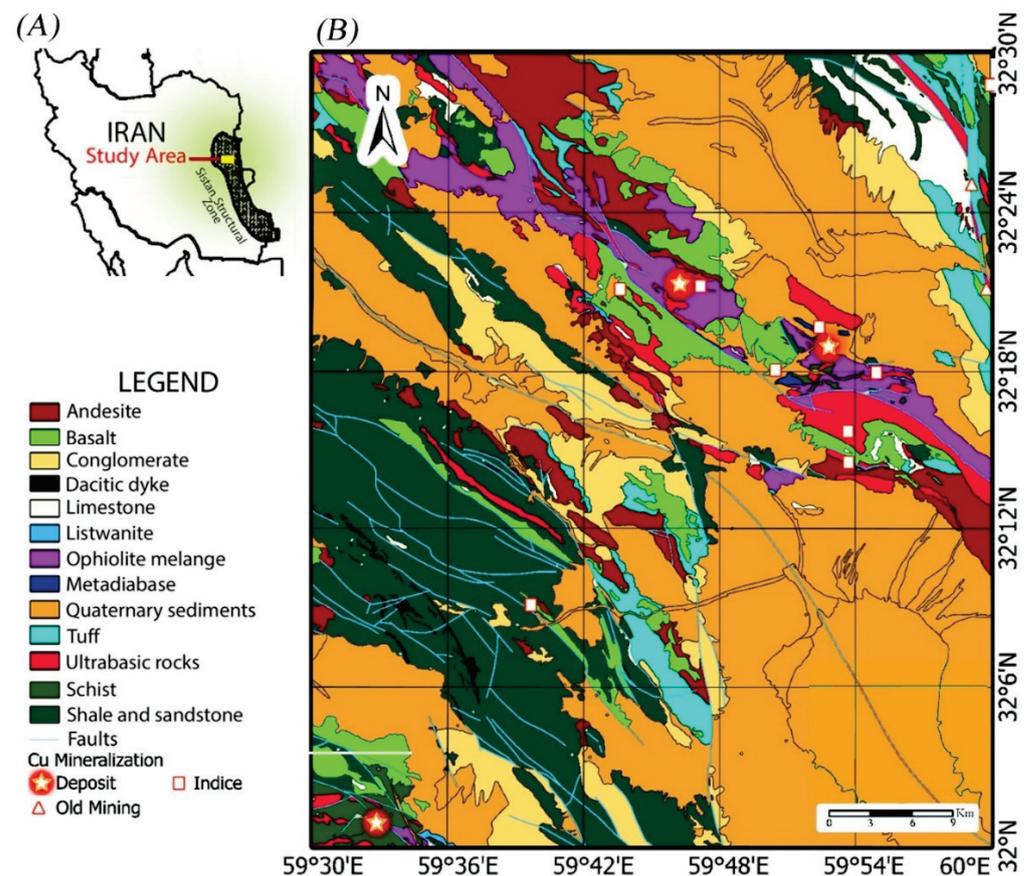
Figure 1. (**A**) Geographical location of Sistan Structural Zone (SSZ) and the study area in Iran. (**B**) Detailed geology map of the Sahlabad mining Area (modified from [23]).

Table 1. Copper mineralization, host-rock lithologies and alteration zones in Sahlabad mining area [31]. Abbreviations: Cpy = Chalcopyrite, Py = Pyrite, Mch = Malachite, Ch = Chalcocite, Az = Azorite, Ba = Basalt, Phy = Phyllic, An = Andesite, Sch = Schist, Mtd = Metadiabase, Chl = Chlorite Alteration, Db = Diabase, Anb = Andesite-Basalt, Qtz = Quartz Alteration, Cab = Carbonate Alteration, Pp = Propylitic Alteration, Arg = Argillic Alteration, Ub = Ultrabasic, Alteration, Sep = Serpentine Alteration, Hem = Hematite Alteration, Lm = Limonite Alteration, Goe = Goethite Alteration.

Copper Occurrences	Center Coordinates		Anomaly Area (Km^2)	Mineralization	Alterations	Host-Rock Lithologies
	Longitude (E)	Latitude (N)				
Mesgaran Deposit	59°52′49″	32°18′58″	8	Cpy + Mch	Phy + Arg + Pp + Chl + Qtz	Ba + Anb
Chah-Rasteh Deposit	59°46′15″	32°21′19″	4	Ch + Mch	Phy + Arg + Pp + Chl + Cab	An + Anb
Zahri Deposit	59°32′52″	32°00′50″	2	Cpy + Ch + Mch	Phy + Arg + Pp + Hem	Ub + Sch
Kasrab Abandoned Mine	59°59′45″	32°21′05″	3.8	Mch	Phy + Arg + Pp + Sep	Ub
Cheshme-Zangi Abandoned Mine	59°59′08″	32°25′02″	2.5	Cpy + Mch	Phy + Arg + Pp + Silicification	Limestone shale + Listwanite
Shir-Shotor Indice	59°53′50″	32°14′28″	1	Mch + Az	Arg + Pp + Sep	An + Serpentinite (Ub)
Dastgerd Indice	59°43′39″	32°21′03″	2	Mch	Arg + Pp + Sep + Hem	Harzburgite
Torshaab Indice	59°59′56″	32°28′48″	5	Mch + Az	Phy + Arg + Pp + Hem + Lm	Sch
Chah-Anjir Indice	59°53′37″	32°15′44″	2	Mch + Az	Pp + Sep	Serpentinite (Ub)
Zargaran Indice	59°47′09″	32°21′14″	1	Mch + Az	Phy + Arg + Pp + Lm + Goe + Hem	An + Db
West Mesgaran Indice	59°52′26″	32°19′36″	1.5	Cpy + Mch + Az	Arg + Pp + Hem + Lm	Mtd
Mirsimin Indice	59°54′58″	32°17′53″	9	Cpy + Mch + Az	Arg + Pp + Hem	Db
Kuharod Indice	59°50′31″	32°18′01″	1	Mch	Phy + Arg + Pp + Hem	Db
Barghan Indice	59°39′38″	32°09′05″	2	Mch	Arg + Pp + Lm + Geo + Hem	Db + Limestone

3. Materials and Methods

3.1. Information Layers

Several information layers were considered and analyzed to map high potential zones of VMS copper mineralization in the Sahlabad mining area. The information layers containing significant information for VMS copper mineralization in the study area were mainly selected from remote sensing and geological datasets. ASTER remote sensing data were processed using band ratio and Selective Principal Components Analysis (SPCA) techniques [20–22] to map hydrothermal alteration zones such as the gossan, argillic, phyllic and propylitic associated with massive sulfide copper mineralization in the study area. The geological information is typically derived from lithological, geochronological, geochemical and structural information that is documented (as digital maps) for the study area [32]. Subsequently, information layers were evaluated using the BPNN algorithm and fused using the Fuzzy-AHP method to generate a potential map of VMS copper mineralization for the Sahlabad mining area. Figure 2 shows the methodological flowchart implemented in this study.
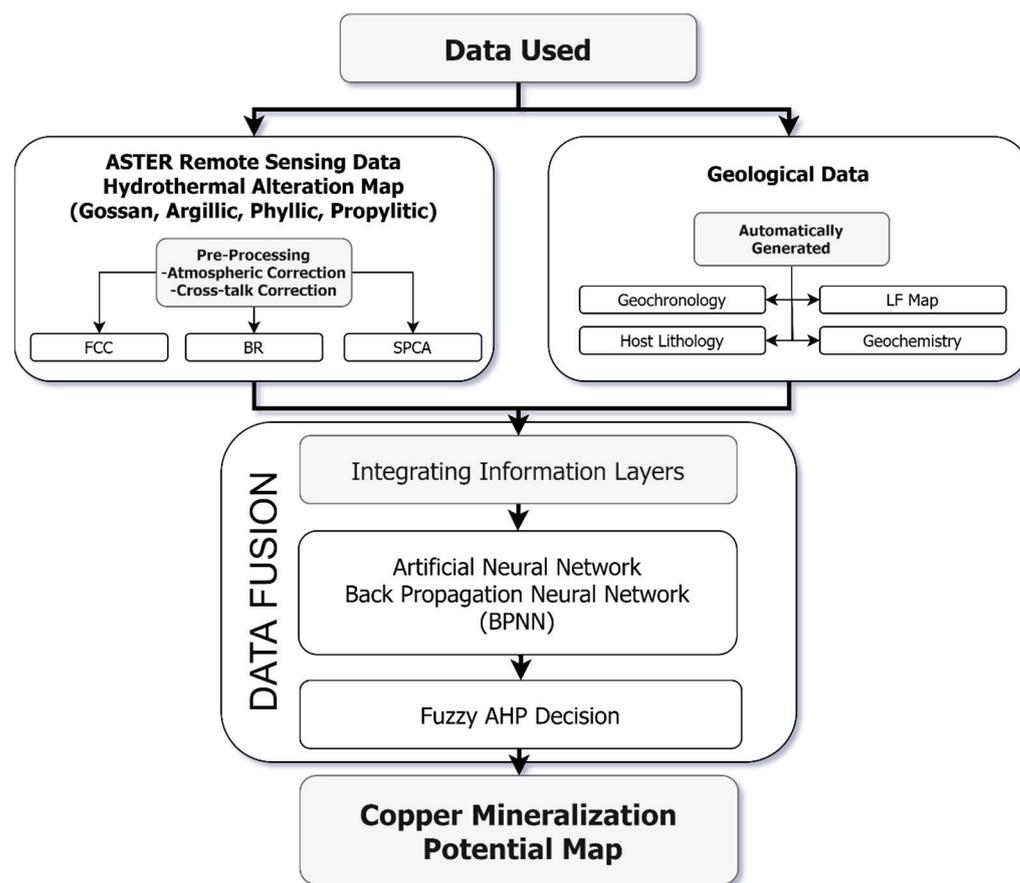
Figure 2. The methodological flowchart implemented in this study.

3.2. Remote sensing Data Characteristics and Processing

Three ASTER scenes covering the study area acquired on 15 July 2002 were used in this study. They are level 1B product and cloud-free and were obtained from the USGS Earth Explorer (http://earthexplorer.usgs.gov, accessed on 1 June 2021) website. The data were pre-georeferenced to UTM zone 40 North projection using WGS-84 datum. ASTER level 1B data were mosaic and preprocessed using the Cross-Talk correction [33]. Also, atmospheric correction was executed by Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm to VNIR and SWIR subsystems [34]. Band ratio [35] and Selective Principal Components Analysis (SPCA) [36] were applied to VNIR+SWIR bands (1 to 9) for mapping gossan, argillic, phyllic and propylitic alteration zones. The ENVI (Environment for Visualizing Images, http://www.exelisvis.com, accessed on 1 June 2021) version 5.2 and ArcGIS version 10.3 software (Esri, Redlands, CA, USA) packages were used to process the remote sensing datasets.

3.3. Geological Data

The main host-rocks of copper mineralization are typically andesite, andesite-basalt and basalt rocks [23,32,37]. However, other geological units such as ultrabasic unit, ophiolite mélange and schist could host the copper mineralization in the study area due to structural controls of ore mineralization [23,31]. These lithological units were used as sub-criteria for the geology information layer and decision-making process. Furthermore, the geological ages were considered sub-criteria of the geochronological information layer. The geological ages of the lithological units are shown in Table 2.

The relationship between the fault systems and copper mineralization in the study area is documented [31]. Lithological trends as well as spatial distribution of copper occurrences are related to structural features and the fault systems acted as a controller

of the host lithology trend. As a result, the Lineament Factor (LF) map of the study area can be considered as one of the most important keys for identifying the potential zones of copper mineralization. The LF map shows three important factors, including (i) frequency of faults, (ii) frequency of fault intersection and (iii) fault length. The LF map scores the factors related to the faults based on the grade of significance and ultimately shows the areas that are important for fault activity. Initially, the network of the Sahlabad area was divided into 100-square-meter cells to study the faults and generate an LF map using the RockWorks software package (Version 17, RockWare, Golden City, CO, USA). The scores of these factors were considered from top to bottom, 1, 2 and 3, respectively [31]. In this analysis, high intensity areas (LF more than 30) in the LF map were also considered as one of the sub-criteria in identifying copper mineralization. The results of applying factor analysis method on the geochemical data of stream sediments (706 samples) showed that copper is among the first principal factors with 27% variance justification [38]. The elements associated with Cu among the first principal components are Pb, Zn, Sn, Ag and Mo. For this analysis, the geochemical map of the study area is produced based on the scores of Cu, Pb, Zn, Sn, Ag and Mo and the kriging interpolation. These anomalies were also considered as sub-criteria in the geochemical information layer and consequence decision-making process.

Table 2. Geological ages of host lithological units related to copper occurrences in the study area.

| Lithological Unit	Geological Age
Andesite (An) Schist (Sch)	Paleogene
Basalt (Ba) Ophiolite Mélange (Ml) Ultrabasic (Ub)	Upper Cretaceous

3.4. Data Fusion

3.4.1. Back Propagation Neural Network (BPNN)

Back Propagation Neural Network (BPNN) is an algorithm for neural network supervised learning using a reduction gradient. In this method, for an Artificial Neural Network (ANN) and a specific error function, the gradient of the error function relative to the weights of the neural network is calculated [9–13]. The purpose of applying the BPNN algorithm is estimation and validation of the results of the analysis of information layers related to copper mineralization based on copper deposits, old mines, and mineral indices in the Sahlabad mining area. In other words: provide the answer to the question of whether the exploration factors studied so far have sufficient credibility to continue the decision-making process and identify copper mineralization in the area or not. At this stage, if the input factors can be estimated with high accuracy based on each other, the information layers are used as input for AHP-Fuzzy method.

In the BPNN algorithm, parameters such as the type of training, the choice of the number of neurons in different layers and the type of neurons are important [11]. In the obtained results, the type of training is based on binary rules and the network efficiency was estimated using the mean squared error. The type and number of neurons as well as the type of activation functions are shown in Figure 3.

The general structure of the artificial neural network presented in Figure 3 contains two hidden layers, the input layer and the output layer. The first and second hidden layers have two neurons and one neuron, respectively. The activation functions used in the first and second layers are the linear function and the sigmoid logistic function, respectively. This choice was made due to the appropriate results in the mineral exploration data [39,40]. There are 14 copper occurrences in the Sahlabad mining area, including copper deposits, old mines and mineral indices. Therefore, the economic mineralization certainty of the points in the area was defined as the BPNN input algorithm with probability percentages

of 100% for the deposits, 70% for the old mines, and 50% for the mineral indices. Then, information about each copper mineralization point in the area was used as a BPNN input. Table 3 lists the inputs for the BPNN algorithm, which are the exploration factors used to generate information layers.
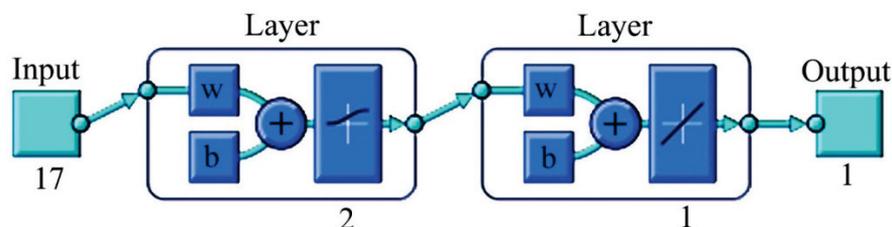
Figure 3. Schematic view of the BPNN structure used in this analysis. (Numbers are neurons in the layers, W: Weight, b: bias value, +: combination).

Table 3. Factors of the copper occurrences in the Sahlabad mining area used as the input for the BPNN algorithm.

Lithology	Geochronology	Alterations	Dominant Mineralization	Structural Geology
An	Upper Cretaceous	Argillic	Malachite	
Ba		Phyllic	Azurite	
Ub		Propylitic	Chalcopyrite	Distance from high intensity LF community
Ml	Paleogene			
Sch		Iron Oxides	Chalcocite	

3.4.2. Hybrid Fuzzy-Analytic Hierarchy Process (Fuzzy-AHP) Method

Fuzzy-AHP method was used as a knowledge-based method. In fact, the process used is a decision-making method based on priorities [41–43]. Each of the information layers, including maps of geology, geochronology, geochemistry, structural geology, and hydrothermal alterations were weighted. Subsequently, all the fuzzified information layers were fused based on the assigned weights, and the map of the copper mineralization potential was produced. To validate the model obtained from the Fuzzy-AHP method, copper deposits, old mines and indices in the area were positioned on the map. Generally, the processing is comprised of four main steps [44,45]: (i) criteria and sub-criteria determination to use in modeling; (ii) criteria and sub-criteria weight calculation; (iii) information layers fuzzification; and (iv) final integration of information layers based on the calculated weights.

Based on the available data and parameters and the exploration target (identification of copper mineralization), the main criteria and sub-criteria were determined. The hierarchical structure of the criteria and sub-criteria determined based on the exploratory layers is presented in Figure 4. A total of five main criteria and eleven sub-criteria were determined for Fuzzy-AHP modeling in this study. It should be noted that all coding and analysis related to data fusion was done in the MATLAB software (R2020b, Portola Valley, CA, USA) environment.
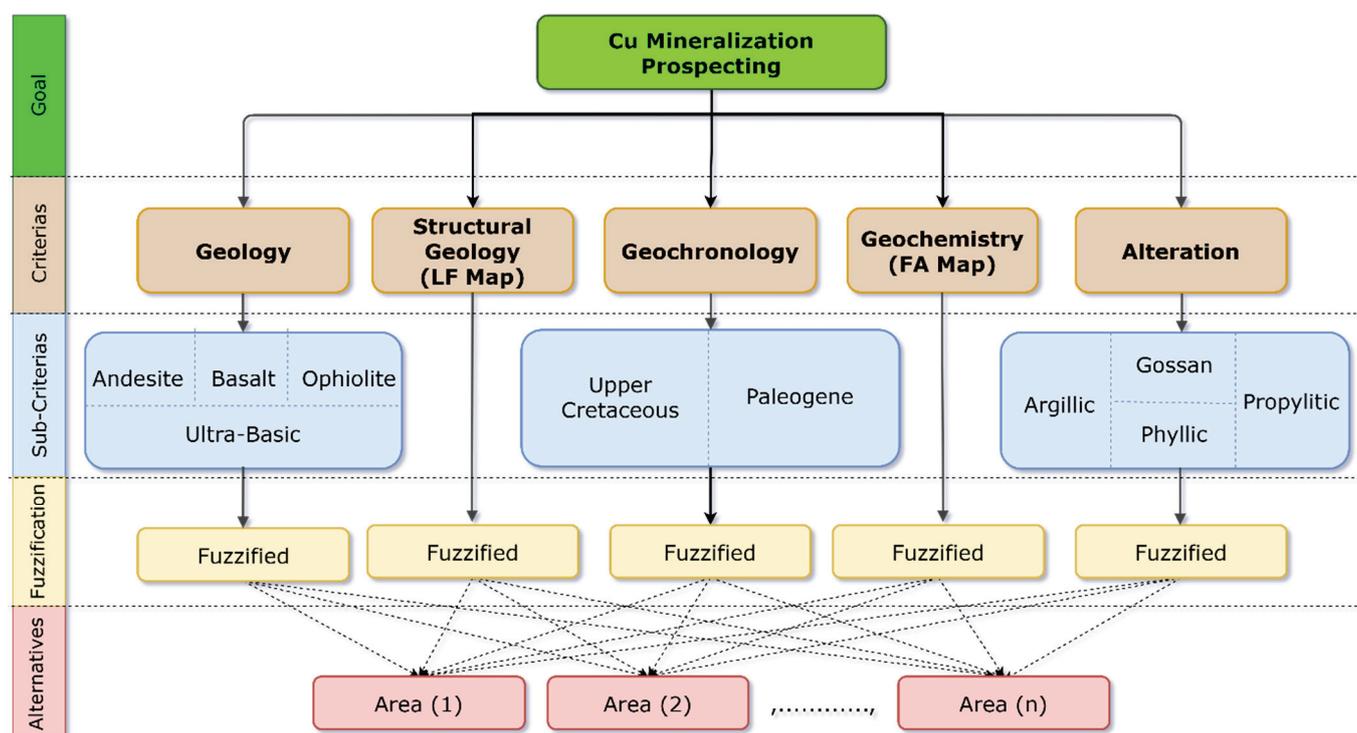
Figure 4. Hierarchical structure for integrated modeling of copper mineralization in the Sahlabad mining area based on the Fuzzy-AHP method.

4. Results

4.1. Alteration Mapping Using ASTER Data

An overview of the lithological units and alteration zones in the Sahlabad mining area was shown using false color composites (FCC) of bands 4 (red), 6 (green), and (blue) 8 of ASTER (Figure 5). The selected bands of ASTER have corresponded to the reflection (1.60–1.70 μm: band 4) and absorption properties (2.185–2.225 μm: band 6 and 2.295–2365 μm: band 8) related to Al-OH and Mg-Fe-OH mineral assemblages [20,46]. The FCC helps to distinguish some of the main lithological units such as ultrabasic rocks and ophiolite mélange (black ton), andesite (grey to brown hue), shale and sandstone (gray to brown shade), tuff (dark grey color) and limestone (whitish pink shade). Moreover, the altered zones might be depicted in pink to magenta color due to high reflectance of OH-bearing minerals in band 4 of ASTER [20,46].

The phyllic and argillic alteration zones have strong Al-OH adsorption properties (illite, montmorillonite, kaolinite, alunite, and muscovite) and can be represented as yellowish-pink shade (Figure 5). The propylitic alteration regions might be mostly manifested in green tone (Figure 5) due to the absorption properties of Fe-Mg-OH (chlorite and epidote) in band 8 [47]. In the southeastern part of the study area, Quaternary alluvium appeared in a red to dark color (Figure 5) attributed to high abundance of clay minerals [47]. Considering the geological map of the study area, the rocks that have been subjected to phyllic and argillic alteration are generally associated with felsic to moderate igneous units (i.e., granodiorite and andesite) and sedimentary rocks such as sandstone and conglomerate. Altered propylitic rocks in the study area are generally associated with mafic igneous units such as basalt, ultrabasic units, andesite-basalt and andesitic lavas.

For detailed mapping of hydrothermal alteration minerals, band ratios of 4/2 (to detect iron oxide/hydroxides: gossan), 5/6 (kaolinite and alunite: argillic alteration), 7/6 (muscovite and jarosite: phyllic alteration) and 9/8 (chlorite, epidote and calcite: propylitic alteration) [19,42,43] were assigned and implemented (Figure 6A–D). Subsequently, the alteration zones were mapped in the study area. The results show that moderate to high surface distribution of gossan is typically detected in ophiolite mélange, basalt,

shale and sandstone, limestone, schist, and conglomerate (Figure 6A). Argillic alteration is widespread in the study area, which is mainly associated with shale and sandstone, ophiolite mélange, andesitic, ultrabasic rocks and conglomerate (Figure 6B). The phyllic alteration zones show moderate to high surface distribution in many parts of the conglomerate, dacitic dyke, limestone and Quaternary sediments (Figure 6C). Moderate to high surface distribution of propylitic alteration was mapped associated with ultrabasic rocks, basalt, andesite, ophiolite mélange shale and sandstone (Figure 6D). Moderate to high abundance of the alteration zones, especially propylitic, gossan and argillic zones were found with copper deposits, old mines, and indices (Figure 6A–D).
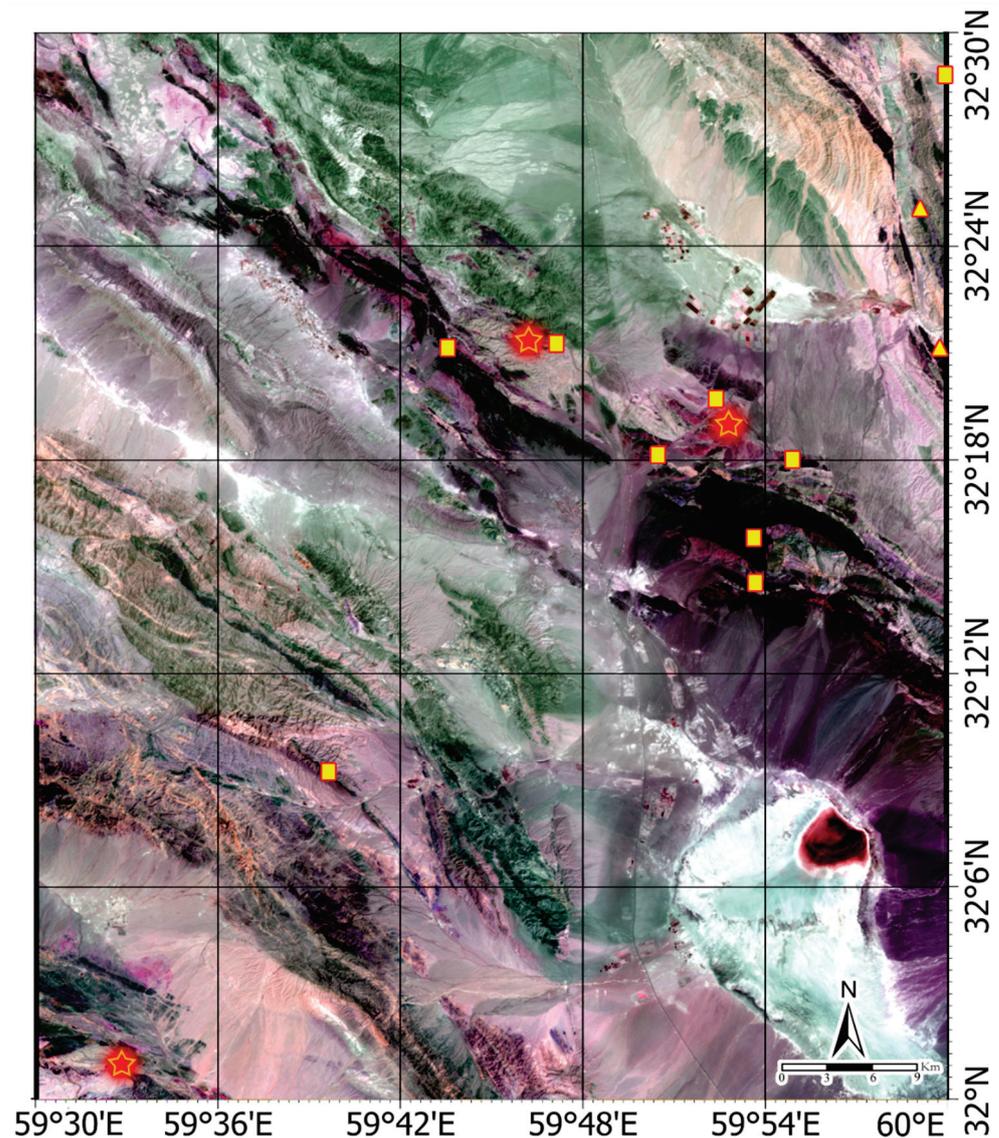
Figure 5. False Color Combination (FCC) of bands 4, 8 and 6 as RGB color composite for Sahlabad mining area. The location of copper deposits, old mines and indices are shown with red stars and yellow triangles and cubes, respectively.

Additionally, the selective principal components analysis (SPCA) method was used to detect the spatial distribution of gossan, argillic, phyllic and propylitic alteration regions in this study. Bands 1, 2, 3 and 4 were selected to map gossan. Analyzing the eigenvector matrix of the selected bands shows the PC2 contains strong loading in band 2 (0.586388) and band 4 (−0.908873) with opposite signs (Table 4). The positive loadings of the eigenvectors in the reflective bands (band 4) show the alterations in the form of bright pixels, while the

negative loadings display the alterations in the dark pixels [36]. Therefore, iron oxides (gossan) can be mapped as dark pixels in the PC2 image. This image is negated by multiplication to −1 to show iron oxides (gossan) in bright pixels. Figure 7A shows pseudo-color ramp of PC2 image. Moderate to high spatial distribution of iron oxides (gossan) is mapped in andesite, ophiolite mélange, basalt, shale and sandstone, schist, limestone and conglomerate. Results are almost similar to band ratio of 4/2, although some parts of conglomerate and limestone show high surface distribution of iron oxides in the southwestern and northeastern of the study area (Figure 7A).
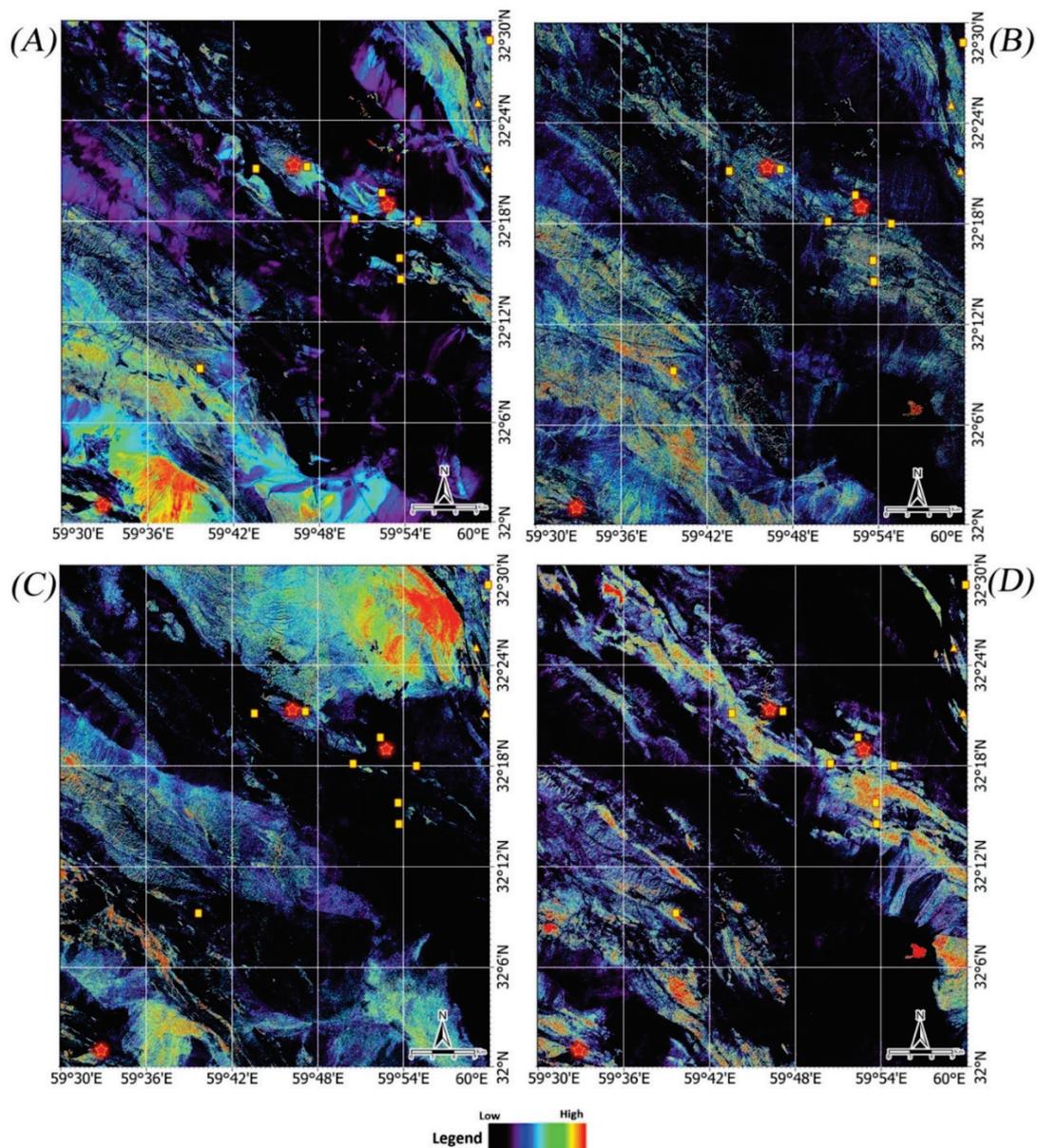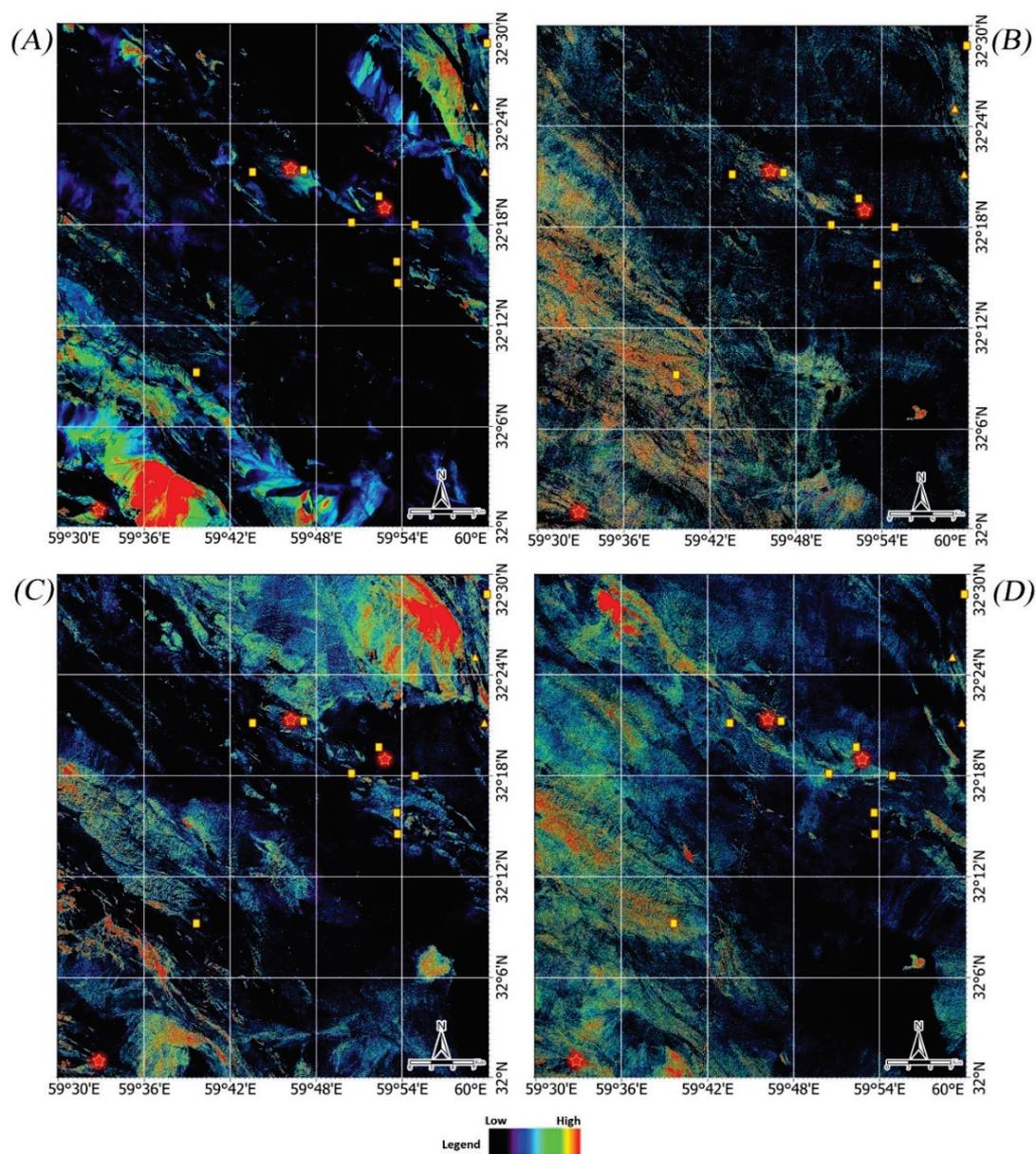

Figure 6. Pseudo-color ramp of band ratio results. (**A**) Band ratio of 4/2 shows the surface distribution of iron oxide/hydroxides; (**B**) band ratio of 5/6 shows the surface distribution of argillic alteration; (**C**) band ratio of 7/6 shows the surface distribution of phyllic alteration; (**D**) band ratio of 9/8 shows the surface distribution of propylitic alteration.

Considering the eigenvector matrix of bands 1, 4, 5 and 6 for mapping argillic zone (Table 5), the PC4 has a strong contribution of band 5 (0.777935) and band 6 (−0.595506) with reverse signs. Argillic alteration zone (kaolinite and alunite) displays absorption in band 5 (2.145–2.185 μm) [20,46,47]. Thus, argillic alteration can be mapped as bright pixels

in the PC4 image. A pseudo-color ramp of image PC4 is shown in Figure 7B. Argillic alteration (strong to moderate) is mainly mapped in the shale and sandstone, ophiolite mélange, ultrabasic, some part andesitic units and conglomerate and Quaternary deposits. The results are almost identical with the band ratio of 5/6 (see Figure 6B).

Table 4. Eigenvector matrix of selected bands for detecting iron oxides (gossan) derived from SPCA.

Eigenvector	Band 1	Band 2	Band 3	Band 4
PC 1	0.546195	0.644888	0.406994	0.346621
PC 2	0.372506	0.586388	−0.021196	−0.908873
PC 3	0.642726	−0.232160	−0.692243	0.231958
PC 4	−0.387056	0.703904	−0.595573	−0.000394

Figure 7. Pseudo-color ramp of SPCA results. (**A**) PC2 shows the surface distribution of iron oxide/hydroxides; (**B**) PC4 shows the surface distribution of argillic alteration; (**C**) PC4 shows the surface distribution of phyllic alteration; (**D**) PC4 shows the surface distribution of propylitic alteration.

Table 5. Eigenvector matrix of selected bands for mapping argillic zone derived from SPCA.

Eigenvector	Band 1	Band 4	Band 5	Band 6
PC 1	0.590462	0.465558	0.452265	0.479653
PC 2	0.805201	−0.296067	−0.381717	−0.343930
PC 3	0.034781	−0.810683	0.211106	0.544994
PC 4	0.042383	−0.195941	0.777935	−0.595506

Looking at the eigenvector matrix of bands 1, 4, 6 and 7 for identifying the phyllic zone (Table 6), it seems that the PC4 contains information for mapping this alteration zone. Strong loading in band 6 (−0.752127) with a negative sign and strong positive loading in band 6 (0.652273) is presented for the PC4. Phyllic alteration exhibits absorption in band 7 (2.235–2.285 µm) and reflectance in band 6 (2.185–2.225 µm) [19]. Accordingly, the phyllic zone will appear as dark pixels in the PC4 image. This image is negated (by multiplication to −1) for converting the dark to bright pixels. Strong spatial distribution of phyllic zone is associated with dacitic dyke, conglomerate, limestone, basalt, tuff, andesite and quaternary sediments (Figure 7C). The SPCA results for mapping phyllic zone are matched to band ratio of 7/6 (see Figure 6C).

Table 6. Eigenvector matrix of selected bands for mapping phyllic zone derived from SPCA.

Eigenvector	Band 1	Band 4	Band 6	Band 7
PC 1	−0.583430	−0.457601	−0.472606	−0.476292
PC 2	−0.811464	0.294887	0.349311	0.364074
PC 3	−0.033560	0.833551	−0.298210	−0.463828
PC 4	0.003041	0.093998	−0.752127	0.652273

Analyzing the eigenvector matrix of bands 1, 4, 8 and 9 is considered for detecting propylitic zone (Table 7). The PC4 has strong loading in band 8 (−0.705894) with a negative sign and strong loading in band 9 (0.707293) with a positive sign. The propylitic zone is characterized by absorption features of Fe, Mg-OH and CO_3 [48]. The absorption features are situated in band 8 (2.295–2.365 µm) [47]. Thus, propylitic zone can be mapped as bright pixels in the PC4 image with considering band 9 as reflectance band. Propylitic alteration is detected in ophiolite mélange, shale and sandstone, ultrabasic rocks, basalt, andesite and some parts of conglomerate, limestone, and Quaternary sediments (Figure 7D). The identified propylitic zone is almost identical with the band ratio (9/8) results, however, SPCA shows strong surface distribution in northwestern and western parts of the study area. Propylitic, gossan and argillic zones are typically mapped in the location of copper deposits, old mines, and indices in the study area (see Figures 6 and 7).

Table 7. Eigenvector matrix of selected bands for mapping propylitic zone derived from SPCA.

Eigenvector	Band 1	Band 4	Band 8	Band 9
PC 1	−0.545145	−0.432945	−0.525490	−0.489117
PC 2	−0.830444	0.168516	0.355665	0.394294
PC 3	−0.109009	0.885442	−0.314765	−0.324087
PC 4	0.036012	0.012378	−0.705894	0.707293

4.2. Geological Maps Derived from the Geochronology, Structures and Geochemistry

The host-rock of the known VMS copper deposits in the study area, namely Mesgaran, Chah-Raste and Zahri are documented as andesite, andesite-basalt and basalt units. Ultrabasic unit, ophiolite melange and schist are also related to structurally-controlled copper mineralization in the study area [20,31]. The geological ages of the above mentioned lithologies are Upper Cretaceous and Paleogene (see Tables 2 and 3). A geochronological map of the study area was generated (Figure 8). This map shows the spatial distribution of

lithological units for the geological ages of Neogene, Upper Cretaceous, Quaternary and Paleogene. Copper deposits, old mines, and indices in the study area are mostly hosted in Upper Cretaceous and Paleogene units (Figure 8). The three main copper deposits are placed in Upper Cretaceous units. However, old mines and some of the indices are situated in Paleogene units (see Figure 8).
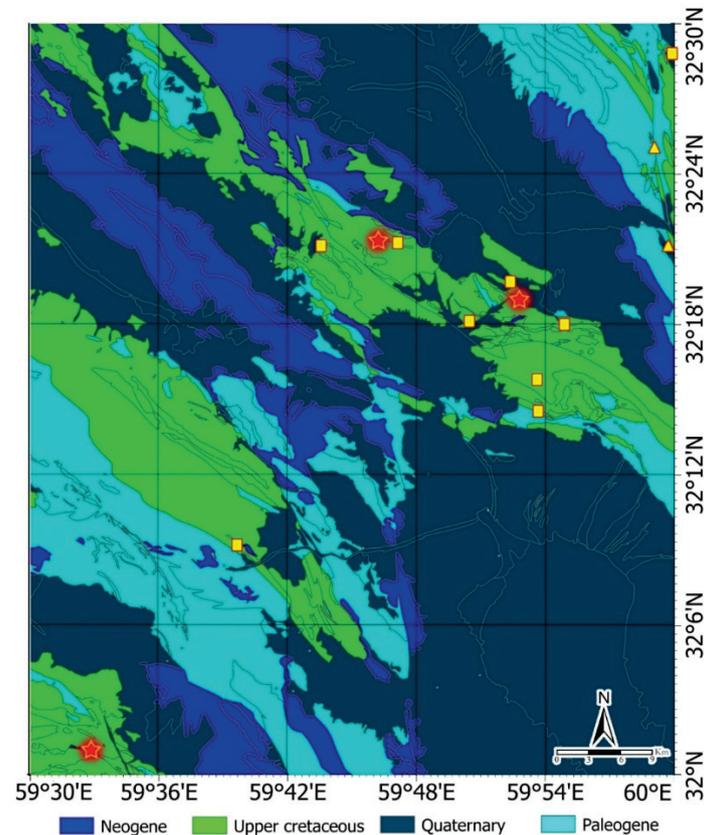
Figure 8. Geochronological map of the study area.

In the study area, fault systems and copper mineralization have a close relationship. The fault system controlled the trend of the host lithology of copper mineralization [31]. Previous research in the study area confirmed that the high intensity areas (LF more than 30) is one of the most important factors in identifying copper mineralization [31]. Accordingly, the Lineament Factor (LF) map of the study area was adopted herein (Figure 9). The NW-SE fault systems are controlling the lithology of the host-rock for copper mineralization. The NW-SE fault systems are consistent with the main trend of lithological units related to copper deposits, old mines, and indices in the area (see Figure 9). Figure 10 shows the geochemical map for geochemical family of copper (Pb, Zn, Sn, Ag and Mo) in the study area. The elements Pb, Zn, Sn, Ag and Mo were considered as trace elements and predictor composition of copper mineralization in the study area. Hence, the group behavior was investigated to clarify the geochemical relationships of trace elements with the Cu. Copper deposits, old mines, and indices in the study area are mainly associated with moderate to high anomaly zones (Figure 10).
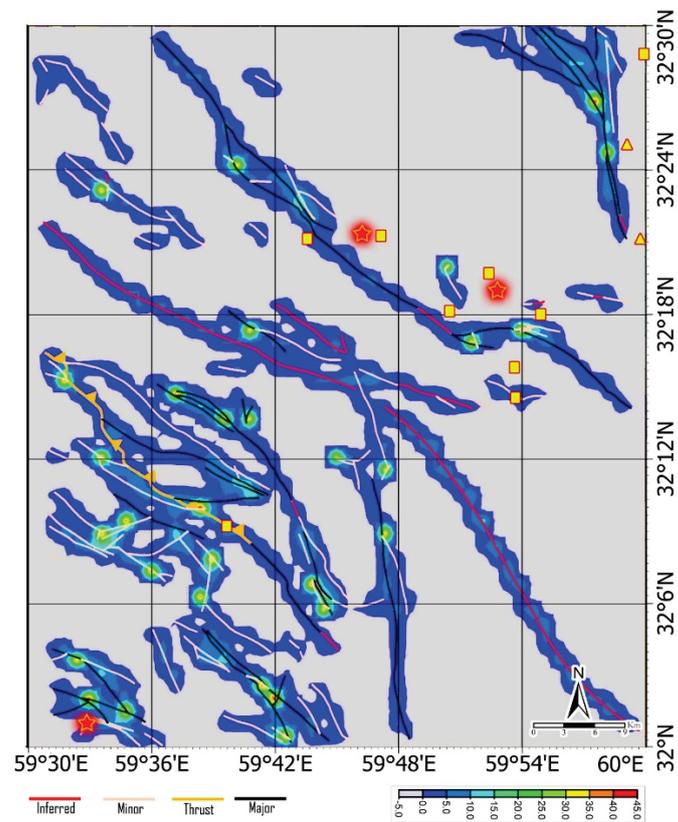
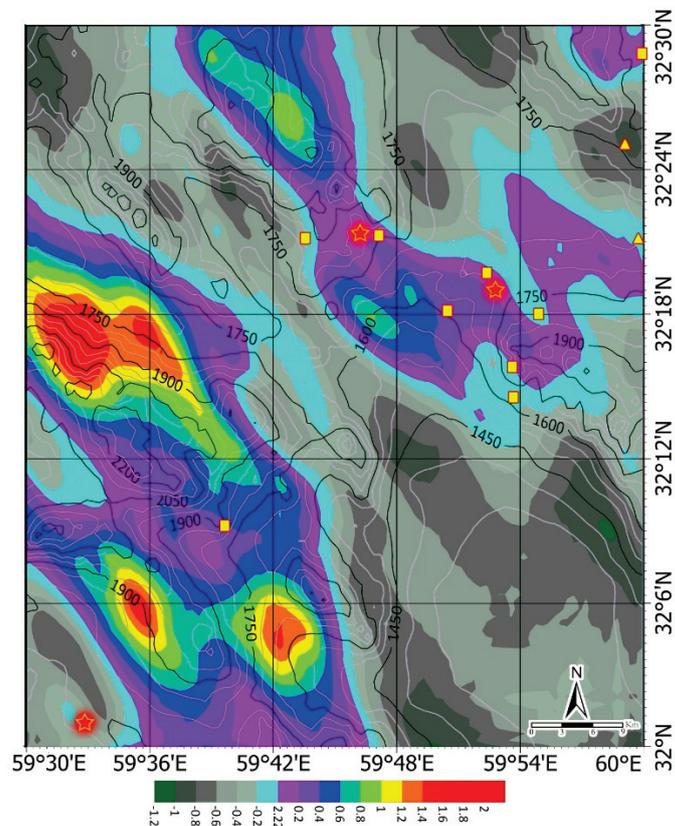**Figure 9.** Lineament Factor (LF) map of the study area.

Figure 10. Geochemical map of high potential areas based on geochemical family of copper (Pb, Zn, Sn, Ag and Mo).

4.3. Fusion of Information Layers

In the BPNN algorithm, the training data (see Table 3) should be divided into two categories (30% to 70%). For this purpose, the training data were randomly divided into two groups and 70% of the group entered the artificial neural network. The result of the estimated line of copper mineralization with real points in the training data and test data is presented in Figure 11A,B and the accuracy of the training data and test data was estimated (Figure 11C,D). According to the results, the accuracy value for training data and test data is over 99%. This result indicates a high performance of the BPNN algorithm in predicting information about copper mineralization in the Sahlabad mining area. In other words, the information obtained from copper exploration studies in the area is practically 99% reliable and is sufficiently fit to be used in the decision-making process and the integration of information layers.
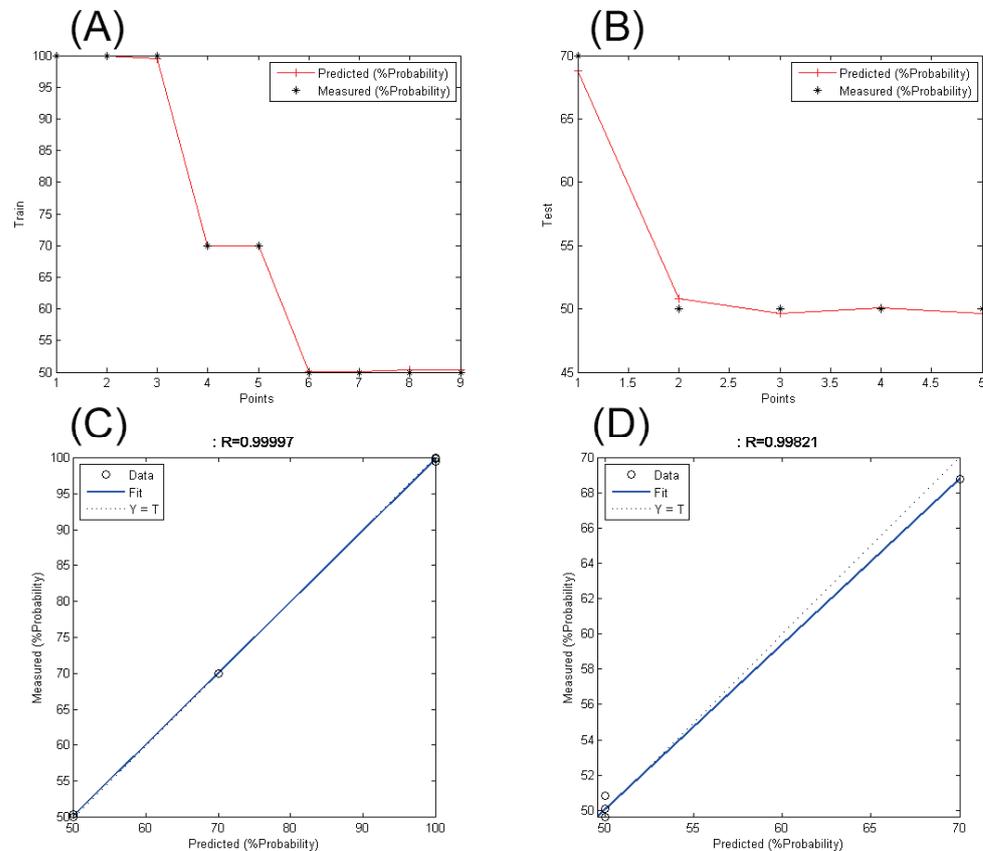
Figure 11. (**A**) Copper occurrences estimation line based on copper occurrences in training data; (**B**) copper occurrences estimation line based on copper occurrences in test data; (**C**) regression of estimated data versus training data; (**D**) regression of estimated data versus test data.

After determining the main criteria and sub-criteria in the Fuzzy-AHP method (see Figure 4), their weights were calculated successively. A paired comparison matrix was formed for the main criteria and sub-criteria and based on expert opinions and prioritization table [45,46]. The main criteria and sub-criteria weights were calculated by paired comparison. Paired comparison matrices for the main criteria and sub-criteria are presented in Tables 8–11, respectively. According to the calculations performed for each of the above paired comparison matrices, the Inconsistency Rate (IR) (%) was obtained for the main criteria and sub-criteria matrices (Table 12). Given that all Inconsistency Rates (IRs) are less than 10%, it can be said that the consistency of the expert judgment is accepted [45]. At information layers fuzzification stage, each information layer, which was defined as the main criteria, was fuzzified based on fuzzy logic. Fuzzy maps are shown in Figure 12.

Table 8. Paired comparison matrix and calculated weights for each of the main criteria by AHP method.

Criteria	Geology	Geochronology	Structural Geology	Geochemistry	Alterations	Importance Rank	Weight (%)
Geology	1	4	0.2	3	0.25	3	13.3
Geochronology	0.25	1	0.14	0.2	0.14	5	3.6
Structural Geology	5	7	1	5	2	1	44.6
Geochemistry	0.33	5	0.2	1	0.33	4	9.3
Alterations	4	7	0.5	3	1	2	29.2

Table 9. Paired comparison matrix and calculated weights for sub-criteria of geology criteria by AHP method.

Sub-Criteria of Geology	An	Ba	Ml	Ub	Importance Rank	Weight (%)
An	1	1	0.33	0.2	3	9.9
Ba	1	1	0.33	0.2	3	9.9
Ml	3	3	1	0.5	2	28.4
Ub	5	5	2	1	1	51.8

Table 10. Paired comparison matrix and calculated weights for sub-criteria of geochronology criteria by AHP method.

Sub-Criteria of Geochronology	Upper Cretaceous	Paleogene	Importance Rank	Weight (%)
Upper Cretaceous	1	2	1	66.7
Paleogene	0.5	1	2	33.3

Table 11. Paired comparison matrix and calculated weights for sub-criteria of alterations criteria by AHP method.

Sub-Criteria of Alterations	Argillic	Phyllic	Propylitic	Importance Rank	Weight (%)
Argillic	1	1	1	1	33.3
Phyllic	1	1	1	1	33.3
Propylitic	1	1	1	1	33.3

Table 12. The inconsistency rate (%) of criteria and sub-criteria paired comparison matrices.

	Inconsistency Rate (%)
Main Criteria	8.9
Sub-criteria of Geology	0.2
Sub-criteria of Geochronology	0
Sub-criteria of Alterations	0

In the fusion of information layers based on the calculated weights step, the fuzzified information layers were integrated using AHP and based on the calculated weights. Consequently, the final fused map/potential map of copper mineralization for the Sahlabad mining area was generated (Figure 13). In the potential map of the study area (Figure 13), copper occurrences in the area, including copper deposits, old mines and mineral indices were plotted to validate the results. The map shows the potential copper mineralization zones (Figure 13). The areas with the highest potential are shown in the black color spectrum and the areas with the lowest potential or barren areas are shown in the white spectrum. As can be seen in Figure 13, all copper occurrences in the Sahlabad mining area are located in areas with high potential copper mineralization. This indicates the high

validity of the results. Some new copper potential zones are also identified in northwestern and southeastern parts of the study area (Figure 13). These zones can be considered for subsequent field campaigns and drilling programs in the Sahlabad mining area.
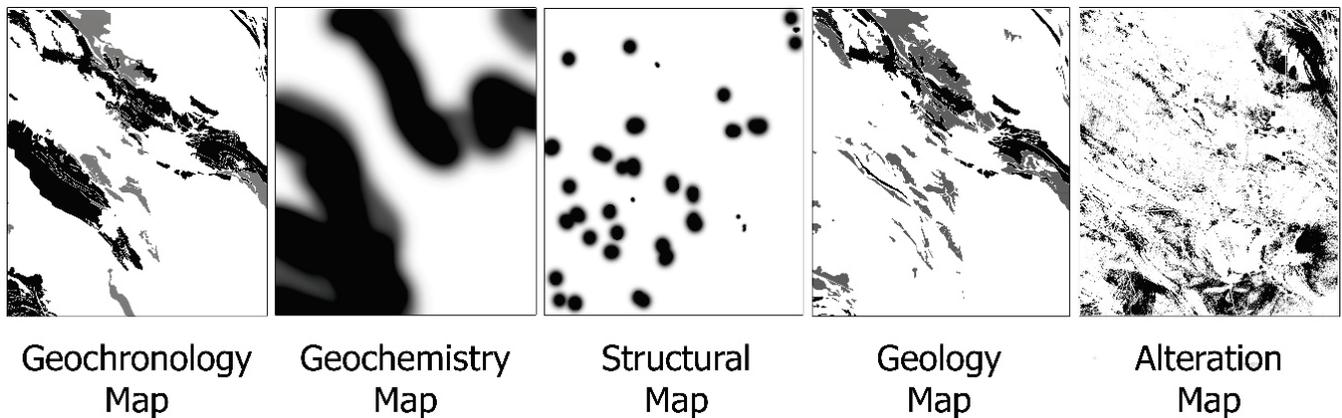
Figure 12. The fuzzified exploratory maps for the Sahlabad mining area.
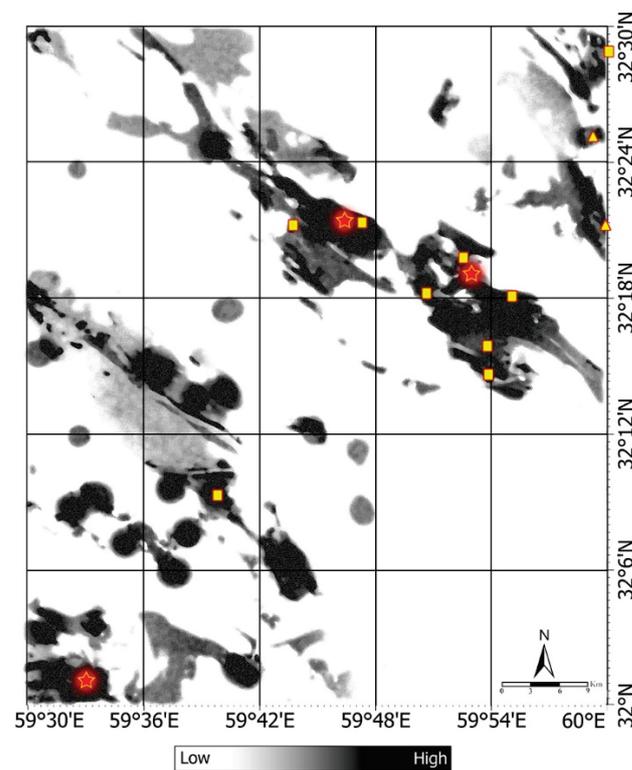
Figure 13. The potential map of copper mineralization for the Sahlabad mining area.

5. Discussion

Identifying potential areas of VMS copper mineralization depends on various parameters such as fault systems, lithological and geochronological units, geochemical anomalies and hydrothermal alteration mineral zones [49–51]. Using information layer integration methods has a great potential to increase the accuracy of the identification of high potential areas. Recently, Artificial Intelligence (AI)-based techniques such as Machine Learning (ML) and Artificial Neural Network (ANN) have been successfully used for mineral exploration [52–55]. In this research, the BPNN method was used before integrating exploratory information layers by Fuzzy-AHP for exploration of massive sulfide copper mineralization in the Sahlabad mining area, east Iran. In fact, the BPNN evaluates the ability to predict

copper occurrences by prospecting parameters that exploratory information layers provided based on them. This integrated technique is called Neuro-Fuzzy-AHP (NFAHP), which is developed in the present study.

To use the proposed BPNN network, exploratory features of each copper occurrence such as host lithology, geochemical anomalies, fault system, ore mineralization, alterations and geochronological units of each copper occurrence should be investigated. The BPNN technique ensures the prediction ability of the information layers before combining them. If the exploratory features cannot predict copper occurrences, they should be removed in the final decision making (Fuzzy-AHP). Another advantage of the combined use of BPNN with Fuzzy-AHP is to help prioritize information layers relative to each other.

In this research, by using the exploratory parameters related to 14 occurrences of massive sulfide copper in the Sahlabad mining area, the prediction of copper mineralization was established with 99% accuracy. Then, by combining exploration layers, the map of copper mineralization potential in the area was prepared using fuzzy-AHP method. After plotting the copper occurrences, as can be seen in the map in Figure 13, all the documented points are located in areas with high potential for massive sulfide copper mineralization. The NFAHP approach presented in this study first evaluates the input data by means of the BPNN algorithm to determine whether the mineralizations in question can be estimated based on the factors or not. Then, if the accuracy of the factors is confirmed, it combines the information layers with each other (by Fuzzy-AHP). Therefore, using the NFAHP, any type of mineralization can be prospected, provided that the exploration factors related to target mineralization type, are taken precisely. In fact, using a greater number of information layers and exploration factors, more accurate results will be obtained.

6. Conclusions

To identify the potential areas of VMS copper mineralization in Sahlabad mining area, east of Iran, hydrothermal alterations including argillic, phyllic and propylitic, were identified using band ratios and SPCA that were implemented to ASTER remote sensing data. The integration of faults and the alteration map showed that the faults are one of the controlling factors of the alterations and ore mineralization trend in the study area. Information layers, including alterations, geochemistry, host lithology, LF map and geochronology have been prepared based on exploration factors related to copper occurrences in the study area. Subsequently, using the BPNN, the predictive power of copper occurrences estimation was evaluated based on exploration factors. According to the results, the available exploration factors with an accuracy of 99% were able to estimate copper mineralizations in the Sahlabad mining area. After the validity of the available information had been confirmed by BPNN, the information layers were fused using the Fuzzy-AHP decision method and a copper mineralization potential map was produced. Copper occurrences points were plotted on the final map. In this way, the validation of the final map was examined. As a result, all points were located in high potential areas, which shows the high reliability of this map. It is recommended that the Neuro-Fuzzy-AHP (NFAHP) technique can be considered for mineral exploration in other metallogenic provinces in Iran and other regions around the world.

References

1. Bachri, I.; Hakdaoui, M.; Raji, M.; Teodoro, A.C.; Benbouziane, A. Machine Learning Algorithms for Automatic Lithological Mapping Using Remote Sensing Data: A Case Study from Souk Arbaa Sahel, Sidi Ifni Inlier, Western Anti-Atlas, Morocco. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 248. [CrossRef]
2. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [CrossRef]
3. Shirmard, H.; Farahbakhsh, E.; Heidari, E.; Pour, A.B.; Pradhan, B.; Müller, D.; Chandra, R. A Comparative Study of Convolutional Neural Networks and Conventional Machine Learning Models for Lithological Mapping Using Remote Sensing Data. *Remote Sens.* **2022**, *14*, 819. [CrossRef]
4. Shirmard, H.; Farahbakhsh, E.; Müller, R.D.; Chandra, R. A Review of Machine Learning in Processing Remote Sensing Data for Mineral Exploration. *Remote Sens. Environ.* **2022**, *268*, 112750. [CrossRef]
5. Sagi, O.; Rokach, L. Ensemble Learning: A Survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1249. [CrossRef]
6. Tahmasebi, P.; Kamrava, S.; Bai, T.; Sahimi, M. Machine Learning in Geo-and Environmental Sciences: From Small to Large Scale. *Adv. Water Resour.* **2020**, *142*, 103619. [CrossRef]
7. Holloway, J.; Mengersen, K. Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review. *Remote Sens.* **2018**, *10*, 1365. [CrossRef]
8. Brown, W.M.; Gedeon, T.D.; Groves, D.I.; Barnes, R.G. Artificial Neural Networks: A New Method for Mineral Prospectivity Mapping. *Aust. J. Earth Sci.* **2000**, *47*, 757–770. [CrossRef]
9. Sun, T.; Li, H.; Wu, K.; Chen, F.; Zhu, Z.; Hu, Z. Data-Driven Predictive Modelling of Mineral Prospectivity Using Machine Learning and Deep Learning Methods: A Case Study from Southern Jiangxi Province, China. *Minerals* **2020**, *10*, 102. [CrossRef]
10. Shirazy, A.; Hezarkhani, A.; Timkin, T.; Shirazi, A. Investigation of Magneto-/Radio-Metric Behavior in Order to Identify an Estimator Model Using K-Means Clustering and Artificial Neural Network (Ann)(Iron Ore Deposit, Yazd, Iran). *Minerals* **2021**, *11*, 1304. [CrossRef]
11. Zhang, N.; Zhou, K.; Li, D. Back-Propagation Neural Network and Support Vector Machines for Gold Mineral Prospectivity Mapping in the Hatu Region, Xinjiang, China. *Earth Sci. Inform.* **2018**, *11*, 553–566. [CrossRef]
12. Kakar, S.A.; Sheikh, N.; Naseem, A.; Iqbal, S.; Rehman, A.; Ullah, A.; Ahmad, B.; Ali, H.; Khan, B. Artificial Neural Network Based Weather Prediction Using Back Propagation Technique. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 462–470. [CrossRef]
13. Shahabi, H.; Shirzadi, A.; Ronoud, S.; Asadi, S.; Pham, B.T.; Mansouripour, F.; Geertsema, M.; Clague, J.J.; Bui, D.T. Flash Flood Susceptibility Mapping Using a Novel Deep Learning Model Based on Deep Belief Network, Back Propagation and Genetic Algorithm. *Geosci. Front.* **2021**, *12*, 101100. [CrossRef]
14. Wan, Y.-Q.; Fan, Y.-H.; Jin, M.-S. Application of Hyperspectral Remote Sensing for Supplementary Investigation of Polymetallic Deposits in Huaniushan Ore Region, Northwestern China. *Sci. Rep.* **2021**, *11*, 440. [CrossRef] [PubMed]
15. Nazerian, H.; Shirazy, A.; Shirazi, A.; Hezarkhani, A. Design of an Artificial Neural Network (Bpnn) to Predict the Content of Silicon Oxide (Sio2) Based on the Values of the Rock Main Oxides: Glass Factory Feed Case Study. *Int. J. Sci. Eng. Appl.* **2022**, *2*, 41–44. [CrossRef]
16. Shirazy, A.; Ziaii, M.; Hezarkhani, A. Geochemical Behavior Investigation Based on K-Means and Artificial Neural Network Prediction for Titanium and Zinc, Kivi Region, Iran. *Bull. Tomsk Polytech. Univ.* **2021**, *332*, 113–125.
17. Zadeh, L.A. Fuzzy Sets. In *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi a Zadeh*; World Scientific: Singapore, 1996; pp. 394–432.
18. Saaty, T.L. How to Make a Decision: The Analytic Hierarchy Process. *Eur. J. Oper. Res.* **1990**, *48*, 9–26. [CrossRef]
19. Liu, Y.; Eckert, C.M.; Earl, C. A Review of Fuzzy Ahp Methods for Decision-Making with Subjective Judgements. *Expert Syst. Appl.* **2020**, *161*, 113738. [CrossRef]
20. Pour, A.B.; Hashim, M. The Application of Aster Remote Sensing Data to Porphyry Copper and Epithermal Gold Deposits. *Ore Geol. Rev.* **2012**, *44*, 1–9. [CrossRef]
21. Pour, A.B.; Sekandari, M.; Rahmani, O.; Crispini, L.; Läufer, A.; Park, Y.; Hong, J.K.; Pradhan, B.; Hashim, M.; Hossain, M.S.; et al. Identification of Phyllosilicates in the Antarctic Environment Using Aster Satellite Data: Case Study from the Mesa Range, Campbell and Priestley Glaciers, Northern Victoria Land. *Remote Sens.* **2020**, *13*, 38. [CrossRef]
22. Wambo, J.D.T.; Pour, A.B.; Ganno, S.; Asimow, P.D.; Zoheir, B.; Salles, R.D.R.; Nzenti, J.P.; Pradhan, B.; Muslim, A.M. Identifying High Potential Zones of Gold Mineralization in a Sub-Tropical Region Using Landsat-8 and Aster Remote Sensing Data: A Case Study of the Ngoura-Colomines Goldfield, Eastern Cameroon. *Ore Geol. Rev.* **2020**, *122*, 103530. [CrossRef]
23. Monazami Bagherzade, R. *Economic Geology Report (Hammer Exploration) Sahlabad Sheet (Scale 1:100,000)*; Geological Survey and Mineral Exploration of Northeast Iran: Tehran, Iran, 2001; p. 20.
24. Stoecklin, J. Structural History and Tectonics of Iran: A Review. *AAPG Bull.* **1968**, *52*, 1229–1258.

25. Berberian, M.; King, G.C.P. Towards a Paleogeography and Tectonic Evolution of Iran. *Can. J. Earth Sci.* **1981**, *18*, 210–265. [CrossRef]

26. Tirrul, R.; Bell, I.R.; Griffis, R.J.; Camp, V.E. The Sistan Suture Zone of Eastern Iran. *Geol. Soc. Am. Bull.* **1983**, *94*, 134–150. [CrossRef]

27. Şengör, A.M.C. A New Model for the Late Palaeozoic—Mesozoic Tectonic Evolution of Iran and Implications for Oman. *Geol. Soc. Lond. Spec. Publ.* **1990**, *49*, 797–831. [CrossRef]

28. Stampfli, G.; Borel, G. A Plate Tectonic Model for the Paleozoic and Mesozoic Constrained by Dynamic Plate Boundaries and Restored Synthetic Oceanic Isochrons. *Earth Planet. Sci. Lett.* **2002**, *196*, 17–33. [CrossRef]

29. Camp, V.; Griffis, R. Character, Genesis and Tectonic Setting of Igneous Rocks in the Sistan Suture Zone, Eastern Iran. *Lithos* **1982**, *15*, 221–239. [CrossRef]

30. Moazzen, M.; Modjarrad, M.; Zarrinkoub, M. Mineral Chemistry, Petrogenesis and P–T Conditions of Formation of Harzburgitic Peridotites from South of Birjand, Eastern Iran. *J. Asian Earth Sci.* **2006**, *4*, 56–75. [CrossRef]

31. Shirazi, A.; Hezarkhani, A.; Pour, A.B. Fusion of Lineament Factor (Lf) Map Analysis and Multifractal Technique for Massive Sulfide Copper Exploration: The Sahlabad Area, East Iran. *Minerals* **2022**, *12*, 549. [CrossRef]

32. Shirazi, A. Integration of Intelligent Analytical Methods for Geochemical Modeling and Identification of Exploration Keys for the Sahlabad Applied Research. Ph.D. Thesis, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, 2022.

33. Iwasaki, A.; Tonooka, H. Validation of a Crosstalk Correction Algorithm for Aster/Swir. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2747–2751. [CrossRef]

34. Thome, K.; Palluconi, F.; Takashima, T.; Masuda, K. Atmospheric Correction of Aster. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1119–1211. [CrossRef]

35. Inzana, J.; Kusky, T.; Higgs, G.; Tucker, R. Supervised Classifications of Landsat Tm Band Ratio Images and Landsat Tm Band Ratio Image with Radar for Geological Interpretations of Central Madagascar. *J. Afr. Earth Sci.* **2003**, *37*, 59–72. [CrossRef]

36. Noori, L.; Pour, A.B.; Askari, G.; Taghipour, N.; Pradhan, B.; Lee, C.-W.; Honarmand, M. Comparison of Different Algorithms to Map Hydrothermal Alteration Zones Using Aster Remote Sensing Data for Polymetallic Vein-Type Ore Exploration: Toroud–Chahshirin Magmatic Belt (Tcmb), North Iran. *Remote Sens.* **2019**, *11*, 495. [CrossRef]

37. Agharezaei, M.; Hezarkhani, A. Delineation of Geochemical Anomalies Based on Cu by the Boxplot as an Exploratory Data Analysis (Eda) Method and Concentration-Volume (Cv) Fractal Modeling in Mesgaran Mining Area, Eastern Iran. *Open J. Geol.* **2016**, *6*, 1269–1278. [CrossRef]

38. GSI. *Report of Systematic Geochemical Explorations in the Sahlabad Area (Sheet on Scale 1:100,000—Geochemistry of Stream Sediments)*; Geological Survey of IRAN (GSI): Tehran, Iran, 2001; p. 644.

39. Abdelouhed, F.; Algouti, A.; Algouti, A.; Mohammed, I.; Mourabit, Z. Contribution of Gis and Remote Sensing in Geological Mapping, Lineament Extractions and Hydrothermal Alteration Minerals Mapping Using Aster Satellite Images: Case Study of Central Jebilets-Morocco. *Disaster Adv.* **2021**, *14*, 15–25.

40. Yousefi, M.; Tabatabaei, S.H.; Rikhtehgaran, R.; Pour, A.B.; Pradhan, B. Application of Dirichlet Process and Support Vector Machine Techniques for Mapping Alteration Zones Associated with Porphyry Copper Deposit Using Aster Remote Sensing Imagery. *Minerals* **2021**, *11*, 1235. [CrossRef]

41. Khan, S.A.; Naim, I.; Kusi-Sarpong, S.; Gupta, H.; Idrisi, A.R. A Knowledge-Based Experts' System for Evaluation of Digital Supply Chain Readiness. *Knowl.-Based Syst.* **2021**, *228*, 107262. [CrossRef]

42. Wang, C.-N.; Nguyen, N.-A.; Dang, T.-T.; Lu, C.-M. A Compromised Decision-Making Approach to Third-Party Logistics Selection in Sustainable Supply Chain Using Fuzzy Ahp and Fuzzy Vikor Methods. *Mathematics* **2021**, *9*, 886. [CrossRef]

43. Padma, T.; Shantharajah, S.P.; Ramadoss, P. Hybrid Fuzzy Ahp and Fuzzy Topsis Decision Model for Aquaculture Species Selection. *Int. J. Inf. Technol. Decis. Mak.* **2022**, *21*, 999–1030. [CrossRef]

44. Khosravi, V.; Shirazi, A.; Shirazy, A.; Hezarkhani, A.; Pour, A.B. Hybrid Fuzzy-Analytic Hierarchy Process (Ahp) Model for Porphyry Copper Prospecting in Simorgh Area, Eastern Lut Block of Iran. *Mining* **2022**, *2*, 1–12. [CrossRef]

45. Bai, H.; Cao, Y.; Zhang, H.; Zhang, C.; Hou, S.; Wang, W. Combining Fuzzy Analytic Hierarchy Process with Concentration–Area Fractal for Mineral Prospectivity Mapping: A Case Study Involving Qinling Orogenic Belt in Central China. *Appl. Geochem.* **2021**, *126*, 104894. [CrossRef]

46. Mars, J.C.; Rowan, L.C. Spectral Assessment of New Aster Swir Surface Reflectance Data Products for Spectroscopic Mapping of Rocks and Minerals. *Remote Sens. Environ.* **2010**, *114*, 2011–2025. [CrossRef]

47. Mars, J.C.; Rowan, L.C. Regional Mapping of Phyllic-and Argillic-Altered Rocks in the Zagros Magmatic Arc, Iran, Using Advanced Spaceborne Thermal Emission and Reflection Radiometer (Aster) Data and Logical Operator Algorithms. *Geosphere* **2006**, *2*, 161–186. [CrossRef]

48. Spatz, D.M.; Wilson, R.T.; Pierce, F.W.; Bolm, J.G. Remote Sensing Characteristics of Porphyry Copper Systems, Western America Cordillera. *Ariz. Geol. Soc. Dig.* **1995**, *20*, 94–108.

49. Moradpouri, F. A Copper Porphyry Promising Zones Mapping Based on the Exploratory Data, Multivariate Geochemical Analysis and Gis Integration. *Appl. Geochem.* **2021**, *132*, 105051. [CrossRef]

50. Heidari, S.M.; Afzal, P.; Ghaderi, M.; Sadeghi, B. Detection of Mineralization Stages Using Zonality and Multifractal Modeling Based on Geological and Geochemical Data in the Au-(Cu) Intrusion-Related Gouzal-Bolagh Deposit, Nw Iran. *Ore Geol. Rev.* **2021**, *139*, 104561. [CrossRef]

51. Mirsepahvand, F.; Jafari, M.R.; Afzal, P.; Arian, M.A. Identification of Alteration Zones Using Aster Data for Metallic Mineralization in Ahar Region, Nw Iran. *J. Min. Environ.* **2022**, *13*, 309–324.
52. Chen, G.; Huang, N.; Wu, G.; Luo, L.; Wang, D.; Cheng, Q. Mineral Prospectivity Mapping Based on Wavelet Neural Network and Monte Carlo Simulations in the Nanling W-Sn Metallogenic Province. *Ore Geol. Rev.* **2022**, *143*, 104765. [CrossRef]
53. Dumakor-Dupey, N.; Arya, S. Machine Learning—A Review of Applications in Mineral Resource Estimation. *Energies* **2021**, *14*, 4079. [CrossRef]
54. Liu, L.; Cao, W.; Liu, H.; Ord, A.; Qin, Y.; Zhou, F.; Bi, C. Applying Benefits and Avoiding Pitfalls of 3d Computational Modeling-Based Machine Learning Prediction for Exploration Targeting: Lessons from Two Mines in the Tongling-Anqing District, Eastern China. *Ore Geol. Rev.* **2022**, *142*, 104712. [CrossRef]
55. Agrawal, N.; Govil, H.; Chatterjee, S.; Mishra, G.; Mukherjee, S. Evaluation of Machine Learning Techniques with Aviris-Ng Dataset in the Identification and Mapping of Minerals. *Adv. Space Res.* **2022**, *in press*.

Recognition of Sago Palm Trees Based on Transfer Learning

Sri Murniani Angelina Letsoin [1,2,*]**, Ratna Chrismiari Purwestri** [3]**, Fajar Rahmawan** [4] **and David Herak** [1]

1   Department of Mechanical Engineering, Faculty of Engineering, Czech University of Life Sciences Prague, Kamýcká 129, Suchdol, 16500 Praha, Czech Republic
2   Faculty of Engineering, University of Musamus, Merauke 99611, Indonesia
3   Department of Excellent Research EVA 4.0, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Kamýcká 129, Suchdol, 16500 Praha, Czech Republic
4   INTSIA Foundation of Papua Province, Furia 3 Number 116 Abepura, Jayapura City 99225, Indonesia
*   Correspondence: letsoin@tf.czu.cz

Abstract: Sago palm tree, known as *Metroxylon Sagu Rottb*, is one of the priority commodities in Indonesia. Based on our previous research, the potential habitat of the plant has been decreasing. On the other hand, while the use of remote sensing is now widely developed, it is rarely applied for detection and classification purposes, specifically in Indonesia. Considering the potential use of the plant, local farmers identify the harvest time by using human inspection, i.e., by identifying the bloom of the flower. Therefore, this study aims to detect sago palms based on their physical morphology from Unmanned Aerial Vehicle (UAV) RGB imagery. Specifically, this paper endeavors to apply the transfer learning approach using three deep pre-trained networks in sago palm tree detection, namely, SqueezeNet, AlexNet, and ResNet-50. The dataset was collected from nine different groups of plants based on the dominant physical features, i.e., leaves, flowers, fruits, and trunks by using a UAV. Typical classes of plants are randomly selected, like coconut and oil palm trees. As a result, the experiment shows that the ResNet-50 model becomes a preferred base model for sago palm classifiers, with a precision of 75%, 78%, and 83% for sago flowers (SF), sago leaves (SL), and sago trunk (ST), respectively. Generally, all of the models perform well for coconut trees, but they still tend to perform less effectively for sago palm and oil palm detection, which is explained by the similarity of the physical appearance of these two palms. Therefore, based our findings, we recommend improving the optimized parameters, thereby providing more varied sago datasets with the same substituted layers designed in this study.

Keywords: classification; deep learning; models; sago

1. Introduction

Sago palm from the genus *Metroxylon* grows naturally in Asian countries such as Indonesia, specifically in Papua or West Papua Province. This palm has become more important recently since the Indonesian Government is concerned about the role of this palm in various sectors, such as the food industry, as well as other uses [1,2]. Nevertheless, the detection of sago palm trees tends to be challenging due to their comparable features with other plants, for instance, coconut tree or oil palm tree, especially in natural sago forests, where they commonly live together with other particular plants. Therefore, appropriate assessment should be based on their spatial need considerations [3,4]. Results of the previous research regarding land cover changes in the Papua Province of Indonesia and the impact on sago palm areas in the region confirmed that 12 of 20 districts of Merauke Regency in Papua Province tended to lose their potential sago palm habitats. Therefore, one of the recommendations is to attempt to detect and recognize the sago palm [5]. The palm has made significant contributions to supporting local households, for instance, low bioethanol, particularly waste from washing and purifying of sago processing, and food security, specifically from the starches [4,6]. When the harvest time begins, as indicated

by the flowers on the top center of the tree, local people will cut the tree and remove the bark, followed by processing to extract the starch. Figure 1a describes the sago palm tree that was captured by using a UAV from our fieldwork, whereas (b) and (c) represent the traditional processes of local farmers at Mappi Regency of Papua Province in Indonesia. The general activity consists of bark removal, pulping, washing, purifying, and subsequent sieving. The visual interpretation will be more demanding as a consequence of the height of the plant, which can be more than 15 m in swampy areas, along with swampy shrubs [7].

(**a**) Sago palm tree from the fieldwork

(**b**) Traditional process during the harvest time

(**c**) Traditional washing and purifying

(**d**) Dry starches

Figure 1. (**a**) Sago palm tree in the fieldwork. (**b**) Traditional bark removal and pulping. (**c**) Washing, purifying, and sieving to get the starch. (**d**) Dried sago starches are ready to use.

Conversely, the advancement of remote sensing technology is quite preferable for solving particular situations, such as detection or recognition. For example, using Unmanned Aerial Vehicle (UAV) data to identify multiple land crops and then classifying these data according to their area, or utilizing nonproprietary satellite imagery tools such as Google Earth Pro to detect the selected object [8,9]. The previous study developed object-based image analysis (OBIA) and image processing using high-resolution satellite imagery for sago palm classification. Nevertheless, the study pointed out some challenges for sago palm classification, for instance, asymmetrical spatial shape due to the semi-wild stand palm, various clumps, and overlapped palm trees [3]. Remote sensing technology using satellite imagery or UAV has been combined with artificial intelligence algorithms or image analytics, supported by various methods, including a deep learning model. As established by [10], detecting the particular species in wetland areas using transfer learning in the stem density system for potatoes [11], or applying deep learning in UAV images to obtain the expected attributes from kiwi fruit, such as location and canopy chain [12], are possible. The deep learning and transfer learning environment have not only been applied in the agricultural sector, but have also been applied towards other objectives, for instance, discovering turbine blade damage [13], crime monitoring systems based on face recognition [14], or energy forecasting using transfer learning and particle swarm optimization (PSO) [15]. Deep learning is a sub-field of machine learning where the model is derived by an artificial neural network structure [16]. Recently, deep learning has been successfully applied in plant detection, for instance, tree crown detection, which generally could be performed by three approaches, i.e., Convolutional Neural Network (CNN), semantic segmentation, and object detection using YOLO, etc. [17]. This current study uses deep learning based on a CNN, which consists of an input layer, convolution layers, pooling layers, fully connected layers, and an output layer. This network's detection system proved superior to other machine learning methods [18,19]. In an image classification task, the machine learning model takes different feature extraction of images, for instance, shape, height, etc., then moves to the classification step. Meanwhile, a medium or huge amount of dataset should be available. In contrast, deep learning obtains the images without a manual feature extraction step. Feature extraction and classification work through model layers; therefore, deep

learning requires a large amount of data to achieve a good performance. It takes a long time to train the model and less time to test it. Since machine learning and deep learning need a more extensive dataset for training, they will require a higher hardware memory capacity [20]. To overcome the lack of data and the time cost consumed during training, transfer learning can be applied by using a deep learning model. Several earlier studies defined transfer learning as a technique using a model that has been trained for one task, which is then used as a baseline to train a model for another typical assignment; as long as the target model is in the same domain [21]. There are three main strategies for doing transfer learning on deep learning models, i.e., using pre-trained models, applying feature extraction by discarding fully connected output layers, and fine-tuning the last layers on pre-trained models [22–24].

Numerous deep learning networks based on CNN have been widely elaborated, for instance, GoogLeNet and DenseNet. Nonetheless, as explained before, the existing model can be modified for other purposes but has not yet been investigated for sago palm detection. Furthermore, transfer learning is acceptable with fewer data and could reduce training time and computer resources, as concluded by an earlier study. Therefore, the current study will use a transfer learning strategy to predict the plants based on their physical appearances, such as leaves, trunks, flowers, or fruits. In order to address this, three different pre-trained networks based on CNN were customized for detection and prediction; namely, SqueezeNet, AlexNet, and ResNet-50, were applied in this study. We modified the last layer and discarded the fully connected output layer to achieve our new task. The study's dataset consists of data training and testing of three plants: sago palm, coconut tree, and oil palm tree. Each class is categorized based on features such as leaves, fruits or flowers, and trunks. The study aims (1) to obtain the prior model based on classification performance, i.e., precision, F1-score, and sensitivity; and (2) to evaluate the transfer learning task in sago palm detection based on leaves, flowers, and trunks.

2. Materials and Methods

2.1. Study Region

The study was performed in Merauke Regency (Figure 2), which is located in the southern part of Papua Province of Indonesia (137°38′52.9692″E–141°0′13.3233″E and 6°27′50.1456″S–9°10′1.2253″S). In the last decade, the population growth in Papua Province was around 18.28%, and approximately 1.20 million people there are economically active in agriculture. Based on weather data, the annual minimum and maximum temperature average deviates between 16–32 °C, while the average rainfall registered is 2900 mm with high humidity from 62% to 95%.
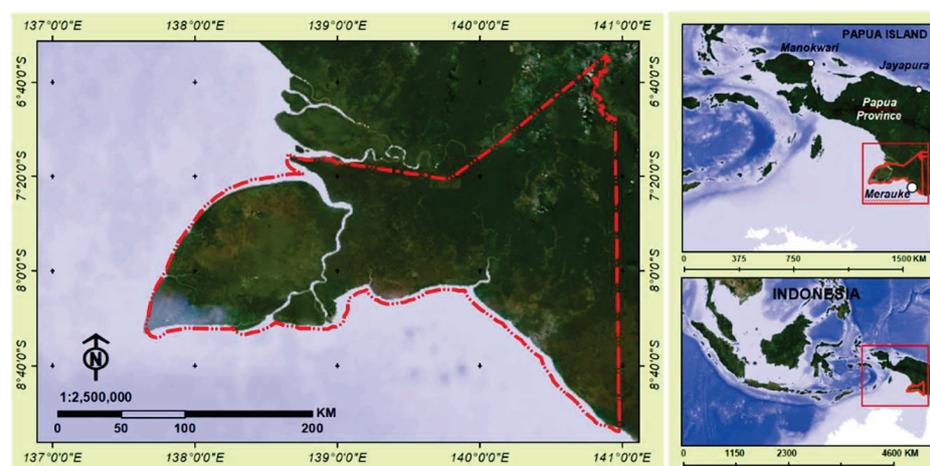
Figure 2. Location of study area: Merauke Regency. Consisting of 20 districts that cover an area of 46,791.63 km^2 with a population of 230,932 in 2020.

2.2. UAV Imagery

Unmanned aerial imagery was captured in Tambat Village, located in Merauke Regency, specifically, the Tanah Miring District. This fieldwork was performed at one of the top sago forest producers in the Merauke Regency. The sago forest images, shown in Figure 3, were captured at a height of 60 m and 100 m, longitude, and latitude of $140°36'46.8498''E$–$8°21'21.2544''S$, respectively. The visible morphology of the sago palm as detected by the UAV is presented in Figure 3b–d. Sago palm in this fieldwork is typically natural sago forest or wild sago. It contains a palm trunk, which stores the starch. At harvest time, the trunk will be cut off and the bark will be opened, followed by further processing to extract the starch. The harvest time of these sago forest areas is commonly identified by the bloom of the flowers, as introduced in Figures 1 and 3, followed by the leaves. The dataset used in this research consists of high-resolution RGB images taken from a UAV by an Autel Robotics multi-copter. Additionally, field survey data were obtained by performing ground photography and a short interview with local sago farmers. Our study focuses on the morphology of sago rather than sago palm health classification or automatic counting; therefore, our dataset also shows other typical plants such as coconut tree and oil palm, based on their leaves, fruits, and trunks.

(**a**)   (**b**)

(**c**)   (**d**)
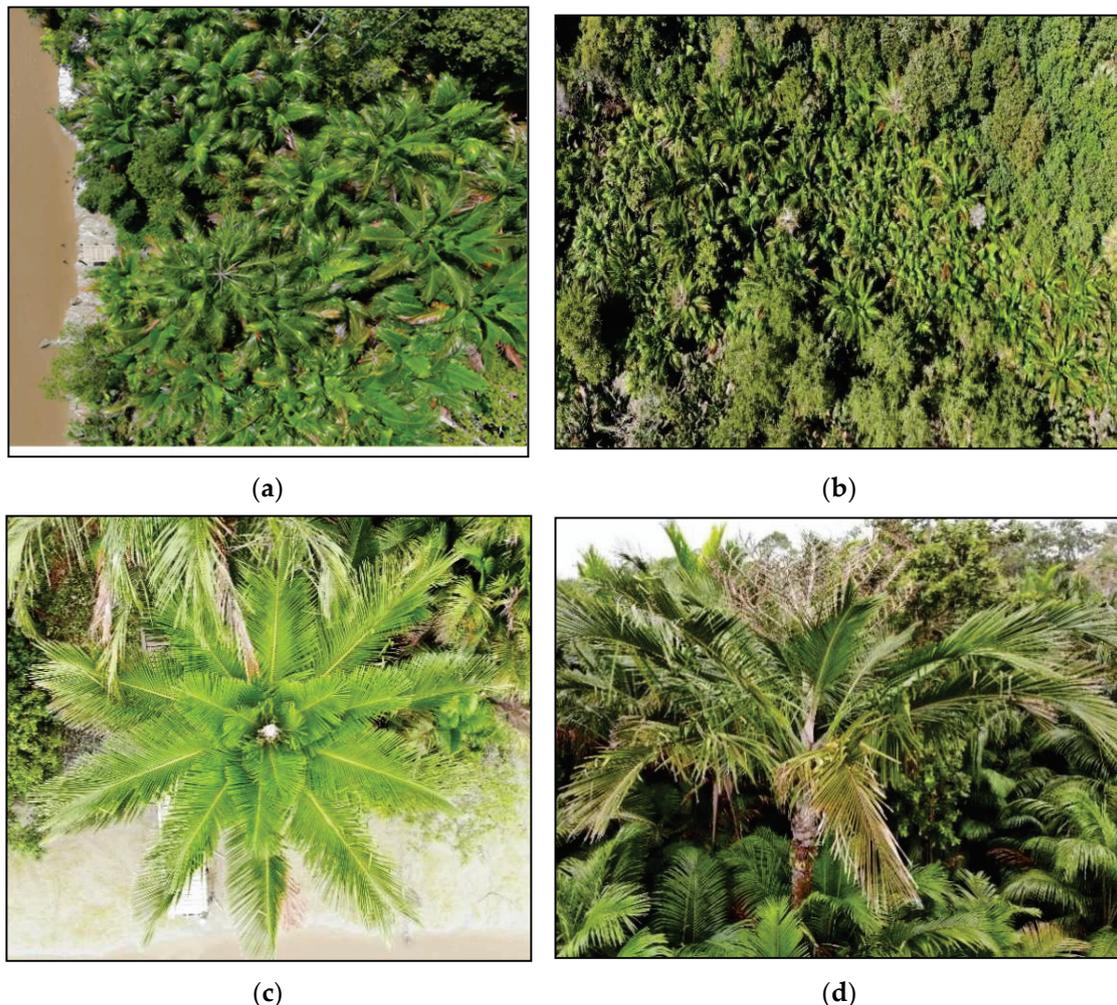
Figure 3. Sago palm captured imagery using UAV RGB and ground photography observation. (**a**,**b**) Sago palm areas in the fieldwork, and other vegetations; (**c**) Sago flowers defined by white flowers at the top center, between leaves; (**d**) Palm tree dominant features: trunk and leaves.

The methodology of this research was developed as presented in Figure 4. First, study preparation is established throughout the field survey and ground photographs around

the fieldwork. We used some tools such as Google Earth engine, and a handheld global positioning system to ensure the location of the fieldwork. Then, we created the mission plan for the UAV. In the next stage, the UAV Red Green Blue (RGB) band images were downloaded and labeled. Next, the region of interest (ROI) was chosen based on each label category and class. The dataset in this study was divided into two types: (1) data trained and (2) data tested. The data trained were categorized into nine classes, namely coconut tree trunks, leaves, and fruits, as well as oil palm trunks, leaves, and fruits. The remaining classes are sago trunks, leaves, and flowers. Considering the classification process and prediction model, we applied deep learning model approaches, namely SqueezeNet, AlexNet, and ResNet-50. Hence, this dataset also involves various sizes of the imagery, blurred and yellowish images with different angles.
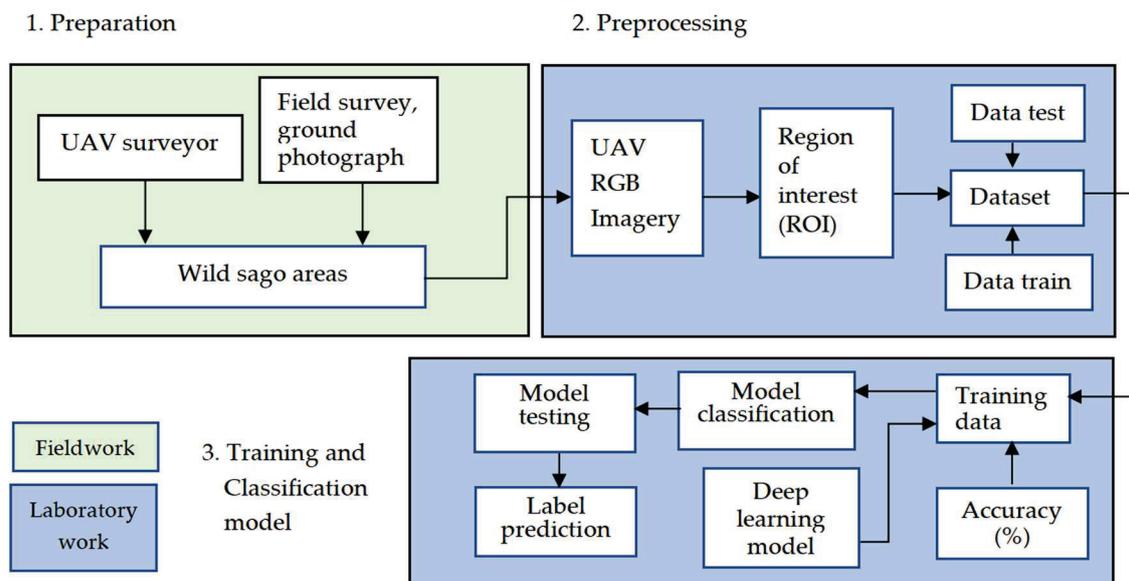

Figure 4. Study workflow. Three stages are practiced with classification and prediction based on three major features of each plant.

The classification and prediction process began after the data collection and training data were developed. Subsequently, the deep learning models were applied, including parameter optimization for instance mini batch size, initial learning rate or epoch. The earlier study has successfully combined the parameter optimization to obtain the higher performance in classification task, for example, learning rate was set up to 0.0001 with ten number of epochs [23,25]. Further, it was trained and tested using a dataset utilized from the previous stage. The accuracy of the sago palm dataset was evaluated by comparing the results of drone imagery with actual data derived from the visual interpretation, and was based on the ground survey and photographs. All of the training and testing procedures were implemented using MATLAB R2021 and deep learning scripts.

2.3. Deep Learning and Transfer Learning Models

Deep Learning (DL) models were defined throughout the pre-trained networks of several common architectures already provided in MATLAB packages for DL, such as AlexNet [26], GoogLeNet [27], ResNet-50 [28], and Inception v3 [29]. This study focused on three networks based on CNN, namely SqueezeNet, AlexNet, and ResNet-50. SqueezeNet consumes small bandwidth to export new models to the cloud or upload them directly to devices. This network is also able to deploy on embedded system or other hardware devices that have low memory [30], while AlexNet shows higher accuracy compared to other different DL, such as GoogleNet or ResNet152, in image classification on the ImageNet dataset [31], however ResNet as the backbone network shows good performance for the segmentation dataset [32]. The transfer learning (TL) strategy used in this study requires

two stages: the base model, which is constructed on pre-training CNN models, and the target model [33], which is tailored to a new, specific task (Figure 5). Three pre-trained networks were used, namely SqueezeNet, ResNet, and AlexNet for the base model, and then we reconstructed the base model to our target model, with nine probability classes. These three models are trained in various datasets, such as the ImageNet Dataset, and are able to classify images into 1000 object categories or 1000 classes [34,35], such as keyboard, mouse, coffee mug, pencil. Nevertheless, TL allows a small dataset, reuse, and extraction of transfer features, and improves the training time of models [21,23,25].
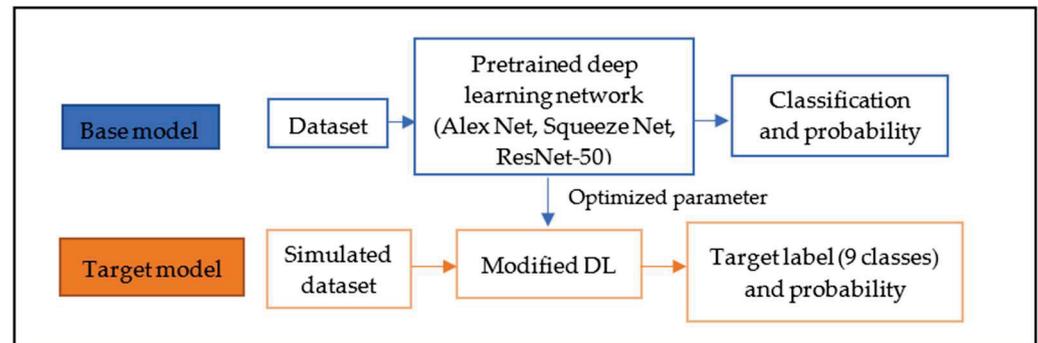
Figure 5. Transfer learning workflow in this study.

AlexNet network involves 5 convolutional layers (conv1–conv5), 3 fully connected layers (fc6, fc7, fc_new) within the ReLu layer are established after every convolution layer (Table 1). Further, the dropout layer (0.5 or 50%) avoids overfitting problems. According to the tools used in this study, the input size is $227 \times 227 \times 3$ or 154,587 values [36], and all layers must be connected.

Table 1. AlexNet designed in this study.

Layer	Layer Name	Layer Type	Layer Details
1	Data	Image input	$227 \times 227 \times 3$ images with zero center normalization
2	Conv1	Convolution	96 $11 \times 11 \times 3$ convolutions with stride [4 4] and padding [0 0 0 0]
3	Relu1	ReLU	ReLU
4	Norm1	Cross channel normalization	Cross channel normalization with 5 channels per elemen
5	Pool1	Max pooling	3×3 max pooling with stride [2 2] and padding [0 0 0 0]
6	Conv2	Grouped convolution	2 groups of 128 $5 \times 5 \times 48$ conv with stride [1 1] and padding [2 2 2 2]
7	Relu2	ReLU	ReLU
8	Norm2	Cross channel normalization	Cross channels normalization with 5 channels per element
9	Pool2	Max pooling	3×3 max pooling with stride [2 2] and padding [0 0 0 0]
10	Conv3	Convolution	384 $3 \times 3 \times 256$ convolutions with stride [1 1] and padding [1 1 1 1]
11	Relu	ReLU	ReLU
12	Conv4	Grouped convolution	2 groups of 192 $3 \times 3 \times 192$ convolutions with stride [1 1] and padding [1 1 1 1]
13	Relu4	ReLU	ReLU
14	Conv5	Grouped convolution	2 groups of 128 $3 \times 3 \times 192$ convoutions with stride [1 1] and padding [1 1 1 1]
15	Relu5	ReLU	ReLU
16	Pool5	Max pooling	3×3 max pooling with stride [2 2] and padding [0 0 0 0]
17	Fc6	Fully connected	4096 fully connected layer

Table 1. *Cont.*

Layer	Layer Name	Layer Type	Layer Details
18	Relu6	ReLU	ReLU
19	Drop6	Dropout	50% dropout
20	Fc7	Fully connected	4096 fully connected layer
21	Relu7	ReLU	ReLU
22	Drop7	Dropout	50% dropout
23	Fc_new	Fully connected	9 fully connected layer
24	Prob	Softmax	
25	Classoutput	Classification output	

SqueezeNet starts the network with an individual convolution layer (conv1), then a rectified linear unit (ReLU), which is a type of activation function, then the max pooling layer (Figure 6). When added to a model, max pooling reduces the dimensionality of images by decreasing the number of pixels in the output from the previous layer. Thus, the Conv+Relu layer is then extended to 8 fire modules, from fire 2 to fire 9, with a filter size combination of 1×1 and 3×3 [30]. Convolution and ReLU layer can be computationally defined as follows:

$$F_l(I) = P_{i-1}(G) \times f \tag{1}$$

where F_l describes an output feature map and l represents the lth convolution layer, while f is defined by filter size or kernel, and then P_{i-1} shows the previous layer output, and I denotes the original data image. Thus, ReLU is denoted through an equation:

$$ReLU(I) = \max(x, 0) \tag{2}$$

x is the input of activation on the lth layer, I denotes a ReLU activation output of the feature maps [37].

Another network included in this study is the Residual network (ResNet-50), the variant of the ResNet model, which has a 50-layer deep convolutional network. It contains one convolution kernel size of 7×7 at the beginning and ends with an average pool, a fully connected layer, and a SoftMax layer in one layer. Between these layers, there are 48 convolutional layers consisting of different kernel sizes [38]. Here, the fully connected layer's purpose is to integrate all of the inputs from one layer connecting to every activation unit of the next layer. Thus, the residual block on the ResNet equation is as follows, where O is the output layer, I is the input layer, and $F(I_i\,W)$ is the residual map function [39].

$$O = F(I_i\,W_i + I) \tag{3}$$

The characteristics of each model are shown in Table 2, as follows.

Table 2. Model comparison [1].

Network Name	Depth	Image Input Size	Parameters (Millions)	Total Memory (MB)
SqueezeNet	18	$227 \times 227 \times 3$	1.24	5.20
ResNet-50	50	$224 \times 224 \times 3$	25.6	96
AlexNet	8	$227 \times 227 \times 3$	61	227

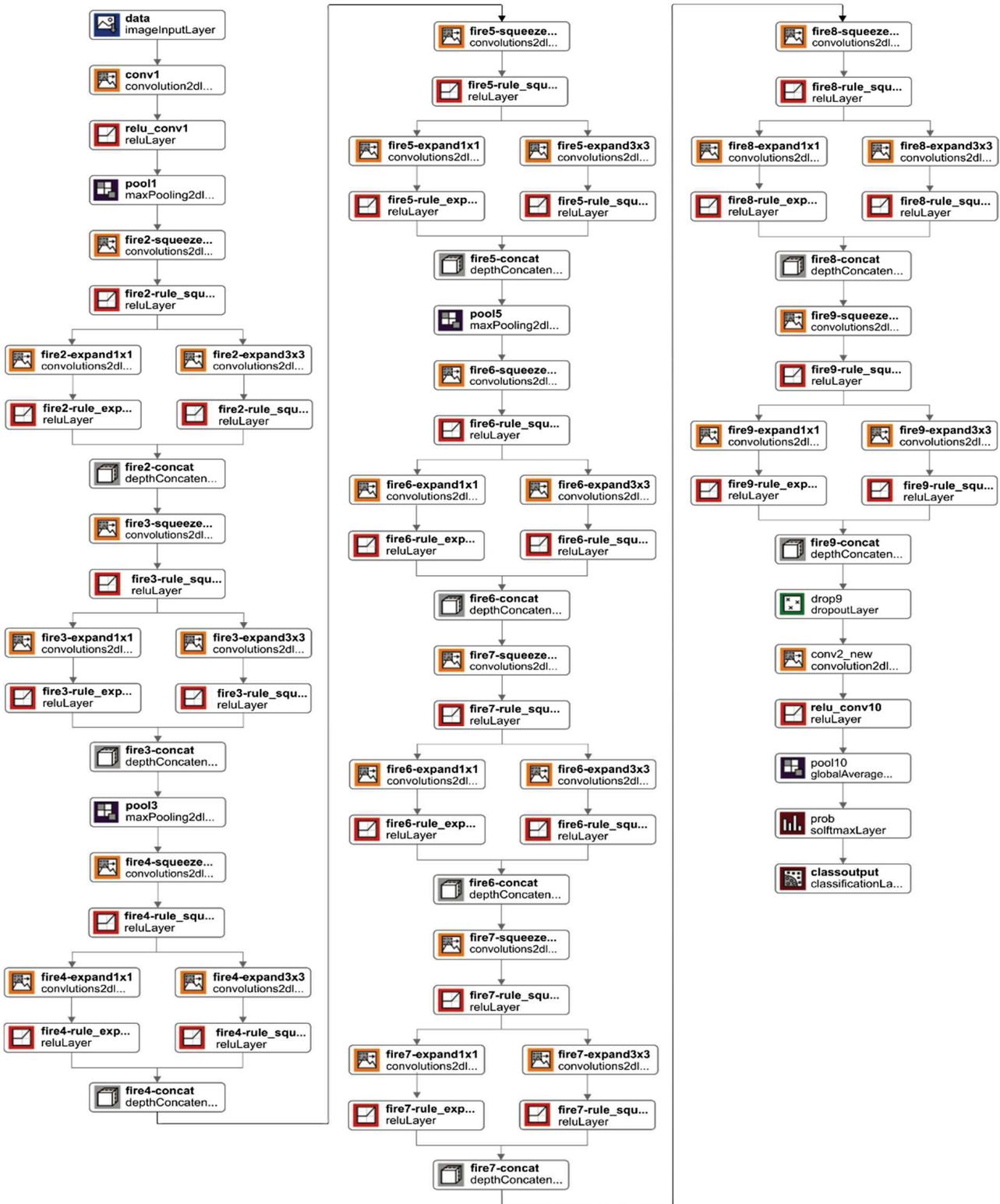[1] Based on tool used in this study, ie., MATLAB.

Figure 6. SqueezeNet used in this study: all layers are connected.

Once the data preparation is ready and the deep learning model has been designed, we can analyze the chosen model and optimize the parameters. If there are no errors in the model, then all of the training data can be imported as data trained (Figure 7).
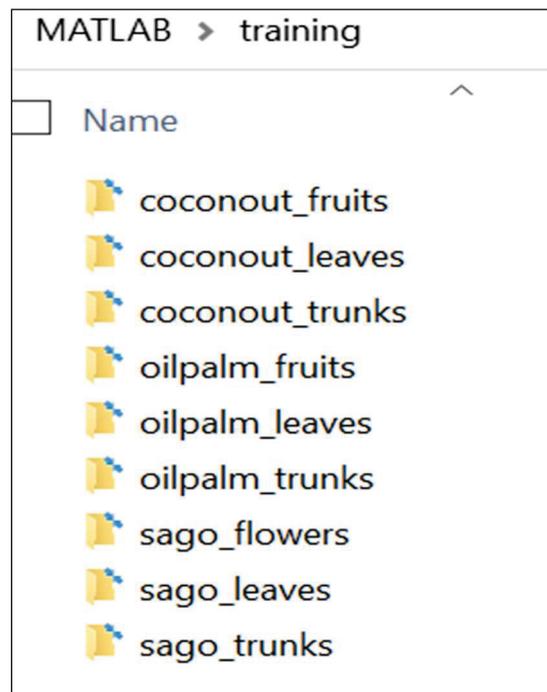
Figure 7. Simulated data train: CF, CL, CT, OPF, OPL, OPT, SF, SL, ST, respectively.

As a result, validation accuracy appears, including the training time (elapsed time) and training cycle, such as the epoch number. Optimized parameters from deep pre-trained networks are transferred to a simulated dataset and then will be trained. These models are compared using the same number of epochs, learning rate, and batch size. At the final stage, 227 × 227 and 224 × 244 image input in the data test will be resized, then a single image as an output will be predicted and categorized (Figure 8).
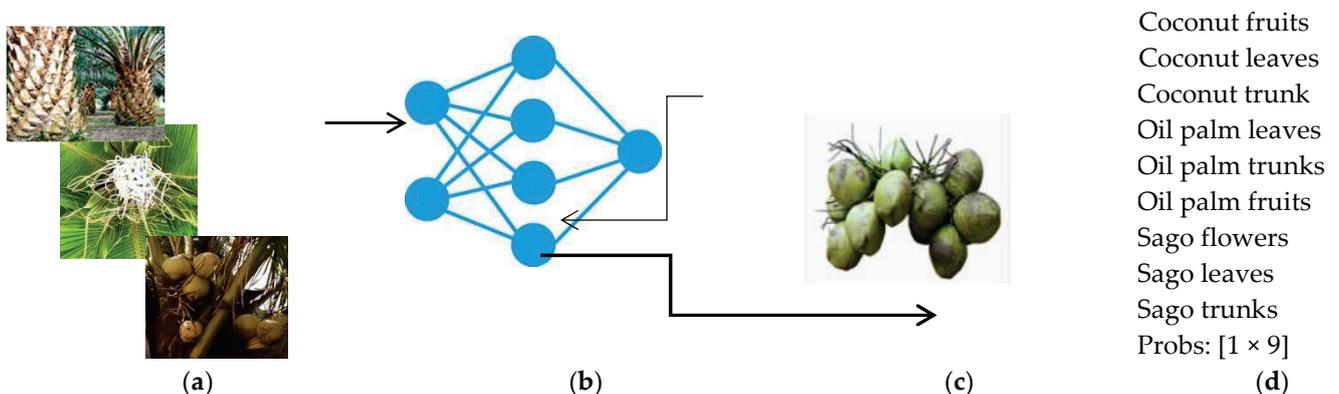

(**a**)	(**b**)	(**c**)	(**d**)

Figure 8. Examples from processing workflow. (**a**) All data trained images. (**b**) Model and optimized parameters via transfer learning. (**c**) Test image cropped (227 × 227) or (224 × 224). (**d**) Classification and probability.

2.4. Performance Evaluation

Four metrics are typically evaluated in DL and TL model evaluation, namely true positive (TP), true negative (TN), false positive (FP), and false negative (FN). In this study, TP and TN describe the correct identification of class, while FP and FN correspond to false identification of class [40]. The evaluation was investigated using an image from the validation set and their specific labels, which were not used for training. The detection ability is assessed based on the precision and the sensitivity criteria, as shown in Table 3, while the optimized parameters applied in this study are presented in Table 4.

Table 3. Evaluation criteria.

Metric	Formula	Criteria
F1-score	$\frac{2 \times (Recall \times Precision)}{Recall + Precision}$	Denotes a high value, which validates the model.
Precision	$\frac{TP}{TP + FP}$	Examines the ability of the model to predict positive label.
Sensitivity (Recall)	$\frac{TP}{TP + FN}$	Defines the ability of the model to detect instances of certain classes well.
Specificity	$\frac{TN}{FP + TN}$	Defines the true negatives that are correctly identified by the model.
Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$	Examines the accurately in identifying the images to the classes.

Table 4. Optimized parameters in this study.

Parameter Name	Value
Epochs	10
Initial learning rate	0.0001
Validation frequency	9
Learning rate weight coefficient	10
Learning rate bias coefficient	10
Learning rate schedule	Constant
Momentum	0.9
L2 Regulation	0.0001
Min batch size	10

Multi-class detection can be explained as follows, for instance, in sago flowers:

1. TP, the number of actual images that are displaying sago flowers (true) and are classified as sago flowers (predicted).
2. FP, the number of actual images that are not displaying sago flowers (not true) and are classified as sago flowers (predicted).
3. FN, the number of actual images that are displaying sago flowers (true) and are classified as a different class (predicted).
4. TN, the number of actual images that are not displaying sago flowers (not true) and are classified as a different class (predicted).

Hyperparameters set in the training model of TL (Table 4) were determined from the earlier studies [22,23] by epochs, batch size, and learning rate. A very high learning rate will trigger the loss function to go up, and as a result, the accuracy of classification can be reduced. Conversely, if it is too low, it will reduce the network training speed, the correction of weight parameter correction will be slow, and it will fail to obtain a proper model accuracy. Batch size is also vital to the accuracy of models and the training process performance. Using a larger batch size will require higher GPU memory to store all of the variables (e.g., momentum), and the weights of the network also may cause overfitting; however, using a minimum batch size may lead to slow convergence of the learning algorithm. Another technique to overcome the GPU memory limitation and run large batch sizes is to split the batch into mini-batch sizes. The number of epochs defines the learning algorithm will complete passes through the entire training dataset.

An ANOVA test was employed to compare means between the accuracy (true positive) values of three models in correctly identifying the target trees' morphology. A p value less than 0.05 was designed for a statistically significant difference in all data analyses. A receiver–operating characteristics curve (ROC curve) was employed to identify the sensitivity and 1-specificity (false positive) of the three algorithms in identifying sago (flowers, leaves, and trunks) over coconut and oil palm. All data analyses were performed using the IBM SPSS version 27 (IBM Corp., Armonk, NY, USA). Additionally, an approximate cost of software measurement was estimated using function point (*FP*) analysis [41], which

is described the functionality points through complexity adjustment factor (*CAF*), and unadjusted functional point (*UFP*), as follows [42]:

$$CAF = 0.65 + (0.01 \ x \ \sum_{k=1}^{14} F_i \ x \ rating) \tag{4}$$

$$FP = UFP \times CAF \tag{5}$$

3. Results

3.1. Dataset Development

In our fieldwork, the Autel UAV flew at various altitudes of 60 m up to 100 m, as well as with different forward and sideways overlapping, during the mornings and mid-days of July 2019, August 2021, and July 2022. This stage aims to obtain various shapes as well as to enrich the dataset images, instead of counting the plants. Next, the data collection of sago palm was downloaded and cropped, then allocated according to the labels in Figure 7. The experiments were processed with an Intel Core i7 processor, with the dataset defined in RGB space being categorized in 9 classes. The data train and data test were divided separately, as presented in Figure 4, stage 2; around 70% was allocated for data training, and 30% were used for data testing and validation. This study comprised 231 total images; 68 images for testing and the rest for data training. The same dataset was used to train and test, based on the deep learning networks used. All images were pre-processed based on the three pre-trained networks, as compared in Table 2. Three pre-trained networks of deep learning were examined and then modified. Then, they were transferred to the target as transfer learning, of which the modified version is shown in Figure 5. Regarding the new task via transfer learning, as well as to obtain the aims of this study, the last layers of each model were reformed as follows: the fully connected layer, fc1000 changes to fc_new, then SoftMax layer for converting values into probabilities, and subsequently, the classification layer predictions for 1000 output size were replaced to class_output for categorizing into nine classes. Convolution1000 layer is also restored to the conv2d layer with nine num-filters. Then, the weight learning factor and bias learn rate coefficients, as presented in Table 4. Furthermore, all images were pre-assessed using resizing and normalization, for instance, rescaling, rotation, and augmentation. In addition, the datasets were all evaluated by inspecting the UAV images, visual interpretation, and ground surveys, such as photographing plants.

### 3.2. Training and Testing Data Performance

Considering the TL process as presented as the workflow in Figure 5, all datasets were imported into a specified workplace through MATLAB and followed by other processing, namely, training data in the modified deep learning pre-trained network. As a result, training accuracy and validation lost over ten epochs and ten min batch sizes are introduced in Figure 9. The smoothness of the accuracy and the loss of training process are described by the blue colour and orange colour, respectively. While light blue coloured dots and light orange coloured dots represent the training progress. Additionally, the validation of data trained and the loss are explained by black line coloured dots. The training progress of the three models was not quite as smooth, with accuracies of 76.60%, 76.60%, and 82.98%. However, the ResNet-50 model is more dominant when compared to the others, with the highest accuracy of 82.98%. The training loss values decreased sharply on these models in epoch 5, while the training progress increased. Subsequently, the validation accuracy and loss curves were more eased, especially in ResNet-50 and AlexNet, where the data training loss decreased during the rest of the process. Although SqueezeNet and AlexNet fluctuated after 5 epochs, AlexNet network validation was improved, while the training loss was smaller. This result demonstrates the ability of the three classifiers in recognizing the dataset.

(**a**)
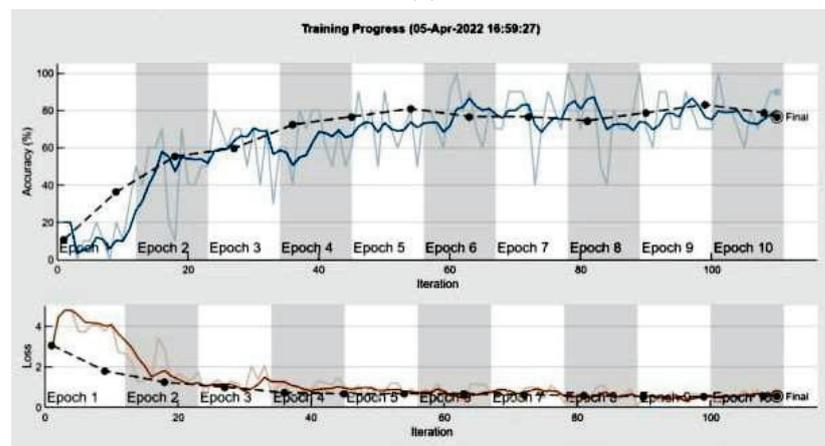
(**b**)
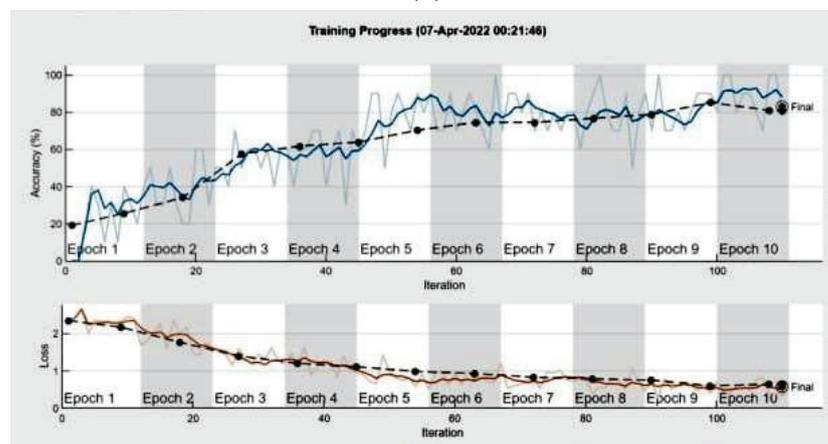
(**c**)

Figure 9. Training accuracy: (**a**) SqueezeNet, accuracy of 76.60%. (**b**) AlexNet, accuracy of 76.60%. (**c**) ResNet-50, the accuracy of 82.98%.
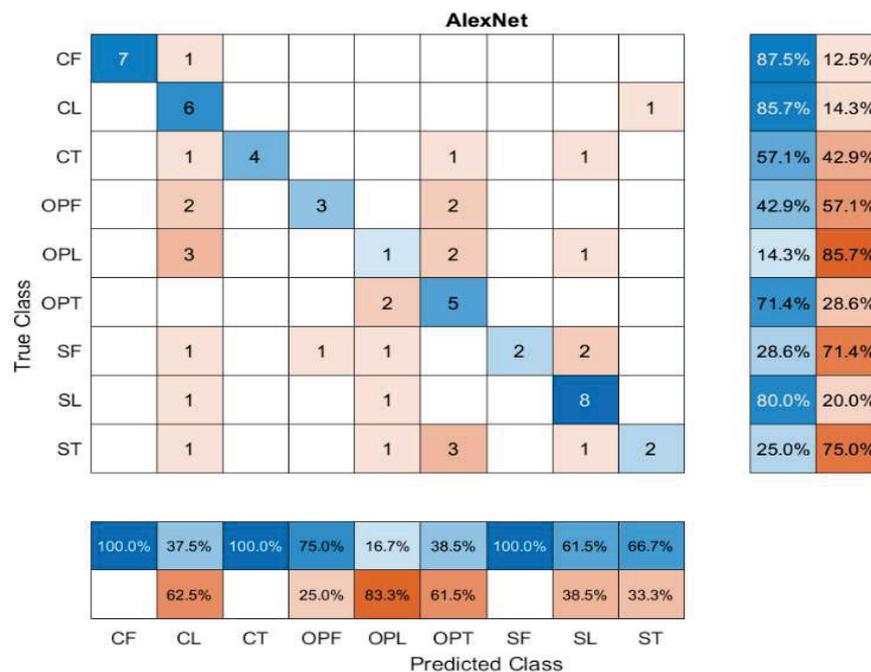
After the training process shown in Figure 9, all models were tested using the same data test, which was prepared and placed differently than the trained data. To support this testing process, we used various types of syntax that were accommodated in MATLAB2021, such as imresize, imshow, prediction, probability, and confusion matrix. The uses of imresize and imshow are basically appropriate preparation for the input test, according to the features of each model in Table 2, while the probability and categorization were generated from each model, specifically in layer name: prob, within SoftMax type (Figure 6). Next, the confusion matrix was calculated for each classification model, and the performance was visualized using the values on the confusion matrix. The confusion matrix in this study

was used to describe each model's performance, consisting of the true class and the class predicted by the model. Then, the metric was calculated based on the formula shown in Table 3.

All models were able to predict the plants with 100% accuracy, such as SqueezeNet (Figure 10) for recognizing the coconut trunk (CT), AlexNet for coconut fruit (CF), coconut trunk (CT), and sago flowers (SF), while ResNet-50 recognized oil palm trunk (OPT). In the case of sago palm classification, the convolution matrix of AlexNet and ResNet-50 were superior to SqueezeNet. Despite the fact that the models were trained with a self-contained dataset and smaller datasets compared to pre-trained deep learning when utilizing 1000 images, the training accuracy rose to 82%. Meanwhile, the models obtained the expected results in the recognition of the plant's physical morphology.

(**a**)

(**b**)

Figure 10. *Cont.*

Figure 10. Confusion matrix of predictions made by TL on (**a**) SqueezeNet; (**b**) AlexNet; (**c**) ResNet-50.

3.3. Model Performance Evaluation

In terms of precision and detection of the sago palm based on leaves (SL), flowers (SF), and trunks (ST), the best performances are highlighted in bold (Table 5). Even though the AlexNet model is 100% able to detect sago flowers (SF), the sensitivity of this model or valid positive rate is only around 29%. Conversely, the ResNet-50 model is quite good as a classifier in SL and ST, with the precision value for SL and ST being 0.78 and 0.83, respectively. Precision and sensitivity should preferably be 1, which is the highest value, or close to 100% if expressed in percentage. Nevertheless, the F1 score turns out to be ideally 1 when both precision and sensitivity are increased or equal to 1. Therefore, this study examines the precision, as well as the recall or sensitivity to evaluate the performance of each model as a good classifier.

Table 5. Classification results of three networks.

Model	Training Accuracy (%)	Training Time	Image Input Size	Class	Recall (Sensitivity)	Precision	F1 Score
SqueezeNet	76.60	3 min 39 s	227×227	CF	1.00	0.80	0.89
				CL	0.83	0.83	0.83
				CT	0.71	1.00	0.84
				OPF	0.71	0.71	0.71
				OPL	0.57	0.33	0.42
				OPT	0.71	0.63	0.67
				SF	0.29	0.67	0.41
				SL	0.70	0.54	0.61
				ST	0.25	0.67	0.36
AlexNet	76.60	5 min 8 s	227×227	CF	0.88	1.00	0.94
				CL	0.86	0.38	0.53
				CT	0.57	1.00	0.73
				OPF	0.43	0.75	0.55
				OPL	0.14	0.17	0.48
				OPT	0.71	0.39	0.51
				SF	0.29	1.00	0.45
				SL	0.80	0.62	0.70
				ST	0.25	0.67	0.36

Table 5. *Cont.*

Model	Training Accuracy (%)	Training Time	Image Input Size	Class	Recall (Sensitivity)	Precision	F1 Score
ResNet-50	82.98	18 min 29 s	224 × 224	CF	0.88	0.88	0.88
				CL	0.71	0.46	0.56
				CT	0.57	0.80	0.67
				OPF	0.57	0.67	0.62
				OPL	0.71	0.39	0.50
				OPT	0.57	1.00	0.73
				SF	**0.43**	**0.75**	**0.55**
				SL	**0.70**	**0.78**	**0.74**
				ST	**0.63**	**0.83**	**0.72**

SqueezeNet performed significantly better in identifying the leaves of oil palm (OPL) than AlexNet (p = 0.046); meanwhile, no statistical significance difference was found between the tree models in accurately recognizing the target tress based on fruits, leaves, and trunks. Based on the accuracy values, the AlexNet had a more surprising performance in the detection of sago flower (SF) than the other models (Figure 11), while ResNet-50 can recognize the sago trunk (ST) and sago leaves (SL) better than other models. These results indicate that the models can distinguish sago palms from other plants used in this study. Based on this evaluation, the AlexNet and ResNet-50 can promote the preliminary detection of the sago palm.
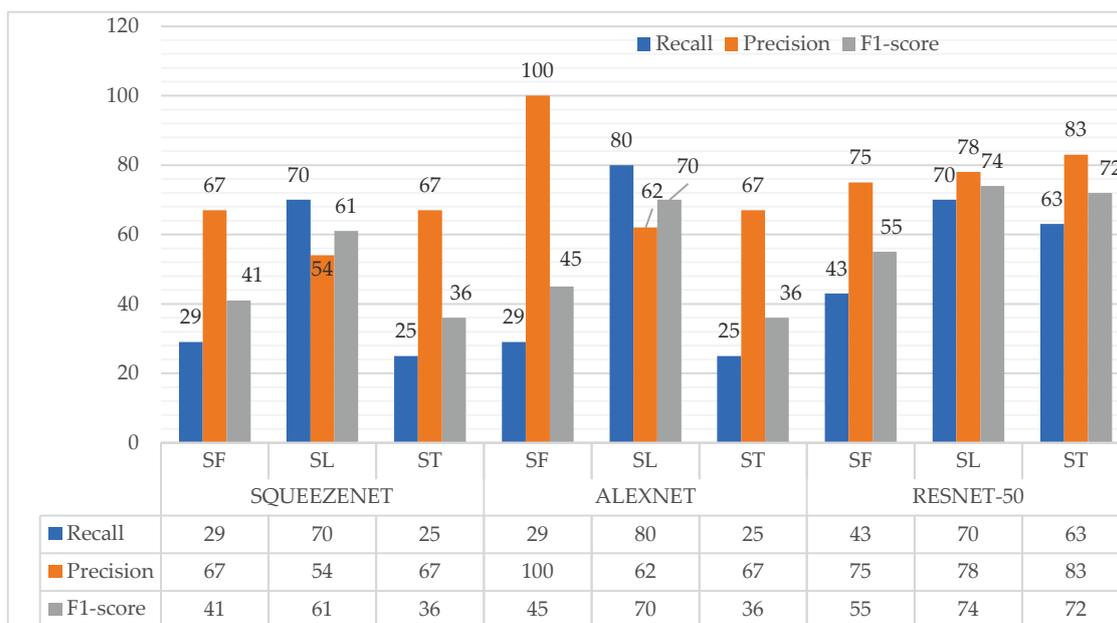

	SF	SL	ST	SF	SL	ST	SF	SL	ST
		SQUEEZENET			ALEXNET			RESNET-50	
Recall	29	70	25	29	80	25	43	70	63
Precision	67	54	67	100	62	67	75	78	83
F1-score	41	61	36	45	70	36	55	74	72

Figure 11. Performance of sago palm classifier in percentage (%).

ROC curves compared all experimented models on the sago palm dataset, the results showed that all algorithms were able to correctly identify sago (flowers and leaves) over coconut and oil palm (Figure 12a–c), of which ResNet50 presented the best model for predicting sago trees. SqueezeNet and ResNet 50 could also distinguish between the sago trunks over that of the coconut and palm oil, however, AlexNet was less likely to identify it (as depicted by the line under the reference values) (Figure 12a).

The cost analysis of model implementation according to *FP* was estimated with the result of *CAF* and *FP*, i.e., 1.21 and 1057.54 *FP*, respectively. Comparable other costs revealed similar expenses for the further deployment of the model. The only difference between them was found in the performance of three models, as presented above (Figures 11 and 12).
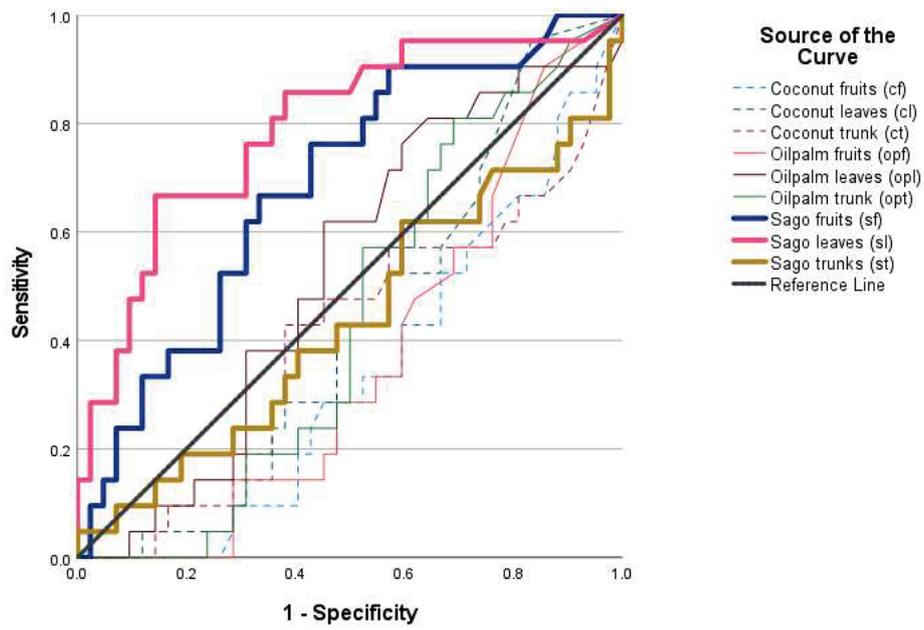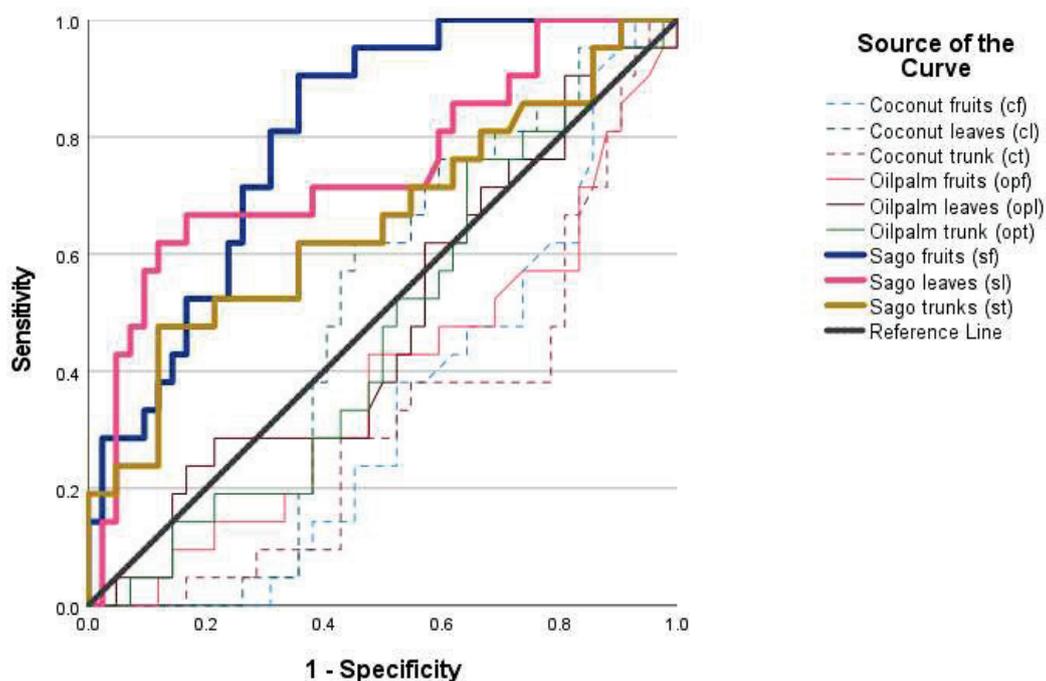
(**a**)
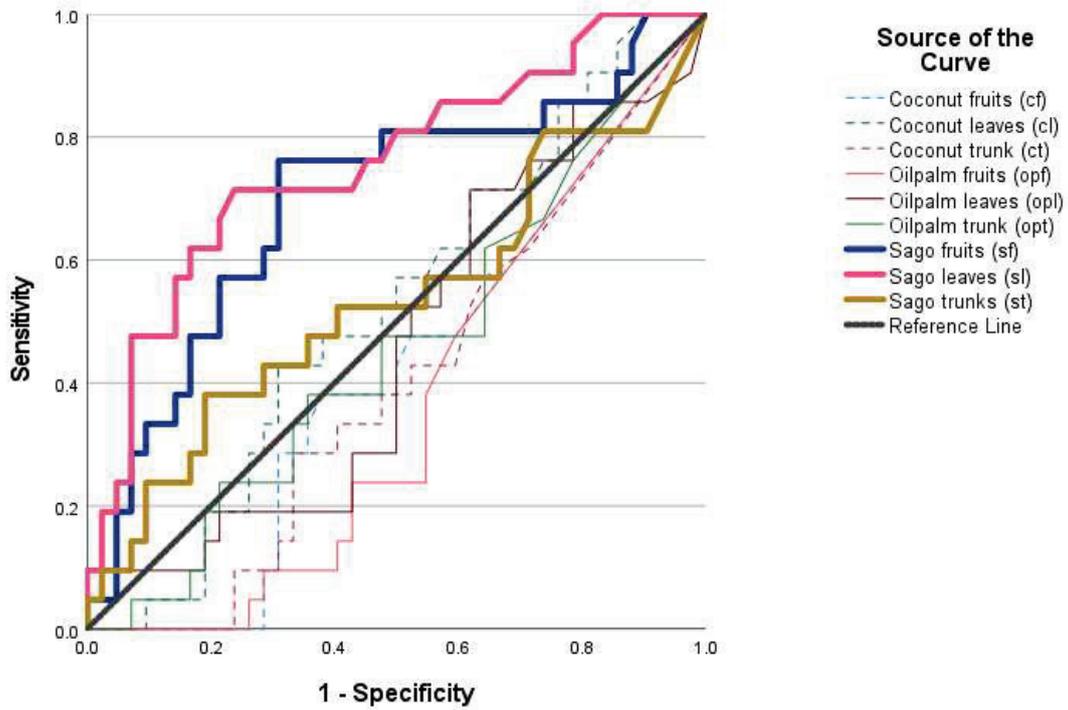
(**b**)

Figure 12. *Cont*.

(**c**)

Figure 12. (**a**) ROC curve of the AlexNet. (**b**) ROC curve of the SqueezeNet. (**c**) ROC curve of the ResNet50.

4. Discussion

The implementation of deep learning can be performed with two methods: (a) a self-designed deep learning model, and (b) transfer learning approaches. In this study, transfer learning based on three models, namely, SqueezeNet, AlexNet, and ResNet-50 were used as transfer objects to categorize and predict the three plants based on their physical morphology. Generally, the three models can detect the morphology of coconut trees well, specifically SqueezeNet, as shown by the higher precision in CF, CL, and CT—80%, 83%, 100%, respectively, when compared to sago palm or oil palm. This happens because the shape of palm oil and sago are similar, as shown in Figures 1 and 3, especially when captured by using a drone or other remote sensing technology at a specific altitude [3]. As investigated by [43], tree classification using UAV imagery and deep learning has confirmed that deep learning and transfer learning can apply to the classification of UAV imagery, however, more tree species and various study areas will improve the accuracy of the classifiers. Concerning the performance of sago classifiers, as shown in Figure 11, AlexNet can predict sago flowers (SF) at 100%, while ResNet-50 forecasts sago leaves (SL) and sago trunk (ST) at 78% and 83%, respectively. A different study of wood structure found that the testing performance of ResNet-50 as a transfer object was about 82% from 4200 images of the dataset [44,45]. Additionally, for carrot crop detection, which included 1115 standard-quality carrots and 1330 imperfect carrots using ResNet-50, it was proven that this transfer method is superior compared to the others. Even though TL can predict the class with fewer datasets, it can provide a variety of sago palm datasets that will improve the learning performances [46]. Therefore, providing more datasets with different types, angles, and shapes of the sago is recommended for further work.

Considering the availability of datasets from UAV imagery, findings have been provided by several studies, as mentioned in [40,47], for instance, weed map dataset, VOAI dataset, or other resources, such as ImageNet [48,49]; we found that these do not follow the requirements of this study, especially the dataset provided, for example, a dataset of ImageNet supports the recognition of various images such as vehicles, birds, carnivores,

mammals, furniture, etc., but it is obviously not yet purposed for a sago palm dataset. However, previous studies explained how a proper dataset helps enhance learning performance; therefore, we applied transfer learning as a strategy to overtake the insufficient data, since it could train network models with a small dataset [40]. For the dataset, our study provided its own dataset captured from UAV images and labeled according to each class. The original dataset for nine classes contained 144 images, while the augmentation process obtained 19 images. The augmentation process consists of rotation, scale (Appendix B), and then 68 test images. In total, the dataset used in this study contains 231 images. Nevertheless, the existing data, for instance, UW RGB-D object dataset, provides 300 general objects in 2.5 datasets [32]. At the testing stage, the RGB images were resized based on the model (Table 2), which was also done by the earlier studies [49,50]. Transfer learning-based CNN models using UAV imagery generate one label for one image patch rather than making a class prediction for every pixel [51]. On the one hand, the presence of overlapping plants, for example, wild sago palm (e.g., lives with other vegetation, irregular shape), could be more challenging in pixel-based classification, for instance, semantic segmentation [40,47,51]. Nevertheless, providing a DL-based segmentation dataset of overlap sago combined with other models is essential for different purposes [32]. At the same time, the selected models used in this study performed detection and recognition successfully, as assessed by earlier studies [23,52]. According to the result of the metric evaluation, the ResNet-50 model outperforms, at around 90%, compared to other networks, which was also depicted by its ROC curve (Figure 12c). Nonetheless, the effects of hyperparameters of each model, such as learning rate, epochs, and minibatch sizes require consideration [53]. Consequently, fine-tuning the parameters of each model should be more noticeable, which is also described in the limitations of this current study.

Since the sago palm has become important in Indonesia, and considering that the potential area for sago in Papua province tends to be declining [5], designing a relevant application using reliable methods or algorithms needs to be considered. In the case of sago palm in our study area, the harvesting time is examined by its morphological appearance, as mentioned earlier, through the flowers. Sago palm forests commonly live together with other undistinguishable plants, but unfortunately, due to the height of the sago and the limitation of visible inspection by human or satellite images, especially in the sago area of Papua that are part of the overall ecology, sago palms are difficult to identify. After investigating other areas in Indonesia, such as South Sulawesi, which is also typically a sago or semi-cultivated forest, [3] found that the complexity of morphologic appearances, such as the similarity of typical plants, affects the results. Therefore, the result of this current study can help the local community, as well as the stakeholders to recognize the harvest time and the species properly, whether it is sago or other plants. To support this, the deployment of this current study by using appropriate fine-tuning or integrating with other frameworks to address a variety of target problems, as mentioned previously, must be considered in our further research.

5. Conclusions

This study compared the capabilities of three models for sago palm recognition based on their dominant appearances, such as leaves fruits, flowers, and trunks. Each model is transferred from pre-trained deep learning networks by substituted base layers. Likewise, the fully connected layer becomes an fc_new, SoftMax layer, and output layer; to obtain our target model, which is nine labels from nine classes, and the probabilities as well. The experiment's result, as shown in Figures 11, 12 and Table 5, ResNet-50 model was taken as a prior model for flowers, leaves, and trunks for sago palm detection. In further research, this baseline model designed is the first in its field and is expected to obtain a high accuracy, including training validation accuracy up to 90%, with less elapsed time and an improved number of epochs, which also provides more datasets of sago palms. Moreover, since the similarity of sago morphology is influenced by the current result, further work must be integrated with different environments and various sago palm datasets.

Appendix A

UAV and dataset information
Table A1, Figures A1–A3.

Table A1. Technical data of UAV used.

Index	Specification
Dimensions	42.4 × 35.4 × 11 cm
Battery (life and weight)	Li-Ion 7100 mAh 82 Wh; 40 min; 360 g
Video resolution	6K (5472 × 3076)
ISO range	Video-ISO 100-3200 Cr/100-6400 Manual, Photo-ISO100-3200 Car/100-12800 Manual
Camera resolution	20 Mpx; camera chip: 1′ CMOS IMX383 Sony
Maximum flight time	40 min (single charge)
Field of view	82°
Gesture control, Wi-Fi, GPS, controller control, Mobile App, homecoming, anti-collision sensors, automatic propeller stop	Provided
Speeds	72 km/h to 5 km; winds of 62–74 km/h at up to 7000 m above sea level

Figure A1. The drone used in this study: Autel Robotics EVO II Pro 6K.

Appendix B

Figure A2. Sample dataset: training data.

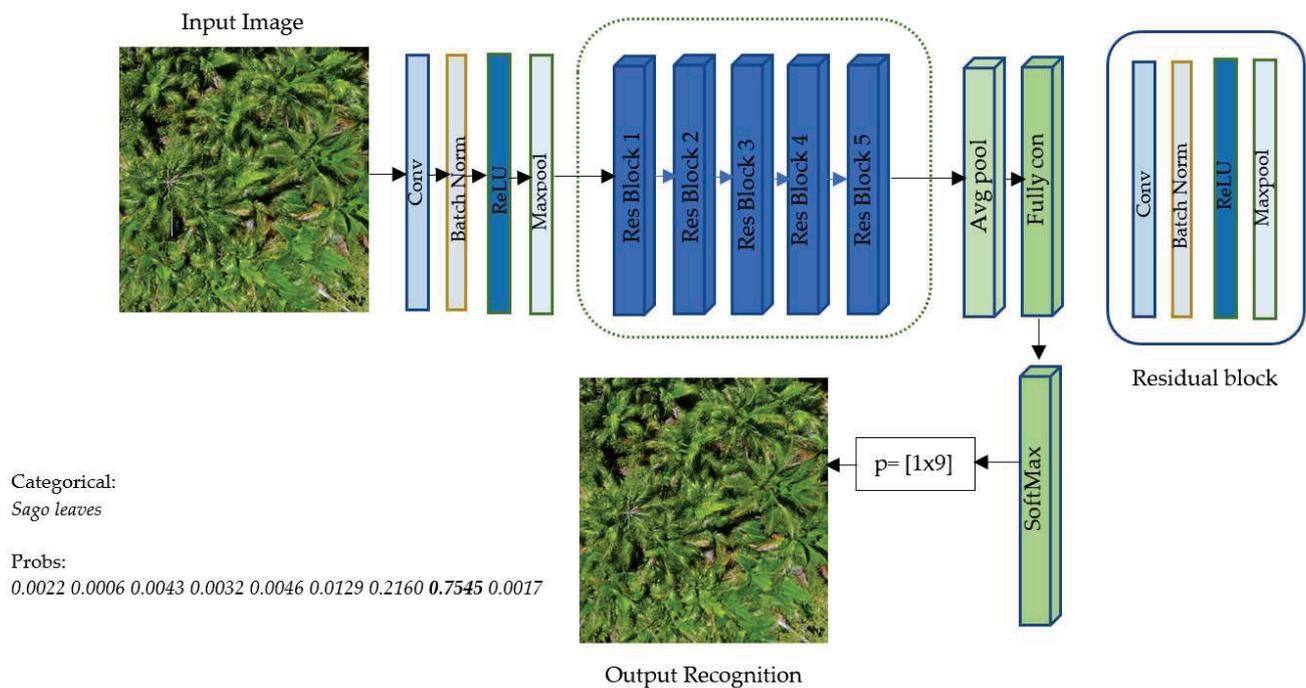**Figure A3.** Sample dataset: testing data.

Figure A4. Model used: ResNet-50 network.

References

1. Chua, S.N.D.; Kho, E.P.; Lim, S.F.; Hussain, M.H. Sago Palm (*Metroxylon sagu*) Starch Yield, Influencing Factors and Estimation from Morphological Traits. *Adv. Mater. Process. Technol.* **2021**, 1–23. [CrossRef]
2. Ehara, H.; Toyoda, Y.; Johnson, D.V. (Eds.) *Sago Palm: Multiple Contributions to Food Security and Sustainable Livelihoods*; Springer: Singapore, 2018; ISBN 978-981-10-5268-2.
3. Hidayat, S.; Matsuoka, M.; Baja, S.; Rampisela, D.A. Object-Based Image Analysis for Sago Palm Classification: The Most Important Features from High-Resolution Satellite Imagery. *Remote Sens.* **2018**, *10*, 1319. [CrossRef]
4. Lim, L.W.K.; Chung, H.H.; Hussain, H.; Bujang, K. Sago Palm (*Metroxylon sagu* Rottb.): Now and Beyond. *Pertanika J. Trop. Agric. Sci.* **2019**, *42*, 435–451.
5. Letsoin, S.M.A.; Herak, D.; Rahmawan, F.; Purwestri, R.C. Land Cover Changes from 1990 to 2019 in Papua, Indonesia: Results of the Remote Sensing Imagery. *Sustainability* **2020**, *12*, 6623. [CrossRef]
6. Jonatan, N.J.; Ekayuliana, A.; Dhiputra, I.M.K.; Nugroho, Y.S. The Utilization of *Metroxylon sago* (Rottb.) Dregs for Low Bioethanol as Fuel Households Needs in Papua Province Indonesia. *KLS* **2017**, *3*, 150. [CrossRef]
7. Nanlohy, L.H.; Gafur, M.A. Potensi Pati Sagu Dan Pendapatan Masyarakat Di Kampung Mega Distrik Mega Kabupaten Sorong. *Median* **2020**, *12*, 21. [CrossRef]
8. Pandey, A.; Jain, K. An Intelligent System for Crop Identification and Classification from UAV Images Using Conjugated Dense Convolutional Neural Network. *Comput. Electron. Agric.* **2022**, *192*, 106543. [CrossRef]
9. Tahir, A.; Munawar, H.S.; Akram, J.; Adil, M.; Ali, S.; Kouzani, A.Z.; Mahmud, M.A.P. Automatic Target Detection from Satellite Imagery Using Machine Learning. *Sensors* **2022**, *22*, 1147. [CrossRef] [PubMed]
10. Kentsch, S.; Cabezas, M.; Tomhave, L.; Groß, J.; Burkhard, B.; Lopez Caceres, M.L.; Waki, K.; Diez, Y. Analysis of UAV-Acquired Wetland Orthomosaics Using GIS, Computer Vision, Computational Topology and Deep Learning. *Sensors* **2021**, *21*, 471. [CrossRef] [PubMed]
11. Mhango, J.K.; Grove, I.G.; Hartley, W.; Harris, E.W.; Monaghan, J.M. Applying Colour-Based Feature Extraction and Transfer Learning to Develop a High Throughput Inference System for Potato (*Solanum tuberosum* L.) Stems with Images from Unmanned Aerial Vehicles after Canopy Consolidation. *Precis. Agric.* **2022**, *23*, 643–669. [CrossRef]
12. Niu, Z.; Deng, J.; Zhang, X.; Zhang, J.; Pan, S.; Mu, H. Identifying the Branch of Kiwifruit Based on Unmanned Aerial Vehicle (UAV) Images Using Deep Learning Method. *Sensors* **2021**, *21*, 4442. [CrossRef]
13. Yang, X.; Zhang, Y.; Lv, W.; Wang, D. Image Recognition of Wind Turbine Blade Damage Based on a Deep Learning Model with Transfer Learning and an Ensemble Learning Classifier. *Renew. Energy* **2021**, *163*, 386–397. [CrossRef]
14. Srivastava, A.; Badal, T.; Saxena, P.; Vidyarthi, A.; Singh, R. UAV Surveillance for Violence Detection and Individual Identification. *Autom. Softw. Eng.* **2022**, *29*, 28. [CrossRef]
15. Liu, Y. Transfer Learning Based Multi-Layer Extreme Learning Machine for Probabilistic Wind Power Forecasting. *Appl. Energy* **2022**, *12*, 118729. [CrossRef]

16. Zhao, W.; Yamada, W.; Li, T.; Digman, M.; Runge, T. Augmenting Crop Detection for Precision Agriculture with Deep Visual Transfer Learning—A Case Study of Bale Detection. *Remote Sens.* **2020**, *13*, 23. [CrossRef]
17. Zheng, J.; Fu, H.; Li, W.; Wu, W.; Yu, L.; Yuan, S.; Tao, W.Y.W.; Pang, T.K.; Kanniah, K.D. Growing Status Observation for Oil Palm Trees Using Unmanned Aerial Vehicle (UAV) Images. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 95–121. [CrossRef]
18. Traore, B.B.; Kamsu-Foguem, B.; Tangara, F. Deep Convolution Neural Network for Image Recognition. *Ecol. Inform.* **2018**, *48*, 257–268. [CrossRef]
19. Altuntaş, Y.; Cömert, Z.; Kocamaz, A.F. Identification of Haploid and Diploid Maize Seeds Using Convolutional Neural Networks and a Transfer Learning Approach. *Comput. Electron. Agric.* **2019**, *163*, 104874. [CrossRef]
20. Omara, E.; Mosa, M.; Ismail, N. Emotion Analysis in Arabic Language Applying Transfer Learning. In Proceedings of the 2019 15th International Computer Engineering Conference (ICENCO), Cairo, Egypt, 29–30 December 2019; pp. 204–209.
21. Xiang, Q.; Wang, X.; Li, R.; Zhang, G.; Lai, J.; Hu, Q. Fruit Image Classification Based on MobileNetV2 with Transfer Learning Technique. In Proceedings of the 3rd International Conference on Computer Science and Application Engineering—CSAE 2019, Sanya, China, 22–24 October 2019; ACM Press: New York, NY, USA, 2019; pp. 1–7.
22. Huang, J.; Lu, X.; Chen, L.; Sun, H.; Wang, S.; Fang, G. Accurate Identification of Pine Wood Nematode Disease with a Deep Convolution Neural Network. *Remote Sens.* **2022**, *14*, 913. [CrossRef]
23. Thenmozhi, K.; Reddy, U.S. Crop Pest Classification Based on Deep Convolutional Neural Network and Transfer Learning. *Comput. Electron. Agric.* **2019**, *164*, 104906. [CrossRef]
24. Shaha, M.; Pawar, M. Transfer Learning for Image Classification. In Proceedings of the 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 29–31 March 2018; pp. 656–660.
25. Gao, C.; Gong, Z.; Ji, X.; Dang, M.; He, Q.; Sun, H.; Guo, W. Estimation of Fusarium Head Blight Severity Based on Transfer Learning. *Agronomy* **2022**, *12*, 1876. [CrossRef]
26. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
27. Minowa, Y.; Kubota, Y. Identification of Broad-Leaf Trees Using Deep Learning Based on Field Photographs of Multiple Leaves. *J. For. Res.* **2022**, *27*, 246–254. [CrossRef]
28. Jin, B.; Zhang, C.; Jia, L.; Tang, Q.; Gao, L.; Zhao, G.; Qi, H. Identification of Rice Seed Varieties Based on Near-Infrared Hyperspectral Imaging Technology Combined with Deep Learning. *ACS Omega* **2022**, *7*, 4735–4749. [CrossRef] [PubMed]
29. Jahandad; Sam, S.M.; Kamardin, K.; Amir Sjarif, N.N.; Mohamed, N. Offline Signature Verification Using Deep Learning Convolutional Neural Network (CNN) Architectures GoogLeNet Inception-v1 and Inception-V3. *Procedia Comput. Sci.* **2019**, *161*, 475–483. [CrossRef]
30. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and <0.5 MB Model Size. *arXiv* **2016**, arXiv:1602.07360.
31. Alom, Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S. The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. *arXiv* **2018**, arXiv:1803.01164.
32. Minaee, S.; Boykov, Y.Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [CrossRef] [PubMed]
33. Yasmeen, U.; Khan, M.A.; Tariq, U.; Khan, J.A.; Yar, M.A.E.; Hanif, C.A.; Mey, S.; Nam, Y. Citrus Diseases Recognition Using Deep Improved Genetic Algorithm. *Comput. Mater. Contin.* **2022**, *71*, 3667–3684. [CrossRef]
34. Zhang, X.; Pan, W.; Xiao, P. In-Vivo Skin Capacitive Image Classification Using AlexNet Convolution Neural Network. In Proceedings of the 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, China, 27–29 June 2018; pp. 439–443.
35. Sun, P.; Feng, W.; Han, R.; Yan, S.; Wen, Y. Optimizing Network Performance for Distributed DNN Training on GPU Clusters: ImageNet/AlexNet Training in 1.5 Minutes. *arXiv* **2019**, arXiv:1902.06855.
36. Izzo, S.; Prezioso, E.; Giampaolo, F.; Mele, V.; Di Somma, V.; Mei, G. Classification of Urban Functional Zones through Deep Learning. *Neural Comput. Appl.* **2022**, *34*, 6973–6990. [CrossRef]
37. Muhammad, W.; Aramvith, S. Multi-Scale Inception Based Super-Resolution Using Deep Learning Approach. *Electronics* **2019**, *8*, 892. [CrossRef]
38. Sarwinda, D.; Paradisa, R.H.; Bustamam, A.; Anggia, P. Deep Learning in Image Classification Using Residual Network (ResNet) Variants for Detection of Colorectal Cancer. *Procedia Comput. Sci.* **2021**, *179*, 423–431. [CrossRef]
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.
40. Bouguettaya, A.; Zarzour, H.; Kechida, A.; Taberkit, A.M. Deep Learning Techniques to Classify Agricultural Crops through UAV Imagery: A Review. *Neural Comput. Appl.* **2022**, *34*, 9511–9536. [CrossRef] [PubMed]
41. Chirra, S.M.R.; Reza, H. A Survey on Software Cost Estimation Techniques. *JSEA* **2019**, *12*, 226–248. [CrossRef]
42. Hai, V.V.; Nhung, H.L.T.K.; Prokopova, Z.; Silhavy, R.; Silhavy, P. A New Approach to Calibrating Functional Complexity Weight in Software Development Effort Estimation. *Computers* **2022**, *11*, 15. [CrossRef]
43. Zhang, C.; Xia, K.; Feng, H.; Yang, Y.; Du, X. Tree Species Classification Using Deep Learning and RGB Optical Images Obtained by an Unmanned Aerial Vehicle. *J. For. Res.* **2021**, *32*, 1879–1888. [CrossRef]

44. Barmpoutis, P.; Dimitropoulos, K.; Barboutis, I.; Grammalidis, N.; Lefakis, P. Wood Species Recognition through Multidimensional Texture Analysis. *Comput. Electron. Agric.* **2018**, *144*, 241–248. [CrossRef]

45. Huang, P.; Zhao, F.; Zhu, Z.; Zhang, Y.; Li, X.; Wu, Z. Application of Variant Transfer Learning in Wood Recognition. *BioRes* **2021**, *16*, 2557–2569. [CrossRef]

46. Xie, W.; Wei, S.; Zheng, Z.; Jiang, Y.; Yang, D. Recognition of Defective Carrots Based on Deep Learning and Transfer Learning. *Food Bioprocess Technol.* **2021**, *14*, 1361–1374. [CrossRef]

47. Diez, Y.; Kentsch, S.; Fukuda, M.; Caceres, M.L.L.; Moritake, K.; Cabezas, M. Deep Learning in Forestry Using UAV-Acquired RGB Data: A Practical Review. *Remote Sens.* **2021**, *13*, 2837. [CrossRef]

48. 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, Miami, FL, USA, 20–25 June 2009.

49. Chakraborty, S.; Mondal, R.; Singh, P.K.; Sarkar, R.; Bhattacharjee, D. Transfer Learning with Fine Tuning for Human Action Recognition from Still Images. *Multimed. Tools Appl.* **2021**, *80*, 20547–20578. [CrossRef]

50. Han, X.; Zhong, Y.; Cao, L.; Zhang, L. Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification. *Remote Sens.* **2017**, *9*, 848. [CrossRef]

51. Chew, R.; Rineer, J.; Beach, R.; O'Neil, M.; Ujeneza, N.; Lapidus, D.; Miano, T.; Hegarty-Craver, M.; Polly, J.; Temple, D.S. Deep Neural Networks and Transfer Learning for Food Crop Identification in UAV Images. *Drones* **2020**, *4*, 7. [CrossRef]

52. Cengil, E.; Cinar, A. Multiple Classification of Flower Images Using Transfer Learning. In Proceedings of the 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Turkey, 21–22 September 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.

53. Feng, J.; Wang, Z.; Zha, M.; Cao, X. Flower Recognition Based on Transfer Learning and Adam Deep Learning Optimization Algorithm. In Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence—RICAI 2019, Shanghai, China, 20–22 September 2019; ACM Press: New York, NY, USA, 2019; pp. 598–604.

Retrieval of Water Quality from UAV-Borne Hyperspectral Imagery: A Comparative Study of Machine Learning Algorithms

Qikai Lu [1,2], Wei Si [1,2,*], Lifei Wei [1,2,3], Zhongqiang Li [1,2], Zhihong Xia [4], Song Ye [5] and Yu Xia [5]

1 Faculty of Resources and Environmental Science, Hubei University, Wuhan 430062, China; luqikai@hubu.edu.cn (Q.L.); weilifei2508@hubu.edu.cn (L.W.); lizhq@hubu.edu.cn (Z.L.)
2 Hubei Key Laboratory of Regional Development and Environmental Response, Hubei University, Wuhan 430062, China
3 Key Laboratory of Urban Land Resources Monitoring and Simulation, MNR, Shenzhen 518034, China
4 Wuhan Regional Climate Center, Wuhan 430074, China; lzy@stu.hubu.edu.cn
5 Changjiang River Scientific Research Institute, Changjiang Water Resources Commission, Wuhan 430010, China; ysyesong@gmail.com (S.Y.); xiayu@mail.crsri.cn (Y.X.)
* Correspondence: 201911110811280@stu.hubu.edu.cn; Tel.: +86-151-7171-4207

Abstract: The rapidly increasing world population and human activities accelerate the crisis of the limited freshwater resources. Water quality must be monitored for the sustainability of freshwater resources. Unmanned aerial vehicle (UAV)-borne hyperspectral data can capture fine features of water bodies, which have been widely used for monitoring water quality. In this study, nine machine learning algorithms are systematically evaluated for the inversion of water quality parameters including chlorophyll-a (Chl-a) and suspended solids (SS) with UAV-borne hyperspectral data. In comparing the experimental results of the machine learning model on the water quality parameters, we can observe that the prediction performance of the Catboost regression (CBR) model is the best. However, the prediction performances of the Multi-layer Perceptron regression (MLPR) and Elastic net (EN) models are very unsatisfactory, indicating that the MLPR and EN models are not suitable for the inversion of water quality parameters. In addition, the water quality distribution map is generated, which can be used to identify polluted areas of water bodies.

Keywords: water quality parameters inversion; machine learning; UAV-borne hyperspectral data; water quality mapping

1. Introduction

Inland water is the most significant freshwater resource for humans. It has various functions, including water storage, irrigation, and power generation. Inland water bodies are usually close to human settlements. They are easily subject to the combined pressures caused by intensive human activities and environmental changes. Pollution, such as agricultural activities, aquaculture, and industrial discharge, will lead to the accumulation of nutrients in the water and will cause eutrophication [1]. The eutrophication will further lead to the occurrence of algal blooms [2]. These destroy the aquatic ecological structure and consume a large amount of dissolved oxygen, leading to hypoxia and cause the death of aquatic animals and plants. Therefore, for the sustainable development of water resources, water quality must be monitored by statutes on a chemical, physical, and biological basis, according to the Environmental Quality Standards for Surface Water (GB3838-2000) in China [3]. Water quality parameters are the most commonly used evaluation measurements to characterize the water quality of inland water bodies.

Chlorophyll-a (Chl-a) concentration and Suspended solids (SS) concentration are the most common water quality parameters. Chl-a is a typical optical active parameter widely existing in algae and cyanobacteria, as well as in other aquatic plants. The concentration of

Chl-a is an indicator for the eutrophication of water bodies based on nutrient availability, quantifying the nutritional status of water bodies [4]. High concentrations of SS will reduce the light transmittance of the water and increase the water-leaving reflectance in the visible wavelength. The concentration of SS in water is also directly related to the migration of pollutants such as heavy metals and organics [5]. Therefore, it is essential to monitor the concentration of SS in the aquatic system.

The conventional methods for water quality monitoring are mainly based on field sampling and laboratory analysis, which are expensive, time-consuming, and labor-intensive [6]. With the development of remote sensing technology, remote sensing has been used as a supplement for traditional methods in aquatic ecosystem monitoring because of its convenient acquisition, long-term dynamic monitoring, and inexpensive qualities. Huang et al. evaluated the spatial variation of Chl-a using MODIS data for different river flow conditions [7]. Doña et al. predicted and assessed the dynamics the diversification of Chl-a and transparency using MODIS, TM, and ETM+ data [8]. Du et al. investigated the tempo-spatial dynamics pattern of water quality in the Taihu Lake estuary using GOCI imagery [9]. Syariz et al. used spectral and spatial information from Sentinel-3 images to retrieval the concentration of Chl-a [10]. Rajesh et al. predicted the heavy metal concentration in water including Arsenic (As), cadmium (Cd), chromium (Cr), copper (Cu), iron (Fe), lead (Pb), nickel (Ni), zinc (Zn), aluminum (Al), cobalt (Co), manganese (Mg), beryllium (Be), boron (B), lithium (Li), molybdenum (Mo), selenium (Se), and vanadium (V), using Cartosat-2 data and measuring data [11]. Rostom et al. predicted and assessed the concentration of heavy metals including Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd, and Pb in water using hyperspectral remote sensing data [12]. The studies mentioned above have researched the application of satellite remote sensing data for water quality monitoring. However, it is challenging to assess water quality in inland water bodies with the coarse spatial resolution and spectral resolution of satellite data. UAV-borne remote sensing hyperspectral data with high spatial resolution, spectral resolution, and timeliness is superior to satellite data. Therefore, the water quality parameters' concentration can be predicted and retrieved using UAV-borne hyperspectral data.

The spectral reflectance information extracted from remote sensing imagery can be used to estimate the concentrations of water quality parameters. Meanwhile, artificial intelligence technology is widely used in the water quality monitoring field in recent years. Quan et al. used a genetic algorithm (GA) to optimize the parameters of the support vector machine regression (SVR) model for the prediction of vertical water temperature and water temperature structure [13]. Leong et al. used the support vector machine (SVM) and least-squares support vector machine (LS-SVM) to predict water quality parameters including dissolved oxygen (DO), biological oxygen demand (BOD), chemical oxygen demand (COD), suspended solids (SS), pH value (PH), and ammoniacal nitrogen (AN) [14]. Lu et al. used the extreme gradient boosting (XGBoost) model and random forest (RF) model to predict water quality parameters including temperature, DO, specific conductance, pH value, turbidity, and fluorescent dissolved organic matter [15]. Najah Ahmed et al. proposed a neuro-fuzzy inference system (ANFIS)-based augmented wavelet de-noising technique (WDT) to predict water quality parameters including AN, SS, and pH [16]. Sharafati et al. used adaboost regression (ABR), gradient boost regression (GBR), and random forest regression (RF) to predict water quality parameters including total dissolved solids (TDS), five-day biochemical oxygen demand (BOD_5), and COD on a daily scale [17]. Parsimehr et al. used the multilayer perceptron artificial neural network to predict and simulate the COD of the Gamasiab river [18]. Xiaojuan et al. used ensemble learning models based on four models, including the k-nearest neighbors (KNN), artificial neural network (ANN), SVR, and RF, to retrieve Chl-a and TN in water [19]. Although researchers have conducted numerous studies, there is still a lack of analyses on different machine learning algorithms using UAV-borne hyperspectral data.

The machine learning model can quickly acquire the required information from the data but different models have different characteristics. Therefore, the main aims of this

study are (1) to compare and analyze the prediction performance of different machine learning models, including CBR, Adaboost regression (ABR), extreme boost regression (XGBR), random forest (RF), extremely randomized trees (ERT), support vector regression (SVR), MLPR, and EN, using the evaluation index including R^2, RMSE, and MAE; (2) to use typical water quality parameters, including Chl-a and SS, to evaluate the machine learning model; and (3) to verify the potential of machine learning models combined with UAV-borne hyperspectral data in water quality mapping.

2. Materials and Methods

2.1. Study Area

The study area was Beigong Reservoir (114°21′14.15″ E, 30°35′5.52″ N) with a catchment area of 6.8 km^2 and a storage capacity of 12.2×10^6 m^3, which is the primary supply source for the tributary of the Liu River, namely the Pearl River basin of the Xijiang river system. Beigong Reservoir is located in the Beigong village southwest of Liuzhou, Guangxi Zhuang Autonomous Region, China. Additionally, Liuzhou is an essential industrial city in Guangxi and is key to the "three wastes" emissions of exhaust gas, wastewater, and waste residue. Moreover, the Liuzhou Municipal Government attaches great importance to the local water pollution problem. Beigong Reservoir is a medium-sized reservoir, utilized for irrigation, flood control, and power generation. It is a famous tourist attraction with a beautiful scenery surrounded by mountains that integrates entertainment, tourism, and life functions. Hence, the Beigong Reservoir is of great significance for the residents and policymakers.

2.2. Data Collection

Liuzhou's summers are long, hot, sultry, humid, and cloudy, and the winters are short, cold, and mostly sunny. During the year, the temperature usually varies between 6 °C and 33 °C, and is rarely below 2 °C or above 36 °C. Therefore, the best time to conduct a water quality sampling survey is from the last ten-day period of September to October. Therefore, field investigations were uniformly conducted in Beigong Reservoir from 9 to 10 September 2018. On the basis of the field data collection regulations, 33 sample points at Beigong Reservoir were collected for Chl-a and SS inversion. The field sampling data were analyzed in the laboratory. The statistical information of the water sampling data is shown in Table 1. Detailed information about the Beigong Reservoir is presented in Figure 1, in which the water quality sampling points, ground water surface spectrum point, the UAV-flight routes, and the obtained UAV-borne flight strips are shown.

Table 1. Statistical information of the water sampling data in Beigong Reservoir including Chl-a (mg/m^3) and SS (mg/L).

Water Quality Parameters	Range	Average Value	Standard Deviation
Chl-a (n = 33)	3.54~14.2 (mg/m^3)	8.03	2.32
SS (n = 33)	2~18 (mg/L)	5.86	4.54
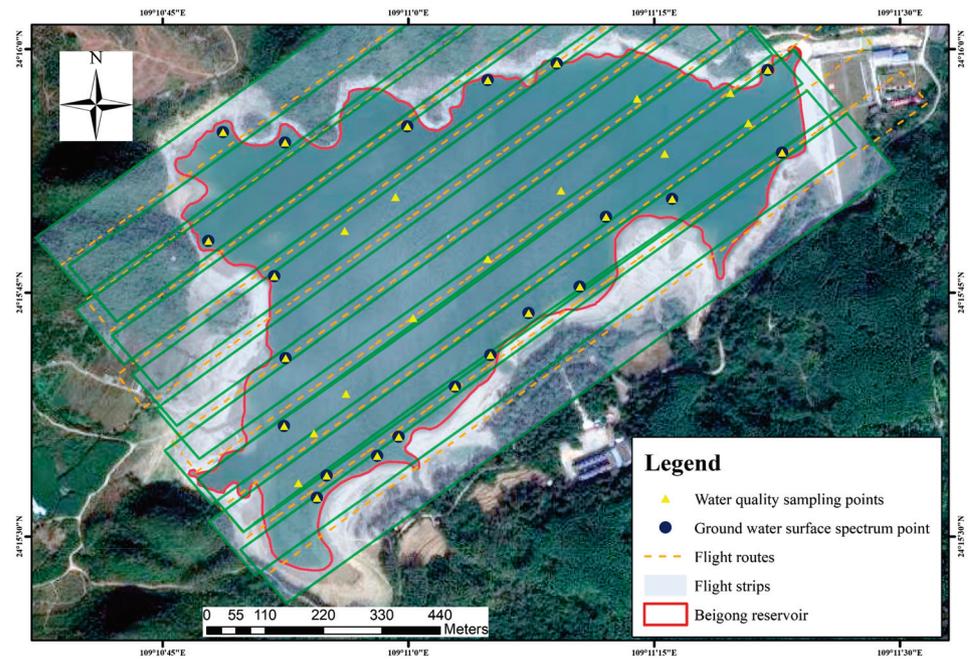
Figure 1. The study area and sampling information.

The ground water surface spectral reflectance data will be discussed next. The ground water surface spectrum was collected by the ASD FieldSpec 3 field-portable spectrometer with a wavelength range of 350–2500 nm. The spectrometer was provided by the China University of Geosciences (Wuhan, China). The measurement of the ground water surface spectrum was based on the "above-water surface method" [20]. The reference board with a reflectivity of nearly 1 was used for radiometric calibration. In windless weather, the water surface was flat. The water surface spectrum, sky spectrum, and synchronously the spectral data of the reference board were collected. The reference board was utilized to perform the calibration of the water surface spectrum and to obtain the water-leaving reflectance data. The ground water surface spectral reflectance collection was repeated three times in situ and the average value of the three collected data was used as the final reflectance data. The total radiance received by the spectrometer can be expressed as:

$$L\_sw = L_w + rL_{sky} \qquad (1)$$

where L_sw is the total radiance received by the spectrometer. L_w is the water-leaving radiance. L_{sky} is the diffuse radiance of the sky. r is the air–water interface reflectance rate. When the water surface is flat, r can be set as 0.022; when the wind speed is about 5 m/s, r can be set as 0.025; and when the wind speed is about 10 m/s, r can be set as 0.026–0.028. L_w can be expressed as follows [21]:

$$R_{rs} = \frac{L_w}{E_d(0^+)} = \frac{(L\_sw - rL_{sky})}{\pi L_p}\rho_p \qquad (2)$$

$$E_d(0^+) = L_p \frac{\pi}{\rho_p} \qquad (3)$$

where $E_d(0^+)$ is the incident total irradiance. L_p is the 100% converted value of the reference board. R_{rs} is the water-leaving reflectance.

The UAV-borne hyperspectral data will be discussed next. We adopted the six-rotor DJ M600 Pro UAV as the airborne platform and the sensor installed on it was the Headwall NANO-Hyperspec manufactured by Headwall Photonics Lnc. The spectral resolution was 6.0 nm [22,23]. The resampling interval was set to 2.2 nm, which is the sensor parameter. At the UAV flight process, the field of view was 16° and the flying height of the UAV was 400

m with a real-time wind speed of 5.2 m/s. According to the area of the reservoir, 10 routes of flight have been designed, where the along-track overlap was 80% and the side overlap was 60% [24]. With 270 spectral bands in the range of 400–1000 nm, the spatial resolution of the hyperspectral imagery was 0.173 m/pixel. Due to the low flight altitude, atmospheric influence can be ignored [25]. The UAV-borne hyperspectral image preprocessing was conducted, including water body extraction, sensor calibration, geometric correction, and in situ radiation correction. Firstly, the normalized difference water index (NDWI) was used to extract the water information in the UAV-borne image [26–28]. Secondly, we performed geometric correction for the image. The NANO-Hyperspec hyperspectral imaging spectrometer has a global positioning system and an inertial measurement unit (GPS/IMU) navigation system that can contribute to geometrically correcting the image by recording the position and attitude information of the spectrometer. Thirdly, regarding the calibration of the sensors, the signal output by the sensor unit was converted into the actual radiation intensity value. Finally, by constructing the linear relationship between the pixel spectrum of the UAV hyperspectral image and the ground water spectrum, the in situ radiation calibration was performed [29]. The water quality sampling data, the ground water surface spectral reflectance data, and UAV-borne hyperspectral data were all collected at the same time.

2.3. Method

2.3.1. Machine Learning Algorithms Used for the Estimation of Water Quality Parameters Adaboost Regression (ABR)

ABR is a typical boosting algorithm introduced by Freund [30]. ABR trains the weak learners and then integrates the trained weak learners to obtain a final model [31]. ABR assigns different weights to each sample according to the prediction error rate of the learner, then adjusts the weight of the sample, and finally accumulates and weights the prediction results of all learners to generate a predicted value.

Gradient Boost Regression tree (GBRT)

GBRT is a machine learning algorithm based on ensemble decision trees [32], which is the regression form of gradient boost decision trees (GBDT). The GBRT model first builds a regression tree with equal weights based on the original data. It evaluates the prediction results by minimizing the square error. The smaller the mean square error, the lower the weight of the decision tree. The GBRT model uses the negative gradient of the loss function in the current model to approximate the residual between the current model's predicted value and the observed value, so that the model optimizes the weight of the regression tree along the direction of the negative gradient of the loss function. In each round of the training process, the model reduces the loss function and accelerates the convergence to reach the local optimal solution or the global optimal solution. Through continuous iteration, the predicted values of all the regression trees are combined to obtain the final prediction result.

Extreme Gradient Boosting Regression (XGBR)

XGBR is an improved decision tree algorithm based on the GBDT algorithm [33]. The core of the algorithm is to continuously add and train new decision trees to fit the residuals of the previous iteration [34] and the prediction values of all the decision trees are accumulated to obtain the final prediction result. XGBR improves prediction performance by reducing model bias. Compared with the traditional GBDT algorithm, XGBR modifies the objective function of the GBDT algorithm. The formula of the loss function is defined as follows:

$$\text{Loss}^{(t)} = \sum_{i=1}^{T} [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \tag{4}$$

$$g_i = \partial_{\hat{y}^{(t-1)}} l(y_{i,} \hat{y}_i^{(t-1)}), h_i = \partial_{\hat{y}^{(t-1)}}^2 l(y_{i,} \hat{y}_i^{(t-1)}) \tag{5}$$

$$\Omega(f_{\text{t}}) = \gamma T + \frac{1}{2}\lambda \sum_{n=1}^{T} w_n^2 \tag{6}$$

where g_i and h_i are the first and second partial derivatives of the loss function. $\hat{y}_i$ is the predicted value of the model, while y_i is the observed value. $f_{\text{t}}(x_i)$ represents the score of the i-th sample in the t-th decision tree, $\Omega(f_{\text{t}})$ is the regular term of the model, l represents the number of trees, γ represents the complexity of the leaves, T represents the number of the leaves,λ represents the scaler factor, and w_n represents the weight of the n-th leaf node in the tree.

After removing the constant term, the final objective function can be expressed as:

$$\text{Loss} = -\frac{1}{2}\sum_{j=1}^{T} \frac{G_n^2}{H_n + \lambda} + \gamma T \tag{7}$$

where $\frac{G_n^2}{H_n+\lambda}$ represents the contribution of each leaf node to the current model loss function. XGBR uses a greedy algorithm to traverse all split leaf nodes in the model. When the gain of the target after the split is less than the set threshold, we can ignore the split.

$$\text{Gain} = \frac{1}{2}\Big[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\Big] - \gamma > 0 \tag{8}$$

where $\frac{G_L^2}{H_L+\lambda}$ represents the left subtree score, $\frac{G_R^2}{H_R+\lambda}$ represents the right subtree score, and $\frac{(G_L+G_R)^2}{H_L+H_R+\lambda}$ represents the score when it is not divided.

Catboost Regression (CBR)

CBR is a new decision tree based on a gradient boosting frame [35] and uses oblivious decision trees as a based learner. Oblivious trees use the same criteria for splitting at each level of the tree. Each leaf index can be encoded as a binary vector whose length is equal to the depth of the tree, which helps to avoid overfitting and speed up the prediction of the model. CBR differs from the traditional GBDT algorithm in three aspects: (1) CBR can efficiently process categorical features. The categorical feature refers to a category or label. Unlike other numerical characteristics, the numerical variables of categorical characteristics cannot be compared with other numerical variables. Therefore, they are also called non-ordered features. Discrete numbers are also categorical features. In the traditional GBDT algorithm, when the structure and distribution of the training data set and the test data set are different, the conditional offset problem will appear. Furthermore, CBR uses the improved greedy target statistics method to add prior distribution items to reduce the influence of noise and low-frequency data on the data distribution. For regression, prior items can take the mean of the data set label. (2) When CBR constructs a new split node for the current tree, it uses a greedy method to consider all combinations which combine different types of features into new features and dynamically transform the new composite categorical features into numeric features. (3) CBR replaces the gradient estimation method in the traditional algorithm by ordered boosting, which helps to overcome the prediction shift caused by gradient bias.

Random Forest (RF)

RF uses the bootstrap method to randomly select n samples from the original data to construct a decision tree. Each sample has M attributes. In the node split of the decision tree, m attributes are randomly selected from the M attributes using the information gain method, where in the attribute with the largest gain is selected as the best split attribute of the node. Then, the prediction results of multiple decision trees are averaged to obtain the final prediction result [36].

Extremely Randomized Trees (ERT)

The structure of the ERT [37] is similar to the RF. The difference between the ERT and RF is that the ERT uses all the samples to construct a decision tree in the training process. For node splitting, the RF algorithm selects the best attribute split, while the ERT randomly selects the attribute split [38], which results in the size of the generated decision tree being larger than that generated by the RF model. Therefore, the variance of the ERT model is reduced compared to the RF model.

Support Vector Machine (SVM)

SVR is a kernel-based algorithm that improves the model's generalization ability by seeking the minimum structured risk and realizing the experience risk minimization. SVR can obtain good prediction results with a small sample size [39].

Multi-Layer Perceptron Regression (MLPR)

MLPR is the most commonly used artificial neural network model, which is composed of three types of layers: an input layer, an output layer, and one or more hidden layers with activation functions [40]. It uses a subset of the training set to adjust the weight and biases on each node of layers. MLPR takes input data, multiplies them with weights, and then inputs them into the activation function to produce final results. MLPR can obtain non-linear relationships and real-time learning. However, MLPR requires many hyperparameters to be adjusted, which is time-consuming.

Elastic Net (EN)

EN is a mixture of the Lasso regression (LR) and the Ridge regression (RR) [41], and the optimization objective function of elastic net regression is defined as follows:

$$\underset{\beta \in R^p}{\mathrm{argmin}}\{\| y - X\beta \|^2 + \lambda[(1-\alpha) \| \beta \|_2 + \alpha \| \beta \|_1]\} \tag{9}$$

where $\| \cdot \|_2$ represents the L2 norm and $\| \cdot \|_1$ represents the L1 norm. The EN regression penalty function uses the convex combination of the L1 norm and L2 norm, which is equally the convex combination of the RR penalty function and LR penalty function. Therefore, the EN has the advantages of both RR and LR, which not only achieves the purpose of variable selection but also improves the stability of the model. It automatically selects variables and retains important features, as well as eliminates irrelevant features.

2.3.2. Model Evaluation

In this study, we evaluate the models' performance via three indicators, including the coefficient of determination (R^2), root mean square errors (RMSE) and mean absolute error (MAE). These indicator metrics can be calculated as follows:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y}_i)^2} \tag{10}$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{11}$$

$$MAE(y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{12}$$

where y_i is the observed value, $\overline{y}_i$ is the average of the observed value, and N is the number of valid samples used for the evaluation. $\hat{y}_i$ is the predicted value. The value of R^2 ranges from 0 to 1. An R^2 score of 1 indicates perfect precision, while a score of 0 indicates that the model has the worst prediction performance. The value range of RMSE is $(0,+\infty)$. If the dispersion of the predicted value of the model is high, the RMSE will be enlarged. MAE is

the mean of the absolute value of the error between the predicted value and the observed value. The value range of MAE is $(0,+\infty)$. Model with high R^2, low RMSE, and low MAE is deemed as a suitable model for quantitative inversion.

The data set was divided into training and validation sets using random split sampling. In total, 80% of the inputting data was used for training the model and 20% of the inputting data was used to assess the prediction accuracy of the model. In this study, all the above model operations were based on the anaconda platform and the modeling of water quality parameters with the ABR, GBRT, RF, ERT, SVM, MLPR, and EN algorithms was implemented with the scikit-learn 0.23.2 machine learning library on the anaconda platform. The XGBR and CBR algorithms were implemented by the xgboost and catboost libraries, respectively.

3. Results

3.1. Spectral Analysis

The spectral signature of Chl-a is characterized by strong absorption in the blue (443 nm) and red wavelengths (near 675 nm), and by high reflectance in the green (550–555 nm) and red-edge spectrum regions (685–710 nm) [42–45]. Existing studies have shown that the most suitable spectral range for monitoring suspended solids in water is 700–800nm [46]. Therefore, the 181 bands with 400 to 800 nm were used for the determination of water constituents in this study. The band ratio method can eliminate background noise and rough water surface interference, and can enhance the fine spectrum characteristics that are beneficial for water quality parameter estimation [47]. This study used the band ratio method to preprocess the original spectrum. The exhaustive method was used to calculate the ratio of the bands. To identify and select the major wavelengths for the estimation of water quality parameters' concentration, we conducted the Pearson correlation analysis.

The spectral curves of the 33 samples in the preprocessed UAV-borne hyperspectral remote sensing imagery are shown in Figure 2. The correlation coefficient between the raw spectral data and the Chl-a concentration value was negative, ranging from 0 to -0.7952. The 84 spectral absolute correlation coefficients were greater than 0.7, which were mainly concentrated in the 400–590 nm wavelength range. The correlation coefficient between the raw spectral data and the SS concentration values ranged from 0.1755 to 0.7685. The 46 spectral correlation coefficients were greater than 0.7, which were mainly concentrated in the 700~800 nm wavelength range.
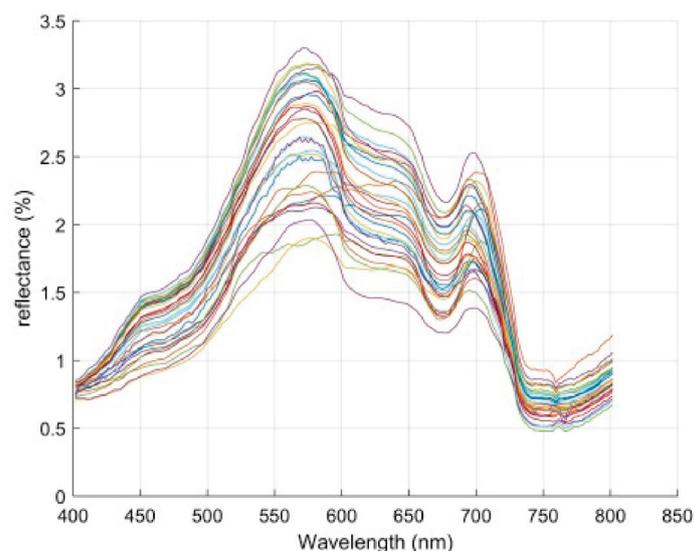
Figure 2. The spectral curves of the 33 samples in the preprocessed UAV-borne hyperspectral remote sensing imagery.

This study calculated the ratio of 181 bands of the original spectra by using the exhaustive method and obtained 32580 characteristic variables. The correlation coeffi-

cients between the characteristic variables and the Chl-a concentration values ranged from −0.8196 to 0.8243. There were 12 ratio variable correlation coefficients greater than 0.81. These variables' spectra were mainly concentrated in the 400~480 nm wavelength range. The band ratio preprocessing method improves the absorption valley characteristics of the original Chl-a spectral. The correlation of processed spectra is significantly improved compared to that of the original spectra. The correlation coefficients between the characteristic variables and the SS concentration values ranged from −0.7653 to 0.7823. There were 43 ratio variable correlation coefficients greater than 0.76. These variables' spectra were mainly concentrated in the 592~606 nm wavelength range. When compared to the raw spectra, the correlation of the processed spectra is considerably enhanced.

3.2. Hyperparameters for the Machine Learning Algorithms

The model performance is influenced by its hyperparameters when estimating the concentration of water quality parameters. Tuning the hyperparameters is a critical step before the quantitative inversion. The key adjusting hyperparameters and their optimal parameter values for each model are shown in Tables 2 and 3.

Table 2. Tuned hyperparameters and their settings for each model in the prediction of Chl-a.

Models	Hyperparameters	Meanings	Search Ranges	Optimal Values
CBR	Learning rate	Shrinkage coefficient of each tree	(0.01,1)	0.26
	Max depth	Maximum depth of a tree	(1,10)	4
	Estimators	Number of trees	(100,1000)	140
	$L_2\_leaf\_reg$	L_2 regularization	(1,10)	7
ABR	Learning rate	Shrinkage coefficient of each tree	(0.01,0.1)	0.1
	Estimators	Number of trees	(0,200)	90
XGBR	Learning rate	Shrinkage coefficient of each tree	(0.01,1)	0.1
	Max depth	Maximum depth of a tree	(1,10)	5
	Estimators	Number of trees	(10,100)	40
	Subsample	Subsample ratio of training samples	(0.1,0.9)	0.8
GBRT	Learning rate	Shrinkage coefficient of each tree	(0.010.1)	0.05
	Estimators	Number of trees	(10,200)	150
	Subsample	Subsample ratio of training samples	(0.5,0.8)	0.5
RF	Estimators	Number of trees	(1,100)	70
	Min samples split	Minimum number of samples for nodes' split	(1,10)	2
ERT	Max depth	Maximum depth of a tree	(1,10)	6
	Estimators	Number of trees	(0,300)	200
SVR	C	Regularization parameter	(10,200)	10
	Gamma	Kernel coefficient	(0.001,0.1)	1
MLPR	Hidden layer size	The number of neurons in the ith hidden layer and the number of hidden layers	$(2^1), (2^{10})$	(2^2)
EN	Alpha	The constant of the mixed penalty term	(0.0001,0.001,0.01,0.1,1,10)	0.1

Table 3. Tuned hyperparameters and their settings for each model in the prediction of SS.

Models	Hyperparameters	Meanings	Search Ranges	Optimal Values
CBR	Learning rate	Shrinkage coefficient of each tree	(0.01,1)	0.1
	Max depth	Maximum depth of a tree	(1,10)	7
	Estimators	Number of trees	(100,1000)	190
	L_2_leaf_reg	L_2 regularization	(1,30)	26
ABR	Learning rate	Shrinkage coefficient of each tree	(0.01,0.1)	0.02
	Estimators	Number of trees	(0,200)	120
XGBR	Learning rate	Shrinkage coefficient of each tree	(0.01,1)	0.05
	Max depth	Maximum depth of a tree	(1,10)	2
	Estimators	Number of trees	(100,1000)	200
GBRT	Learning rate	Shrinkage coefficient of each tree	(0.01,0.1)	0.04
	Estimators	Number of trees	(10,200)	100
	Subsample	Subsample ratio of training samples	(0.5,0.9)	0.8
RF	Estimators	Number of trees	(1,100)	7
	Min samples split	Minimum number of samples for nodes' split	(1,10)	3
ERT	Max depth	Maximum depth of a tree	(1,10)	4
	Estimators	Number of trees	(10,100)	20
SVR	C	Regularization parameter	(10,200)	100
	Gamma	Kernel coefficient	(0.001,10)	1
MLPR	Hidden layer size	The number of neurons in the ith hidden layer and the number of hidden layers	$(2^1,2^1)$, $(2^8,2^8)$	$(2^3,2^3)$
EN	Alpha	The constant of the mixed penalty term	(0.0001,0.001,0.01,0.1,1,10)	0.1

The number and types of hyperparameters in each model are different. Selecting key parameters of the model can reduce the training time of the model and improve the prediction efficiency of the model. Since CBR, ABR, GBRT, XGBR, RF, and ERT are all tree-based models, the number of trees seriously affects the performance of the model. Too many trees will cause over-fitting of the model and an insufficient number of trees will cause underfitting. The L_2 regularization can control the complexity of the model and reduce the generalization error. The GBRT model optimizes the subsample, which controls the random sampling ratio of each tree. If the subsample value is set too small, the result will be underfitting, thus the subsample value range was between 0.5 and 1. For the ERT model, the tree's depth is tuned. The tree's depth determines how the model learns the characteristics of individual samples. The more individual sample features are learned, the worse the generalization ability of the model. In the SVR model, the default radial basis function was selected as the kernel function, tuning the parameters of C and gamma, which control the trade-off between the slack variable penalty and the marginal width. For the MLPR model, we tuned the parameters of the hidden layer size, which include the number of hidden layers and the number of perceptrons contained in each hidden layer.

3.3. Retrieval Results for Different Water Quality
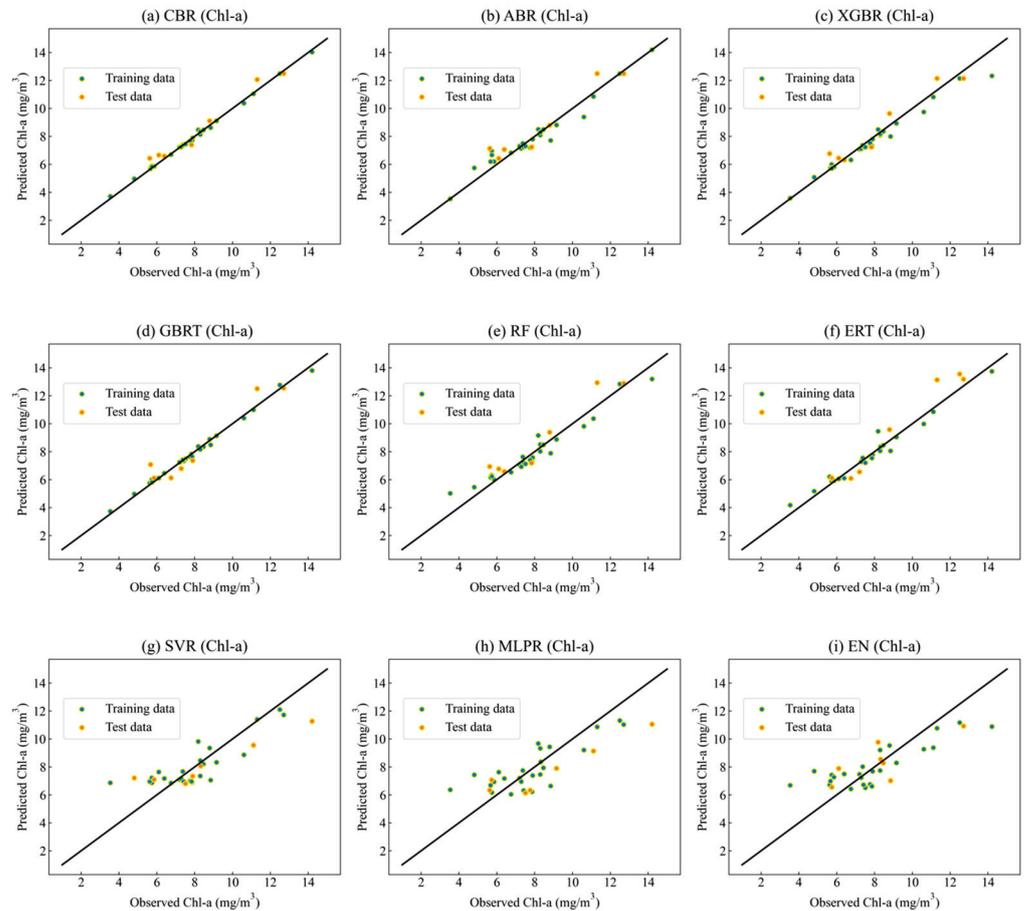
3.3.1. Retrieval Results for Chl-a

The retrieval results of each model for Chl-a are shown in Table 4. We can observe that the CBR model had the best prediction performance (R^2 = 0.96, MAE = 0.53 mg/m^3, RMSE = 0.96 mg/m^3). The prediction accuracy for the CBR validation data set (R^2 = 0.96, MAE = 0.53 mg/m^3, RMSE = 0.96 mg/ m^3) was lower than the prediction accuracy for its training data set (R^2 = 1.00, MAE = 0.09 mg/ m^3, RMSE = 0.12 mg/m^3). For the XGBR validation data set, the prediction accuracy for the XGBR model (R^2 = 0.92, MAE = 0.63 mg/m^3, RMSE = 0.71 mg/m^3) was lower than the CBR model. The prediction accuracy for the GBRT validation data set (R^2 = 0.90, MAE = 0.67 mg/m^3, RMSE = 0.80 mg/m^3) was significantly lower than the prediction accuracy for its training data set (R^2 = 0.99, MAE = 0.13 mg/m^3, RMSE = 0.16 mg/m^3). The prediction accuracy for the ABR on the training data set (R^2 = 0.94, MAE = 0.37 mg/m^3, RMSE = 0.54 mg/m^3) was also significantly higher than the prediction accuracy for its validation data set (R^2 = 0.89, MAE = 0.65 mg/m^3, RMSE = 0.82 mg/m^3). The prediction accuracy for the ERT training data set (R^2 = 0.96, MAE = 0.17 mg/m^3, RMSE = 0.30 mg/m^3) was higher than the prediction accuracy for its validation data set (R^2 = 0.87, MAE = 0.84 mg/m^3, RMSE = 0.95 mg/m^3). The prediction accuracy for the RF training data set (R^2 = 0.93, MAE = 0.48 mg/m^3, RMSE = 0.58 mg/m^3) was higher than the prediction accuracy for its validation data set (R^2 = 0.87, MAE = 0.75 mg/m^3, RMSE = 0.91 mg/m^3). The ERT model was higher than the RF model in terms of RMSE and MAE, while the ERT model and the RF model gave similar R^2. The prediction performance for the SVR validation data set (R^2 = 0.68, MAE = 1.32 mg/m^3, RMSE = 1.52 mg/m^3) was comparable to the prediction performance for its training data set (R^2 = 0.67, MAE = 0.93 mg/m^3, RMSE = 1.27 mg/m^3). The prediction accuracy for the MLPR training data set (R^2 = 0.63, MAE = 1.08 mg/m^3, RMSE = 1.29 mg/m^3) was similar to the prediction accuracy for its validation data set (R^2 = 0.62, MAE = 1.60 mg/m^3, RMSE = 1.75 mg/m^3). The prediction accuracy for the EN training data (R^2 = 0.63, MAE = 1.18 mg/m^3, RMSE = 1.43 mg/m^3) was significantly decreased than the prediction accuracy for its validation data set (R^2 = 0.58, MAE = 1.17 mg/m^3, RMSE = 1.36 mg/m^3). The prediction performance of SVR, MLPR, and EN was too poor for the inversion of the concentration of Chl-a. Since these models, namely SVR, MLPR, and EN, have low R^2 with high MAE and RMSE, it is determined that they are not suitable for Chl-a inversion.

Table 4. Experimental results of Chl-a (mg/m^3) using different models.

Models	Running Time (s)	Training Data Set			Test Data Set		
		MAE (mg/m^3)	RMSE (mg/m^3)	R^2	MAE (mg/m^3)	RMSE (mg/m^3)	R^2
CBR	0.46	0.09	0.12	1.00	0.47	0.53	0.96
ABR	0.15	0.37	0.54	0.94	0.65	0.82	0.89
XGBR	0.47	0.31	0.49	0.95	0.63	0.71	0.92
GBRT	0.06	0.13	0.16	0.99	0.67	0.80	0.90
RF	0.11	0.48	0.58	0.93	0.75	0.91	0.87
ERT	0.17	0.30	0.41	0.96	0.84	0.95	0.87
SVR	0.01	0.92	1.17	0.69	1.38	1.65	0.69
MLPR	0.66	1.08	1.29	0.63	1.60	1.75	0.62
EN	0.01	1.18	1.43	0.63	1.17	1.36	0.58

Scatter plots of the observed and predicted values of the nine machine learning algorithms are presented in Figure 3. The predicted and observed values of the models were evenly distributed on both sides of the regression line, indicating that the models' prediction accuracies are excellent. The difference between the predicted values and the observed values represents the level of the model's prediction deviation, indicating that the model may be overfitting or underfitting. From Figure 3, we can observe that the predicted values of the tree-based ensemble models, including CBR, ABR, XGBR, GBRT, RF, and ERT,

were close to the regression line. Among them, the CBR, XGBR, and GBRT models have the highest prediction accuracy. The difference between the observed and the predicted values of the SVR, MLPR, and EN models is large, indicating that these models' prediction accuracies are extremely poor.

Figure 3. Scatter plot of the observed values and predicted values of Chl-a concentration (mg/m^3) using nine machine learning models including CBR, ABR, XGBR, GBRT, RF, ERT, SVR, MLPR, and EN.

3.3.2. Retrieval Results for SS

The retrieval results of each model for SS are shown in Table 5. We can find that the CBR model had the best prediction performance (R^2 = 0.94, MAE = 1.11 mg/L, RMSE = 1.2 mg/L) in estimating the concentration of SS. The prediction accuracy for the CBR validation data set (R^2 = 0.94, MAE = 1.11 mg/L, RMSE = 1.2 mg/L) was lower than the prediction accuracy for its training data set (R^2 = 0.95, MAE = 0.81 mg/L, RMSE = 1 mg/L). The prediction performance for the RF validation data set (R^2 = 0.93, MAE = 1.23 mg/L, RMSE = 1.39 mg/L) was comparable to its training data set (R^2 = 0.93, MAE = 0.86 mg/L, RMSE = 1.11 mg/L). The prediction accuracy for the XGBR validation data set (R^2 = 0.93, MAE = 1.50 mg/L, RMSE = 1.64 mg/L) was lower than the training data set (R^2 = 0.99, MAE = 0.22 mg/L, RMSE = 0.29 mg/L). The prediction accuracy for the GBRT validation data set (R^2 = 0.91, MAE = 1.22 mg/L, RMSE = 1.48 mg/L) was significantly lower than the prediction accuracy for its training data set (R^2 = 0.99, MAE = 0.39 mg/L, RMSE = 0.44 mg/L). The prediction accuracy for the ABR training data set (R^2 = 0.91, MAE = 1.37 mg/L, RMSE = 1.85 mg/L) was similar to its validation data set (R^2 = 0.92, MAE = 0.81 mg/L, RMSE = 1.10 mg/L). The prediction accuracy for the ERT validation data set (R^2 = 0.90, MAE = 1.41 mg/L, RMSE = 1.54 mg/L) was lower than its training data set (R^2 = 0.93, MAE = 0.85 mg/L, RMSE = 1.17 mg/L). The prediction accuracy for the SVR validation data set (R^2 = 0.89, MAE = 1.42 mg/L, RMSE = 1.57 mg/L) was also lower than its training

data set (R^2 = 0.90, MAE = 0.84 mg/L, RMSE = 1.44 mg/L). The prediction accuracy for the MLPR validation data set (R^2 = 0.59, MAE = 2.31 mg/L, RMSE = 2.95 mg/L) was lower than its training data set (R^2 = 0.63, MAE = 2.09 mg/L, RMSE = 2.78 mg/L). The prediction accuracy for the EN validation data set (R^2 = 0.55, MAE = 2.46 mg/L, RMSE = 2.97 mg/L) was worse than its training data set (R^2 = 0.60, MAE = 2.12 mg/L, RMSE = 2.91 mg/L).

Table 5. Experimental results of SS (mg/L) using different models.

Models	Running Time (s)	Training Data Set			Test Data Set		
		MAE (mg/L)	RMSE (mg/L)	R^2	MAE (mg/L)	RMSE (mg/L)	R^2
CBR	0.31	0.81	1.00	0.95	1.11	1.20	0.94
ABR	0.15	0.81	1.10	0.92	1.37	1.85	0.91
XGBR	0.51	0.22	0.29	0.99	1.50	1.64	0.93
GBRT	0.06	0.39	0.44	0.99	1.22	1.48	0.91
RF	0.02	0.86	1.11	0.93	1.23	1.39	0.93
ERT	0.03	0.85	1.17	0.93	1.41	1.54	0.90
SVR	0.12	0.84	1.44	0.90	1.42	1.57	0.89
MLPR	1.80	2.09	2.78	0.63	2.31	2.95	0.59
EN	0.01	2.12	2.91	0.60	2.46	2.97	0.55

Scatter plots of the observed and predicted values of CBR, ABR, XGBR, GBRT, RF, ERT, SVR, MLPR, and EN are presented in Figure 4. The scatter plots show the relationship between the predicted value of each model and the observed value. A good prediction result will be evenly distributed on both sides of the regression line. We can observe that the tree-based models' (CBR, ABR, XGBR, GBRT, RF, ERT) predicted values and observed values were evenly distributed on both sides of the regression line, indicating that the predicted value of the model is very close to the observed value. The predicted value of the SVR model and the observed value were also evenly distributed on both sides of the regression line, and the accuracy of the SVR model was lower than that of the tree-based model. Similarly, we can observe that there was a significant difference between the predicted value and the observed value of the MLPR and EN models, indicating that the prediction errors of the MLPR and EN models are relatively large.
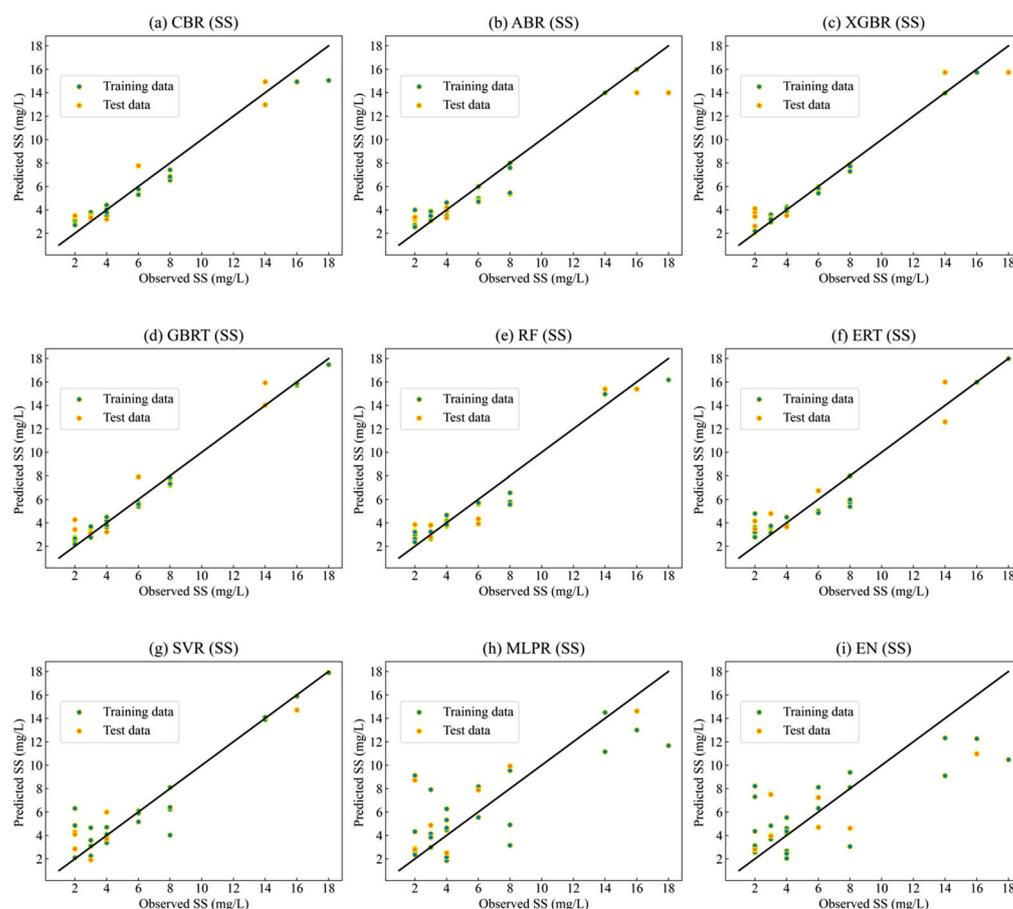
Figure 4. Scatter plot of the observed values and the predicted values of the SS (mg/L) concentration using nine machine learning models including CBR, ABR, XGBR, GBRT, RF, ERT, SVR, MLPR, and EN.

4. Discussion

For a further discussion, the distribution map of Chl-a obtained by CBR model for the Beigong Reservoir hyperspectral imagery is shown in Figure 5. According to the statistics of the inversion results, the maximum value of the inversion result was 14.17 mg/m^3 the minimum value was 3.54 mg/m^3. The observed value ranged from 2.62 mg/m^3 to 14.2 mg/m^3. The inversion map reveals the spatial distribution of Chl-a in the Beigong Reservoir. The concentration of Chl-a was relatively high in the west part of Beigong Reservoir and mainly concentrated along the shore.

The distribution map of SS obtained by CBR model for the Beigong Reservoir hyperspectral imagery is shown in Figure 6. According to the statistics of the inversion results, the minimum value was 2.71 mg/L and the maximum value of the inversion result was 15.04 mg/L. The observed value ranged from 2 mg/L to 18 mg/L. The inversion map shows that the SS concentration in the southwest part of the Beigong Reservoir was significantly high compared to the whole of the reservoir and there was fragmented erythema near the reservoir's border, which may have been produced by transitory human activity.
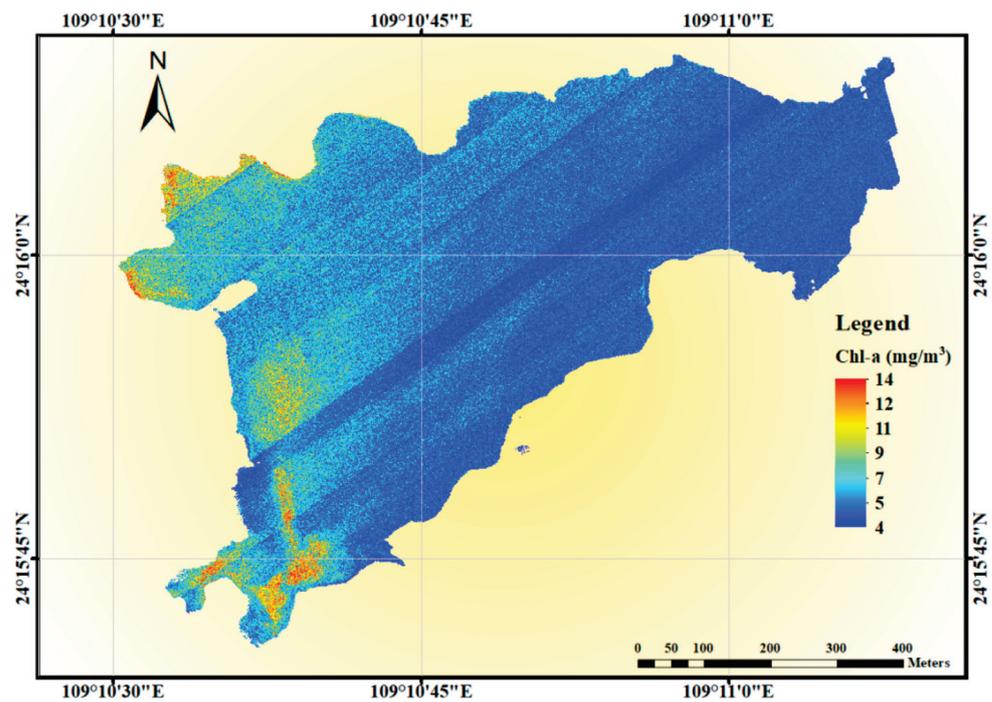
Figure 5. Distribution map of Chl-a (mg/m³) obtained using the CBR model with Beigong UAV-borne hyperspectral imagery.
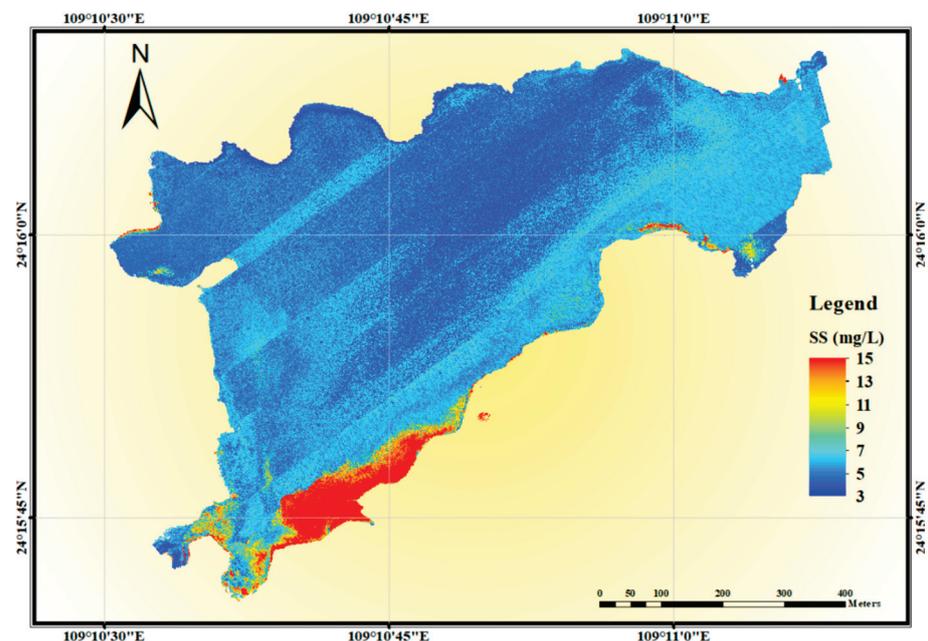
Figure 6. Distribution map of SS (mg/L) obtained using the CBR model with Beigong UAV-borne hyperspectral imagery.

5. Conclusions

The main purpose of this study is to compare the performance of various machine learning algorithms in predicting water quality parameters using UAV-borne hyperspectral data. Through this study, the main conclusions are as follows:

1. The prediction performance of different machine learning algorithms, including CBR, XGBR, GBRT, ABR, ERT, RF, SVR, MLPR, and EN, in predicting water quality were

compared. The overall prediction accuracy of the tree-based models were higher than that of the other three traditional machine learning models.

2. Two water quality parameters, including Chl-a and SS, were analyzed with different machine learning models. For the prediction of Chl-a, the R^2 values of several models ranged from 0.58 to 0.96; the RMSE ranged from 0.53 to 1.75 mg/m^3; and the MAE value ranged from 0.47 to 1.6 mg/m^3. Among them, the CBR model had the highest prediction accuracy and the XGBR model had the second-highest prediction accuracy. For the prediction of SS, the R^2 values of the nine models ranged from 0.59 to 0.94; the RMSE ranged from 1.2 to 2.97 mg/L; and the MAE value ranged from 1.11 to 2.46 mg/L. The prediction accuracy of the CBR model was the highest and the prediction accuracies of the XGBR and RF models were lower than that of the CBR. Notably, the CBR model showed stable and satisfactory performance for predicting water quality parameters, including Chl-a and SS.

3. The water quality distribution map was generated based on the UAV-borne hyperspectral data and machine learning algorithms, which can be used for large-scale and continuous inland water quality monitoring. From the water quality parameter inversion map, we observed that the pollution degree of SS in the west part of Beigong Reservoir was much higher than that in the east part. The areas with the highest Chl-a concentration mainly existed in the southern part of Beigong Reservoir and near the shore area. The management can monitor the water quality from the inversion map, improving the efficiency of water quality maintenance and saving management costs.

To conclude, this study compared and analyzed the predictive performance of nine machine learning models on different water quality parameters. In future research, we will combine multi-temporal UAV-borne hyperspectral images to analyze the dynamic change of inland water quality.

Author Contributions: Conceptualization, W.S.; methodology, W.S.; formal analysis, Y.X. and S.Y.; investigation, Q.L.; resources, Z.L.; supervision, Z.X.; data curation, L.W.; writing—original draft preparation, W.S. All authors have read and agreed to the published version of the manuscript.

References

1. Wang, Y.; Liu, D.; Xiao, W.; Zhou, P.; Tian, C.; Zhang, C.; Du, J.; Guo, H.; Wang, B. Coastal Eutrophication in China: Trend, Sources, and Ecological Effects. *Harmful Algae* **2021**, *107*, 102058. [CrossRef] [PubMed]
2. Ding, Y.; Zhao, J.; Peng, W.; Zhang, J.; Chen, Q.; Fu, Y.; Duan, M. Stochastic Trophic Level Index Model: A New Method for Evaluating Eutrophication State. *J. Environ. Manag.* **2021**, *280*, 111826. [CrossRef] [PubMed]
3. Sun, F.; Mu, Y.; Leung, K.M.Y.; Su, H.; Wu, F.; Chang, H. China Is Establishing Its Water Quality Standards for Enhancing Protection of Aquatic Life in Freshwater Ecosystems. *Environ. Sci. Policy* **2021**, *124*, 413–422. [CrossRef]

4. Moses, W.J.; Gitelson, A.A.; Perk, R.L.; Gurlin, D.; Rundquist, D.C.; Leavitt, B.C.; Barrow, T.M.; Brakhage, P. Estimation of Chlorophyll-a Concentration in Turbid Productive Waters Using Airborne Hyperspectral Data. *Water Res.* **2012**, *46*, 993–1004. [CrossRef] [PubMed]

5. Birtwell, I.K.; Farrell, M.; Jonsson, A. The Validity of Including Turbidity Criteria For Aquatic Resource Protection in Land Development Guideline (Pacific and Yukon Region). In *Canadian Manuscript Report of Fisheries and Aquatic Sciences*; Fisheries and Oceans Canada: Ottawa, ON, Canada, 2008.

6. Bierman, P.; Lewis, M.; Ostendorf, B.; Tanner, J. A Review of Methods for Analysing Spatial and Temporal Patterns in Coastal Water Quality. *Ecol. Indic.* **2011**, *11*, 103–114. [CrossRef]

7. Huang, W.; Chen, S.; Yang, X.; Johnson, E. Assessment of Chlorophyll-a Variations in High- and Low-Flow Seasons in Apalachicola Bay by MODIS 250-m Remote Sensing. *Environ. Monit. Assess.* **2014**, *186*, 8329–8342. [CrossRef]

8. Doña, C.; Chang, N.-B.; Caselles, V.; Sánchez, J.M.; Camacho, A.; Delegido, J.; Vannah, B.W. Integrated Satellite Data Fusion and Mining for Monitoring Lake Water Quality Status of the Albufera de Valencia in Spain. *J. Environ. Manag.* **2015**, *151*, 416–426. [CrossRef]

9. Du, C.; Li, Y.; Wang, Q.; Liu, G.; Zheng, Z.; Mu, M.; Li, Y. Tempo-Spatial Dynamics of Water Quality and Its Response to River Flow in Estuary of Taihu Lake Based on GOCI Imagery. *Environ. Sci. Pollut. Res.* **2017**, *24*, 28079–28101. [CrossRef]

10. Syariz, M.A.; Lin, C.-H.; Nguyen, M.V.; Jaelani, L.M.; Blanco, A.C. WaterNet: A Convolutional Neural Network for Chlorophyll-a Concentration Retrieval. *Remote Sens.* **2020**, *12*, 1966. [CrossRef]

11. Rajesh, A.; Jiji, G.W.; Raj, J.D. Estimating the Pollution Level Based on Heavy Metal Concentration in Water Bodies of Tiruppur District. *J. Indian Soc. Remote Sens.* **2020**, *48*, 47–57. [CrossRef]

12. Rostom, N.G.; Shalaby, A.A.; Issa, Y.M.; Afifi, A.A. Evaluation of Mariut Lake Water Quality Using Hyperspectral Remote Sensing and Laboratory Works. *Egypt. J. Remote. Sens. Space Sci.* **2017**, *20*, S39–S48. [CrossRef]

13. Quan, Q.; Hao, Z.; Xifeng, H.; Jingchun, L. Research on Water Temperature Prediction Based on Improved Support Vector Regression. *Neural Comput. Appl.* **2020**. [CrossRef]

14. Leong, W.C.; Bahadori, A.; Zhang, J.; Ahmad, Z. Prediction of Water Quality Index (WQI) Using Support Vector Machine (SVM) and Least Square-Support Vector Machine (LS-SVM). *Int. J. River Basin Manag.* **2021**, *19*, 149–156. [CrossRef]

15. Lu, H.; Ma, X. Hybrid Decision Tree-Based Machine Learning Models for Short-Term Water Quality Prediction. *Chemosphere* **2020**, *249*, 126169. [CrossRef]

16. Najah Ahmed, A.; Binti Othman, F.; Abdulmohsin Afan, H.; Khaleel Ibrahim, R.; Ming Fai, C.; Shabbir Hossain, M.; Ehteram, M.; Elshafie, A. Machine Learning Methods for Better Water Quality Prediction. *J. Hydrol.* **2019**, *578*, 124084. [CrossRef]

17. Sharafati, A.; Asadollah, S.B.H.S.; Hosseinzadeh, M. The Potential of New Ensemble Machine Learning Models for Effluent Quality Parameters Prediction and Related Uncertainty. *Process. Saf. Environ. Prot.* **2020**, *140*, 68–78. [CrossRef]

18. Parsimehr, M.; Shayesteh, K.; Godini, K.; Bayat Varkeshi, M. Using Multilayer Perceptron Artificial Neural Network for Predicting and Modeling the Chemical Oxygen Demand of the Gamasiab River. *Avicenna J. Environ. Health Eng.* **2018**, *5*, 15–20. [CrossRef]

19. Xiaojuan, L.; Mutao, H.; Jianbao, L. Remote Sensing Inversion of Lake Water Quality Parameters Based on Ensemble Modelling. *E3S Web Conf.* **2020**, *143*, 02007. [CrossRef]

20. Tang, J.W.; Tian, G.L.; Wang, X.Y.; Wang, X.M.; Song, Q.J. The Methods of Water Spectra Measurement and Analysis I: Above-Water Method. *J. Remote. Sens.* **2004**, *8*, 37–44.

21. Mobley, C.D. Estimation of the Remote-Sensing Reflectance from above-Surface Measurements. *Appl. Opt. AO* **1999**, *38*, 7442–7455. [CrossRef] [PubMed]

22. Kelcey, J.; Lucieer, A. Sensor Correction of a 6-Band Multispectral Imaging Sensor for UAV Remote Sensing. *Remote. Sens.* **2012**, *4*, 1462–1493. [CrossRef]

23. Qun'ou, J.; Lidan, X.; Siyang, S.; Meilin, W.; Huijie, X. Retrieval Model for Total Nitrogen Concentration Based on UAV Hyper Spectral Remote Sensing Data and Machine Learning Algorithms—A Case Study in the Miyun Reservoir, China. *Ecol. Indic.* **2021**, *124*, 107356. [CrossRef]

24. He, J.; Lin, J.; Ma, M.; Liao, X. Mapping Topo-Bathymetry of Transparent Tufa Lakes Using UAV-Based Photogrammetry and RGB Imagery. *Geomorphology* **2021**, *389*, 107832. [CrossRef]

25. Zhang, Y.; Wu, L.; Ren, H.; Liu, Y.; Zheng, Y.; Liu, Y.; Dong, J. Mapping Water Quality Parameters in Urban Rivers from Hyperspectral Images Using a New Self-Adapting Selection of Multiple Artificial Neural Networks. *Remote Sens.* **2020**, *12*, 336. [CrossRef]

26. Wang, X.; Xie, S.; Zhang, X.; Chen, C.; Guo, H.; Du, J.; Duan, Z. A Robust Multi-Band Water Index (MBWI) for Automated Extraction of Surface Water from Landsat 8 OLI Imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *68*, 73–91. [CrossRef]

27. Campos, J.C.; Sillero, N.; Brito, J.C. Normalized Difference Water Indexes Have Dissimilar Performances in Detecting Seasonal and Permanent Water in the Sahara–Sahel Transition Zone. *J. Hydrol.* **2012**, *464–465*, 438–446. [CrossRef]

28. Ying, H.; Xia, K.; Huang, X.; Feng, H.; Yang, Y.; Du, X.; Huang, L. Evaluation of Water Quality Based on UAV Images and the IMP-MPP Algorithm. *Ecol. Inform.* **2021**, *61*, 101239. [CrossRef]

29. Wei, L.; Huang, C.; Zhong, Y.; Wang, Z.; Hu, X.; Lin, L. Inland Waters Suspended Solids Concentration Retrieval Based on PSO-LSSVM for UAV-Borne Hyperspectral Remote Sensing Imagery. *Remote Sens.* **2019**, *11*, 1455. [CrossRef]

30. Freund, Y. *Boosting a Weak Learning Algorithm by Majority*; AT&T Laboratories: Murray Hill, NJ, USA, 1995.

31. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
32. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
33. 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, Association for Computing Machinery. New York, NY, USA, 13 August 2016; pp. 785–794.
34. Dong, W.; Huang, Y.; Lehane, B.; Ma, G. XGBoost Algorithm-Based Prediction of Concrete Electrical Resistivity for Structural Health Monitoring. *Autom. Constr.* **2020**, *114*, 103155. [CrossRef]
35. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient Boosting with Categorical Features Support. *arXiv* **2018**, arXiv:1810.11363.
36. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
37. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely Randomized Trees. *Mach. Learn* **2006**, *63*, 3–42. [CrossRef]
38. Cheng, H.; Shi, Y.; Wu, L.; Guo, Y.; Xiong, N. An Intelligent Scheme for Big Data Recovery in Internet of Things Based on Multi-Attribute Assistance and Extremely Randomized Trees. *Inf. Sci.* **2021**, *557*, 66–83. [CrossRef]
39. Raghavendra, N.S.; Deka, P.C. Support Vector Machine Applications in the Field of Hydrology: A Review. *Appl. Soft Comput.* **2014**, *19*, 372–386. [CrossRef]
40. Günther, F.; Fritsch, S. Neuralnet: Training of Neural Networks. *R J.* **2010**, *2*, 30. [CrossRef]
41. Zou, H.; Hastie, T. Regression Shrinkage and Selection via the Elastic Net, with Applications to Microarrays. *JR Stat. Soc. Ser. B* **2004**, *67*, 301–320. [CrossRef]
42. He, J.; Chen, Y.; Wu, J.; Stow, D.A.; Christakos, G. Space-Time Chlorophyll-a Retrieval in Optically Complex Waters That Accounts for Remote Sensing and Modeling Uncertainties and Improves Remote Estimation Accuracy. *Water Res.* **2020**, *171*, 115403. [CrossRef] [PubMed]
43. Beck, R.; Zhan, S.; Liu, H.; Tong, S.; Yang, B.; Xu, M.; Ye, Z.; Huang, Y.; Shu, S.; Wu, Q.; et al. Comparison of Satellite Reflectance Algorithms for Estimating Chlorophyll-a in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. *Remote. Sens. Environ.* **2016**, *178*, 15–30. [CrossRef]
44. Soomets, T.; Uudeberg, K.; Jakovels, D.; Brauns, A.; Zagars, M.; Kutser, T. Validation and Comparison of Water Quality Products in Baltic Lakes Using Sentinel-2 MSI and Sentinel-3 OLCI Data. *Sensors* **2020**, *20*, 742. [CrossRef] [PubMed]
45. Buma, W.G.; Lee, S.-I. Evaluation of Sentinel-2 and Landsat 8 Images for Estimating Chlorophyll-a Concentrations in Lake Chad, Africa. *Remote Sens.* **2020**, *12*, 2437. [CrossRef]

46. Gholizadeh, M.H.; Melesse, A.M.; Reddi, L. A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors* **2016**, *16*, 1298. [CrossRef] [PubMed]
47. Huang, M.; Kim, M.S.; Delwiche, S.R.; Chao, K.; Qin, J.; Mo, C.; Esquerre, C.; Zhu, Q. Quantitative Analysis of Melamine in Milk Powders Using Near-Infrared Hyperspectral Imaging and Band Ratio. *J. Food Eng.* **2016**, *181*, 10–19. [CrossRef]

Article

Potentials of Airborne Hyperspectral AVIRIS-NG Data in the Exploration of Base Metal Deposit—A Study in the Parts of Bhilwara, Rajasthan

Arindam Guha [1,*], Uday Kumar Ghosh [2], Joyasree Sinha [2], Amin Beiranvand Pour [3], Ratnakar Bhaisal [2], Snehamoy Chatterjee [4], Nikhil Kumar Baranval [1], Nisha Rani [2], K. Vinod Kumar [1] and Pamaraju V. N. Rao [5]

1   Geosciences Group, National Remote Sensing Centre, Indian Space Research Organization, Balanagar, Hyderabad 500010, India; nikhilkumar_b@nrsc.gov.in (N.K.B.); vinodkumar_k@nrsc.gov.in (K.V.K.)
2   Photogeology and Remote Sensing Division, 29, Jawaharlal Nehru Road, Geological Survey of India, Kolkata 700016, India; uday.ghosh@gsi.gov.in (U.K.G.); joyasree.sinha@gsi.gov.in (J.S.); ratnakar.bhaisal@gsi.gov.in (R.B.); nisha.rani@gsi.gov.in (N.R.)
3   Institute of Oceanography and Environment (INOS), Universiti Malaysia Terengganu (UMT), Kuala Nerus 21030, Malaysia; beiranvand.pour@umt.edu.my
4   Department of Geological and Mining Engineering and Science, Michigan Technological University, Houghton, MI 49931, USA; schatte1@mtu.edu
5   National Remote Sensing Centre, Indian Space Research Organization, Hyderabad 500010, India; rao_pvn@nrsc.gov.in
*   Correspondence: arindam_g@nrsc.gov.in

Abstract: In this study, we have processed the spectral bands of airborne hyperspectral data of Advanced Visible Infrared Imaging Spectrometer-Next Generation (AVIRIS-NG) data for delineating the surface signatures associated with the base metal mineralization in the Pur-Banera area in the Bhilwara district, Rajasthan, India.The primaryhost rocks of the Cu, Pb, Zn mineralization in the area are Banded Magnetite Quartzite (BMQ), unclassified calcareous silicates, and quartzite. We used ratio images derived from the scale and root mean squares (RMS) error imagesusing the multi-range spectral feature fitting (MRSFF) methodto delineate host rocks from the AVIRIS-NG image. The False Color Composites (FCCs) of different relative band depth images, derived from AVIRIS-NG spectral bands, were also used for delineating few minerals. These minerals areeither associated with the surface alteration resulting from the ore-bearing fluid migration orassociated with the redox-controlled supergene enrichments of the ore deposit.The results show that the AVIRIS-NG image products derived in this study can delineate surface signatures of mineralization in 1:10000 to 1:15000 scales to narrow down the targets for detailed exploration.This study alsoidentified the possible structural control over the knownsurface distribution of alteration and lithocap minerals of base metal mineralizationusing the ground-based residual magnetic anomaly map. This observationstrengthens the importance of the identified surface proxiesas an indicator of mineralization. X-ray fluorescence analysis of samples collectedfromselected locations within the study area confirms the Cu-Pb-Zn enrichment. The sulfide minerals were also identified in the microphotographs of polished sections of rock samples collected from the places where surface proxies of mineralization were observed in the field. This study justified the investigation to utilize surface signatures of mineralization identified using AVIRIS-NG dataand validated using field observations, geophysical, geochemical, and petrographical data.

Keywords: AVIRIS-NG; base metal; continuum removed spectral bands; ground magnetic data; banded magnetite quartzite; multi-rangespectral feature fitting; relative band depth

1. Introduction

Advanced Visible Infrared Imaging Spectrometer-Next Generation (AVIRIS NG) data have been successfully utilized for delineatingthe surface signatures of different mineral de-

posits [1–3]. Utilities of AVIRIS-NG data for deriving spectral index image composites to identify different rock types in the sedimentary province have also been reported [1,4–6]. Studies also indicated the advantages of hyperspectral bands of AVIRIS-NG data over the multispectral bands of sensorssuch as Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) for mineral mapping due to superior spectral, spatial, and radiometric resolution of airborne hyperspectral data like AVIRIS-NG data [7]. In this respect, comparative analysisof AVIRIS-NG and ASTER data confirmed the potential of airborne AVIRIS-NG data over ASTER data for delineating surface exposures of kimberlite and chromite, as these rocks generally have few isolated and small-sized surface exposures [1]. The past studies also reported airborne hyperspectral image (AVIRIS) over spaceborne hyperspectral data EO-1 Hyperion [8,9]. However, there are limited records on the comparative analysis of AVIRIS-NG data with advanced multispectral data such as Landsat-08 and Sentinel-02 sensors. A recent study showed comparative utilization of airborne hyperspectral data, anddifferent spaceborne multispectral data, including Landsat, ASTER, and Sentinel, for mineral mapping [10]. The integrations of multi-sensors data for mineral mapping were also reported in the literature [11,12]. Free availability of AVIRIS-NG data covering significant geological provinces of India would definitely attract researchers to utilize AVIRIS-NG data for geological application for a few important geological provinces [13].

The exploration of the base metal deposit is always challenging using remote sensing data. A few studies were carried out where surface signatures of base metal deposits were delineated using spaceborne hyperspectral datawith moderate spatial resolution [14–16]. Still, records are limited on utilizing high spatial resolution airborne hyperspectral data for detecting subtle or patchy surface signatures. The small and patchy surface signatures of base metal mineralization are difficult to capture by spaceborne multispectral and hyperspectral data due to having low spatial and spectral resolution [1]. Many of the base metal deposits are known for their poor surface expression [17]. Some of them are identified with the surface presence of sporadically developed Gossans. The formation of Gossans above sulfide deposits is governed by the pH of the reaction system and alsoby the presence of Fe minerals with the sulfides [17]. Surface minerals, in an assemblage with goethite, hematite, and malachite, are formed above the pH greater than 6.5, while native Cu are formed at a pH lower than 6 in Gossanized lithocap [17]. However, many times, Gossans are poorly developed [18]. Many of the Cu deposits hosted by Fe-oxides (magnetite and or hematite) are characterized by albitization, formation of K feldspar, and kaolinization [19]. Detection of these patchy and small-sized surface signatures of base metal mineralization is essential to understand the overall distribution of supergene enrichment, and hydrothermal alteration zones for identifying the areas with superior concentrations of metals. The conjugate utilization of surface signatures of mineralization and the subsurface information of ore genetic controls, revealed from ground geophysical data, might be beneficial to identify localized enrichment of metals.

Exploration companies extensively utilize all the modern airborne hyperspectral sensors from globally acclaimed, commercial data providers such as HyMap, Compact Airborne Spectrographic Imager (CASI), SWIR Airborne Spectrographic Imager (SASI), AISA-EAGLE (400–970 nm) and AISA-HAWK (970–2500 nm)—Airborne Imaging Spectrometer for Applicat, and HySpexto prepare large-scale detailed alteration and lithological maps. The role of these datasets in exploration highlights the importance of airborne hyperspectral data in narrowing down an exploration target. The major advantage of the utilization of AVIRIS-NG, or similar high-quality airborne hyperspectral data, is their high spectral, radiometric, and spatial resolution. It helps to map surface signatures of mineralization in its capability to detect the subtle spectral anomalies of localized or spatially restricted targets (e.g., alterations of mineral assemblages, oxide/hydroxide minerals occurring as the cap rock above the mineral enriched zones occurring down the depth). The spectral contrast of these minerals to background rocks will be minimal if the spatial resolution of hyperspectral acquisition is made from spaceborne sensors, which generally have poor spatial resolution, radiometric resolution, and signal-to-noise ratio (SNR)compared to the airborne hyperspectral data. The alteration map prepared from the airborne hyperspectral data isvery detailed. It provides valuable input

for developing a localized exploration model in a specific mineralized province. However, most airborne hyperspectral data have a limited spatial extent, which are not sufficient to identify the spatial extension of geological controlling parameters of the ore genetic process (e.g., the regional extent of the geological structure, the host rock responsible for concentrating and hosting mineral deposits, etc.). Therefore, it is essential that the localized and large-scale map, identifying altered rock or lithocap as the proxy of mineralization, should be analyzed using regional data to improve the exploration model and strategy for mineralexploration. A mineralizedprovince may have a few discrete mineral-enriched blocks and can be controlled by the ore genetic process. This can be achieved if we use airborne hyperspectral data with other geological or spectral datasets witha larger swath.

The potential use of a large-scale map of "surface mineralization signatures" for finding the pathways of localized metal enrichment necessitates the utilization of Aairborne hyperspectral data using AVIRIS-NG sensor to derive some spectrally enhanced products to identify the cap rock and associated altered rock indicative of enrichment of base metal deposit. In this study, an attempt has been made to synergize the high-resolution map of surface signatures of metal enrichment with the high-resolution geophysical anomaly map to identify the locales of metal enrichment. Some of these surface enrichments may lead to the discovery of a pocket load for mining.

The study area is exposed to the tropical dry climate in the state of Rajasthan, which issituated in thenorthwestern part of India. Most of the rocks were exposed on the surface with the limited surface covers (covered with vegetation, soil). Thus, the site is ideal for carrying out remote sensing-based study for detecting surface mineralization signatures using airborne hyperspectral data (Figure 1a). Previous research attributed to the mineralization of Pur-Banera is syngenetic with other sediment-hosted Pb-Zn mineral deposits in Rampur-Agucha, Zawar, and Rajpura-Dariba deposits [13]. It has also been identified that the base metal deposits are associated with the longitudinal structure [20]. Therefore, spectral mapping of supergene and altered minerals are conjugated with the ground magnetic data processing to relate surface proxies of mineralization with the structural control.
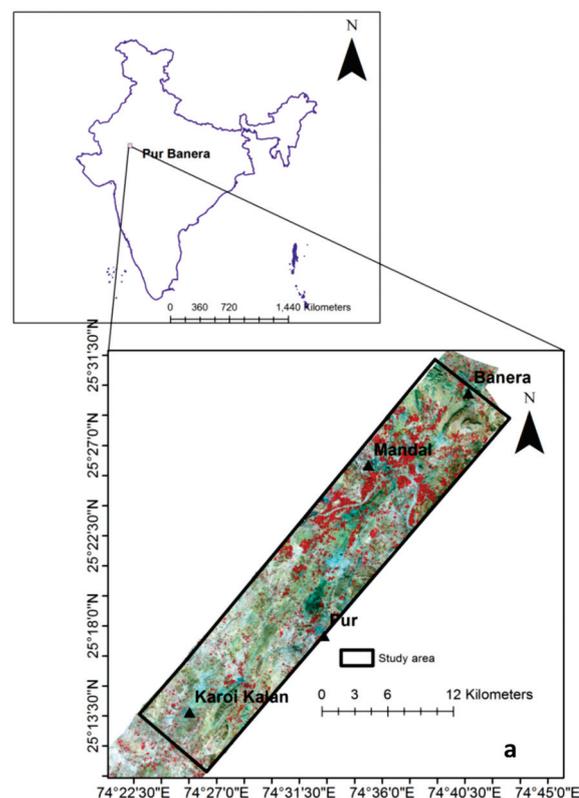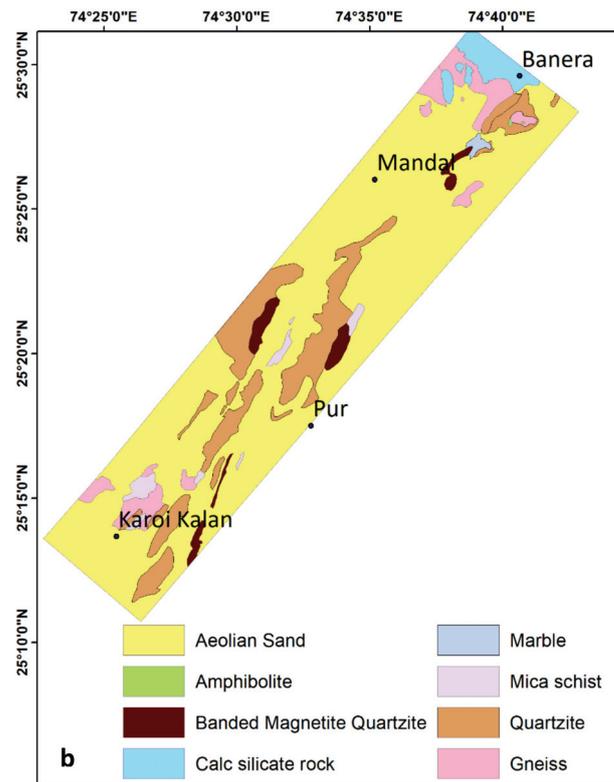

Figure 1. *Cont.*

Figure 1. (**a**) AVIRIS-NG false-color composite (FCC)map of Pur-Banera area. In the figure, AVIRIS-NG image strips from two successive airborne acquisitions are shown. A few important locations within the study area are also shown. For FCC image, Red = 850 nm, Green = 650 nm and Blue = 560 nm bands were used; (**b**) Lithological map showing the disposition of major rock types.

2. Geology

The Pur-Banera belt is characterized by the 80 km long mineralized belt from Banera to Rewara. The rocks of the study area are chemogenic in origin and, the main rock units of the area are calcareous silicate, Banded Magnetite Quartzite (BMQ), quartzite, amphibole rich marble, garnetiferous schist, ferruginous quartzite, amphibolite, garnetiferous mica-schist, etc., [21].

The Pur-Banera Group is part of the Bhilwara Supergroup of the Lower Proterozoic age, and the sediments were believed to be deposited in a pull-apart basin [22]. The lower most stratigraphic unit of the area is known as the Pur Formation. This Formation is constituted with polymictic conglomerate, micaceous quartzite [23]. It is overlain by calcareous silicate of the Rewara Formation. The Rewara Formation, in turn, is overlain by an inter-banded sequence of BMQ, calcareous silicate rock, and garnet biotite schist that constitutes the Tiranga Formation that forms the main hosts for sulfide mineralization (Pb-Zn). The BMQ is regarded as the marker lithounit for Cu, Pb, Zn deposits [23]. The uppermost lithologicalFormation of the study areais known as the Samodi Formation. Major litho-components of this formation are quartzite, mica schist, marble, and calcareous silicate [23]. In addition to the lithological control, the base metal mineralizationis also controlled by fault systems [16]. Base metal deposits of the study area and the reported occurrence of metal inthe adjoining regions are believed to be associated with the north-northeast (NNE)-south-southwest (SSW) to the northeast (NE)-southwest (SW) trending longitudinal structure [13]. Uranium mineralization was also reported in the study area, and it was associated with the quartzite of the Pur Formation [23].

Structurally, the Pur-Banera supracrustals were highly deformed with three phases of deformation (D1–D3); each of these deformation phaseshad imprinted their respective tectonic fabric [23]. The first deformation (D1) resulted in the formation of the NW-SE trending upright fold [23] from NE-SW-directed shortening. Subsequently, NNE–SSW-

directed deformation (D2) contributed to the formation of west-northwest (WNW) – east-southeast (ESE) trending vertical folds [24]. Finally, asymmetrical folds developed due to dextral shearing directed along the north (N)-south (S)/NNE SSW trend at the final stage of deformation (D3) [24]. Therefore, thesegeological structures could have played an important role inremobilizing the ore minerals from host rock to concentrate them along with specific structural fabric [25]. It has been believed that the base metal depositswere formed by convective seawater circulation in zones of crustal extension that traversed the Aravalli Fold belt [25].In the study area, Pb-Zn occurrences were reported at the eastern fringe of the study area, whereas Cu-rich zones were reported in the western fringe [26]. The lithological map of the study area was prepared following the method discussed in [1] and using the information acquired during the field survey, available geological literature, and textural, geomorphic, and tonal contrasts of different rocks observed in the AVIRIS-NG FCC image (Figure 1b). Cu deposits are reported in places such as Pur-Dariba, Gurla, Manpura, Banera forest, Devpura, Banera, Malikhera, Sanganer, Rajpura, Dhulikhera, and Gadan-ka-Khera, etc. [26,27]. On the other hand, known areas for Pb and Zn are Mahua Khurd, Devpura, and Tiranga Hill (South of Arjiya) [26,27].

3. Materials and Methods

3.1. Materials

3.1.1. AVIRIS-NG Data

The AVIRIS-NG hyperspectral sensor of the Jet Propulsion Laboratory (JPL), National Aeronautic and Space Administration (NASA), has been used to acquire hyperspectral data trending along the NE-SW direction in the Pur Banera mineralized belt. AVIRIS-NG has 425 narrow continuous spectral bands within the spectral domain of 380–2510 nm at a 5 nm interval (Table 1). The spectral datasets have high SNR (>2000 @600 nm and >1000 @ 2200 nm) [6]. The sensor has a field of view (FOV) of 34 degrees and an Instantaneous Field of view (IFOV) of 1mrad [6]. The Ground Sampling Distance (GSD) vis-à-vis pixel resolution varies from 4 m from theflightaltitude of 4km with a swath coverage of 10 km.The AVIRIS-NG data were acquired using a B-200 aircraft of the National Remote Sensing Centre, India [6].

Table 1. Specification of AVIRIS-NG data.

S No.	Sensor	Spatial Resolution (m)	Band	Band/Wavelength (μm)
1	Airborne Visible/Infrared Imaging Spectrometer- Next generation (AVIRIS-NG)	4	425	0.38−2.510 (spectral sampling: 5 nm)

Level 02 data products of AVIRIS-NG sensors were used in this study. Level 02 (surface reflectance product) data productswerederived from "at-sensor-radiance data" in real-time [6]. For deriving the Level 02 data product, the pixel-based atmospheric correction method was used with an assumption that the atmosphere is spatially invariant. The pixel-measured aerosol optical depth and water vaporconcentrationwere used for generating surface reflectance [6]. The pixel-based approach first derives atmospheric water vapor and aerosol optical depth at the pixel level from the top of the atmosphere (TOA) radiance, which serves as the input to the atmospheric correction models [6].

Atmospheric calibration was performedby implementinga modified physics-based atmospheric removal (ATREM) algorithm, which utilizes a large lookup table of pre-computed scattering and transmission coefficients, indexed by parameters specifying the aircraft operating conditions at capture time [27].

3.1.2. Spectral Datasets

We used a Field spec3 spectroradiometer, manufactured by Analytical Spectral Devices (ASD) Inc. (Boulder, CO, USA), to collect the laboratory spectra of major rock samples. All

the spectra were collected in the wavelength range of 325–2500 nm. Spectra of the rocks were collected to use them as the reference to process AVIRIS-NG data. Five to six rock samples of each rock type were collected from the prominent surface exposures in the field for spectral data collection. The size of each rock sample was fixed using the size norms of the Jet Propulsion Laboratory [28–32].

3.1.3. Ground Magnetic Data

The Geological Survey of India carried out the magnetic survey under the National Geophysical Mission (NGPM) to support stage II and stage III (i.e., semi detail staged exploration) of exploration. The magnetic survey is a low-cost, rapid geophysical survey and is known for providing an idea of the geological structures and lithological variations. It is also indicative of the presence of certain mineral deposits. In this study, a Proton Precession Magnetometer (GSM 19T) (sensitivity 0.01 nT) was used to carry out a ground-based magnetic survey [33]. At the time of magnetic data collection, special care was taken to avoid different noises induced by certain anthropogenic activities of local origin (e.g., high tension power lines, concealed iron pipes dispersed for water supply, iron-rich elements such as cars, etc.). In this study, a magnetic survey was carried out with a sampling interval of 100 m, and the distance between two subsequent profile lines (which are trending along east-west) is 125 m.

3.1.4. X-ray Fluorescence (XRF) and Petrographical Data

XRFis the fast and accurate non-destructive tool used for estimating the concentration of metals in the rock samples. In the present study, the Malvern Panalytical-Zetium Wavelength dispersive XRF(WDXRF) instrument was used [34]. Pressed powder pellets and fused beads were prepared from powdered samples, and then elements wereanalyzed using WDXRF) having a 4kw Rhodium anodetube. In the WDXRF, the X-ray tube irradiated the sample with high energy X-ray; fluorescence emanated from samples due to the interaction of the sample, and these X-rayswere measured by a wavelength dispersive detection system. The characteristic radiation coming from each element was identified using analyzed crystals, which separate the X-rays based on their wavelength [34]. The intensities of the respective X-rays were also measured sequentially to estimate the concentration of metals.

The petrographical analysis of the rocks wascarried out under refracted light and reflected light conditions using an Olympus trinocular polarizing microscope (model: Olympus BX 51).

3.2. Methods

We have followed three major segments of the methods for identifying the surface signatures of mineralization using AVIRIS-NG data andvalidating the identified surface signatures. These three segments are: 1. collection, processing of rock spectra, and mineralogical analysis of spectral features; 2. AVIRIS-NG data processing and spectral mapping; 3. validation using XRF, ore microscopy, ground magnetic, and field data. Each segment hasdifferent sub-segments. A flow chart is used to illustrate the different segments and sub-segments of the methods to executethis research (Figure 2).

3.2.1. Collection and Analysis of Rock Spectra

We have collected samples of major rock types (e.g., BMQ, calcareoussilicates, quartzite, amphibolite, and mica schist). We analyzed the spectra of these rocks using the Fieldspec 3 spectrometer under laboratory conditions. The spectrometer has a spectral resolution of 3 nm within the spectral range of 350–1000 nm and 10 nm within the spectral domain of 1000–2500 nm [28].We collected the rock samples from the surfaceexposures.We havealso collected the samples of different rock types, which indicated the signature of surface mineralization(e.g., limonite/goethite is regarded as the cap rock of amineralized zone, and sericitised/calcite rich siliceous veins are the indicatorsof hydrothermal or late phase fluid activity).
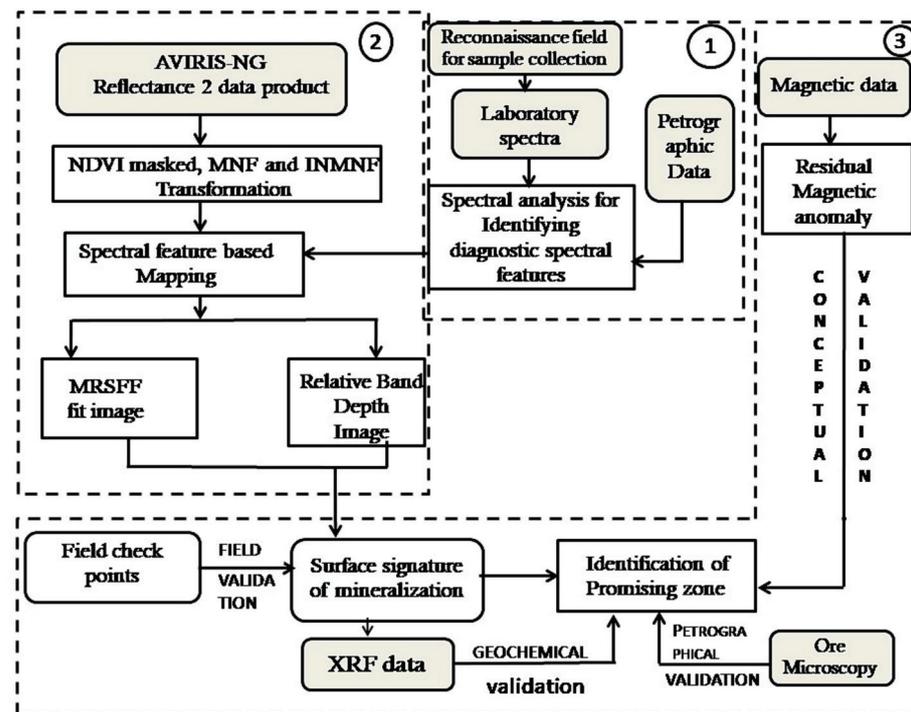
Figure 2. Flow chart showing the major steps followed in the study. In the flow chart, three major segments of the work are shown with dotted outline and numbers. The major segments in the methodology are **1**: collection and rock spectra, and mineralogical analysis of spectral feature; **2**. AVIRIS-NG data processing and spectral mapping; **3**. validation using XRF, Ore microscopy, ground magnetic, and field data.

The rock samples were cut into rectangular blocks of 4 inch × 5 inch to 5 inch × 7 inch. The radiance spectra of the rock samples were collected using a halogen lamp and the foreoptics (i.e., optical fiber) witha 250-degree field of view. Subsequently, reflectance spectra of rock samples were derived by normalizing the sample radiance with the radiance of a fully Lambertian white plate [28–32].

The methodology for collecting the spectral profiles of different types of rocks and the analysis of these spectral profiles are documented in the literature [28,33–35]. Spectral profiles of three to four spots of each sample were recorded to incorporate the variability of spectral profiles across the sample, and subsequently, these observations were averaged to derive representative spectra of each rock sample. Averaging of spectral profiles provides the advantage of reducing the effect of noise imprinted on the spectral profile of therock samples. The representative laboratory spectrum of each rock sample was subsequently resampled to a wavelength range of AVIRIS-NG bands, andthespectra of all these rocks were plotted together to understand the respective spectral features of each rock type. The spectrum of each rock type was also compared with the spectra of dominant constituent minerals to understand how spectra of minerals have contributedto shaping the spectral feature of rock. In this regard, we used the mineralogical information of each rock derived from the petrographic analysis, and analyzed under refracted light (both under plain-polarized and cross-polarized conditions).

Amongstthese rocks, BMQ, calcareoussilicates, and quartzite aremainlyassociated with the base metal deposit. We used the spectra of these rocks andother associated rocks to delineate the spatial distribution of these rocks. We also used spectra of sericitised quartzite and carbonate-rich vein and goethite/Fe-oxides as the reference to identify these minerals on the surface, as these minerals are regarded as the surface proxies of base metal enrichment. These rock samples were collected from various locations whereminerals related to surface alteration were reported.

3.2.2. AVIRIS-NG Data Processing andSpectral Mapping

AVIRIS-NG Level 02 data (reflectance data) were used in the study. Before attempting spectral enhancement of targets, prominent vegetation areas were masked using the normalized difference vegetation index (NDVI). The AVIRIS-NG-bands-NDVI image for each scene was thresholded to derive a mask band which was applied to AVIRIS NG scenes. The threshold value was selected iteratively by applying different thresholds on the scenes to ensure that the deciduous vegetation and crops present in the scenes were masked. We used AVIRIS NG bands with the wavelengths centered around 850 nm (Near infrared band) and 650 nm (for Red band) to derive NDVI. NDVI-based thresholding is used for vegetation masking because it demonstrated promising results in many of the previous studies on mineral exploration [1,36].

Spectra of host rock and altered rock were resampled to the bandwidth of AVIRIS-NG spectral bands, which were used as the reference to implement the multi-range spectral feature fitting (MRSFF) method. Spectral feature fitting is a spectral matching algorithm, which uses the inherent absorption feature of the spectrum of interest and reference endmember spectrum for least squares matching. Different types of spectral matching methods have been discussed in the literature [37]. In MRSFF, the absorption features at various wavelength ranges are considered for estimating similarities between the target spectrum and the reference spectrum [38]. Many minerals have multiple absorption features, and these diagnostic features of different minerals may overlap. Therefore, we used the multi-range spectral feature fitting method to identify the wavelength ranges of diagnosticand non-overlapping spectral features of each rock distinctly. We identified the diagnostic spectral feature of each rock with respect to the spectra of dominant constituent minerals. Mineralogy of rock was ascertained by petrographic analysis. The spectral range of the diagnostic feature of rock is identified from the continuum removed reference spectra of rocks. The spectral range of diagnostic absorption features was delimited by the wavelength of shoulders (localized reflectance maxima) on either side of the absorption minima of the spectral feature. The diagnostic spectral features of each rock were identified based on the comparison of the absorption feature of rock with the corresponding absorption feature of constituent minerals, which have prominent spectral features within the spectral domain of the AVIRIS-NG sensor. We used the width and depth of an absorption feature and the wavelength of absorption minima to estimate the relative abundance of the target [39,40]. Width is the distance measured at the inflection points occurring on either side of the absorption feature. Depth measures the distance between the deepest point of the absorption feature and the central point on a line joining the shoulders of the absorption feature. Once we estimated the least squares fit between the continuum removed AVIRIS-NG image and reference laboratory spectra of rocks using the MRSFF method, we derived the scale image by the fitted least squares model [38]. Subsequently, we derived the ratio of the scale image to root mean squares (RMS) error image of the least squares fit model, defined as "*fit*" image, for delineating the specific rock types. The *fit*image signifies the pixel with high least squares values andminimum RMS error. Furthermore, we prepared the *fit* image foreach rock type to createan FCC image fordelineating rocks in the spatial domain.

Identification of the absorption feature and its spectral range guided by mineralogical knowledge would be helpful to detect the target effectively. In addition to deriving the MRSFF-based *fit* map of different rocks, we deriveda relative band depth image delineating different minerals, which are used as the surface proxy of metal enrichment. In the relative band depth image, spectral bands having their wavelength coinciding with the wavelength of shoulders and absorption minima of the individual diagnostic spectral feature of rock were used. The relative band depth image can be used to detect minerals associated with the lithocap rock and altered rock of sulfide mineralization [41]. Conjugate detection of host rocks and the surface imprints of mineralizationhave been implemented in this study to confirm the spatial association of prominent surface alteration and cap rock minerals within the extent of the host rock. The conjugate approach can then effectively narrow down the area for detailed mineralogical and petrographical analysis for further exploration.

3.2.3. Magnetic Data Processing

The acquired magnetic data were subjected to diurnal and International Geomagnetic Reference Field (IGRF) corrections to derive the total field magnetic anomaly values. The magnetic anomaly values are then presented as a contour map using the minimum curvature gridding technique in Geosoft (Oasis Montaj version 8.3) software [42]. Unlike gravity anomalies, magnetic anomalies do not appear vertically above the anomalous body. It is asymmetric due to the dipolar behavior of the earth's magnetic field as we move away from the poles tothe equator [43,44]. This distortion in magnetic anomalies can be eliminated by a reduction-to-pole (RTP) transformation using various approaches [45–47]. The average inclination and declination of the study area are 37.67 and 0.39, respectively. The total field magnetic anomaly data were reduced to a pole using a Fourier domain filtering technique [48] in Geosoft software to bring the symmetrical anomaly above the causative source. Further, the Gaussian regional/residual filtering technique was used to separate the regional and residual magnetic anomaly [49]. The residual magnetic anomaly was separated using a high pass filter at wavenumber 3 km to delineate the anomaly caused by shallow or near surface subsurface structures.

3.2.4. XRF and Petrographical Analysis

XRF is a widely used technique to quantify the concentration of elements by generatinghigh-energy photons from an X-ray source, which pass through primary filters to normalize their energy variability. Rock samples were milled to less than 230-mesh size for XRF analysis. A 4.5-g powder sample of -200 mesh size was taken in an aluminum cup of 40 mm diameter. A few drops of polyvinyl alcohol weremixedwith the powdered sample; subsequently, the mixture was transferred to a 40 mm aluminum cup and pressed for 5 min at 40,000 psi. Samples were also ignited at above 900 ° C to determine the loss of ignition. Photons are allowed to pass into a sample, and this, in turn, allows the energy of photons to get transferred to an inner-shell electron of an atom. X-Rays displace the electron from its preferred orbit, leaving an unstable atom. An electron from an outer orbital fills the vacancy in the lower orbital by releasing its energy in a fluoresced secondary X-ray, unique to the element. Element concentrations are determined by the rate at which fluorescence of a detector in the spectrometer records secondary X-rays.XRF spectrum analysis was conducted using the methods described in the literature [50,51].

4. Results

We have processed two imaging strips of AVIRIS-NG data for the Pur-Banera area. The AVIRIS-imaging strips are aligned along the NE-SW direction, parallel to the trend of the Aravalli Group of rocks (Figure 1). The detailed steps followed to derive the study results are summarized in a flowchart (Figure 2). We have presented the results of the study in different segments to align with the proposed methods.

4.1. SpectralAnalysis to Identify the Diagnostic Spectral Features of Different Rocks and the Altered Rocks

Spectral analysis was carried out to identify the diagnostic spectral features of rocks and analyze their mineralogical significance [52]. Spectral analysis was used as the basisfor deriving the AVIRIS-NG-based image products (Figure 3). We identified the prominent spectral features of the four most dominating rock types: calcareous silicates, quartzite, amphibolites, and BMQ. These rocks are associated with the surface imprints of mineralization in the study area. We found that calcareous silicates, amphibolite, and sericitized quartzite have their diagnostic absorption features in the short-wave infrared (SWIR) domain, whereas BMQ has its diagnostic spectral feature in the near-infrared domain (Figure 3a).

We found that the spectral feature of the calcite governed the spectral feature of calcareous silicates at 2330 nm (Figure 3b). Similarly, the main diagnostic spectral feature of amphibolite was governed by the hornblende with the absorption feature having wavelength minima at 2320 and 2375 nm, respectively (Figure 3c). Sheared quartzite has sericite

as the major constituent mineral; therefore, the diagnostic spectral feature of sericite was considered as the basis for identifying the spectral feature of sericite bearing quartzite (Figure 3d). We identified that the spectral feature of BMQ was governed by the magnetite; however, the spectral depth of BMQ was smaller than the spectral depth of magnetite at the same wavelength (Figure 3e).
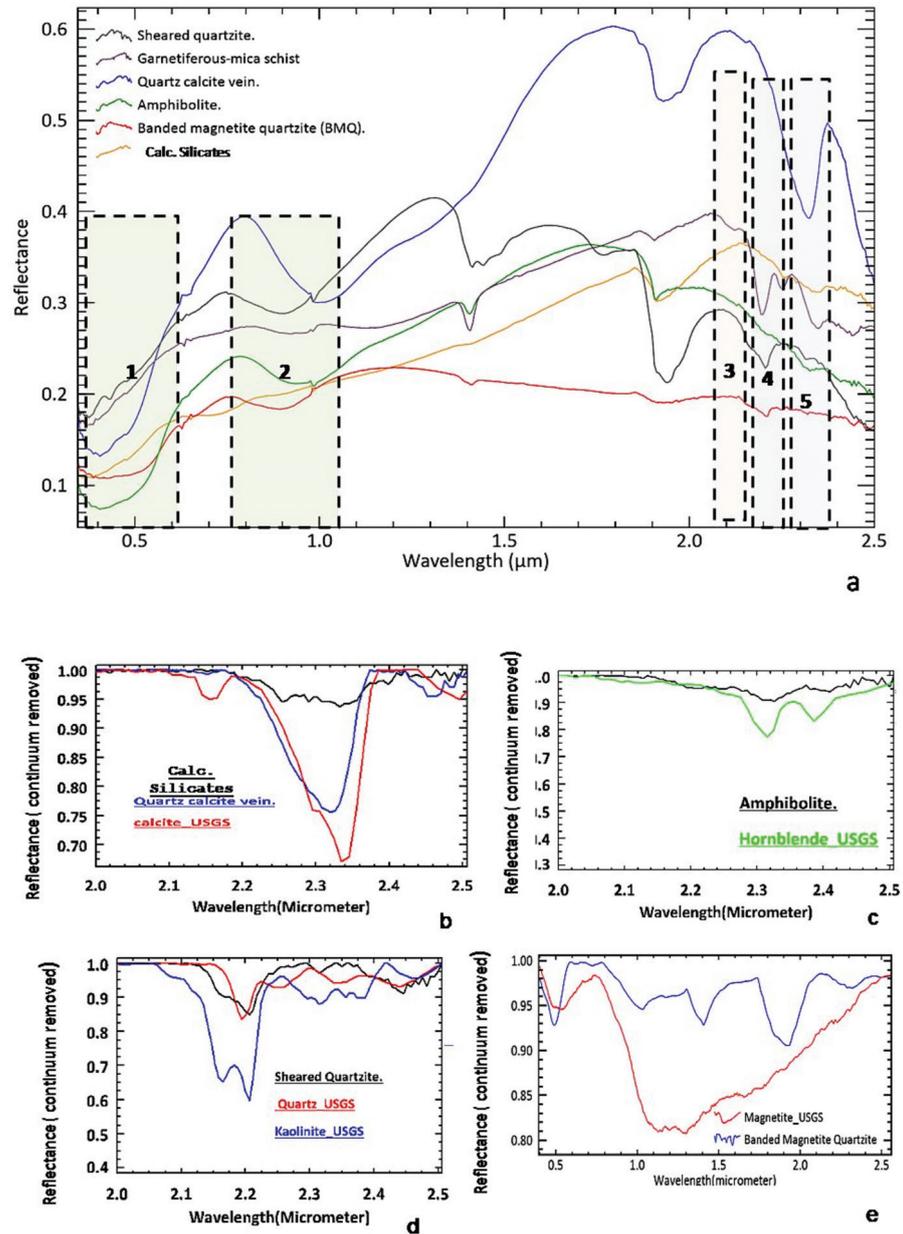

Figure 3. (**a**) Plot of representative spectral profiles of main rock types of the study area. In the plot, each box is used to illustrate schematically the spectral domain of few diagnostic absorption features of main rock type of the study area. Box 1: Wavelength domain of diagnostic spectral feature of BMQ. Box 2: Wavelength domain of diagnostic spectral feature of quartzite (which are sericite rich and sheared). Box 3: Wavelength domain of diagnostic spectral feature of carbonate rich calcareous silicates. Box 4 and 5: Wavelength domain of diagnostic spectral feature of amphibolite. Plots (**b**–**e**) of representative spectrum of each rock type with the spectra of their constituent minerals are presented to show how the spectral features of dominant and spectrally diagnostic minerals influence the spectral features for differentrock types: b. calcareous silicates; c. amphibolite. d = sheared quartzite; and e = banded magnetite quartzite.

We analyzed the optical thin section of representative rock samples of quartzite, amphibolite, BMQ, and calcareoussilicates to ascertain their respective mineral constituents andto relate spectra of constituent minerals with the rock (Figures 4 and 5). We identified the mineral constituent of all four studied rocks from the thin section analysis. The presence of sulfide minerals was confirmed based on optical thick section studies of rock samples collected around the areas with the presence of surface proxies of mineralization (Figures 4 and 5).
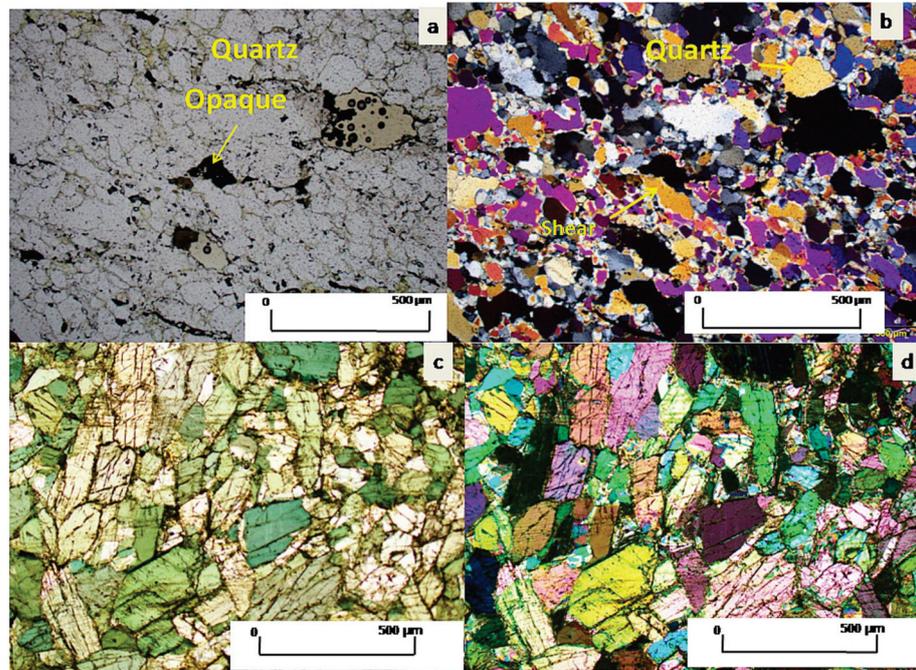
Figure 4. (**a**) Plain-polarized optical thin section of quartzite. (**b**) Cross-polarized optical thin section of quartzite in which evidence of shearing is prominent. (**c**) Plain-polarized optical thinsection of amphibolite. (**d**) Cross-polarized optical thin section of amphibolite.
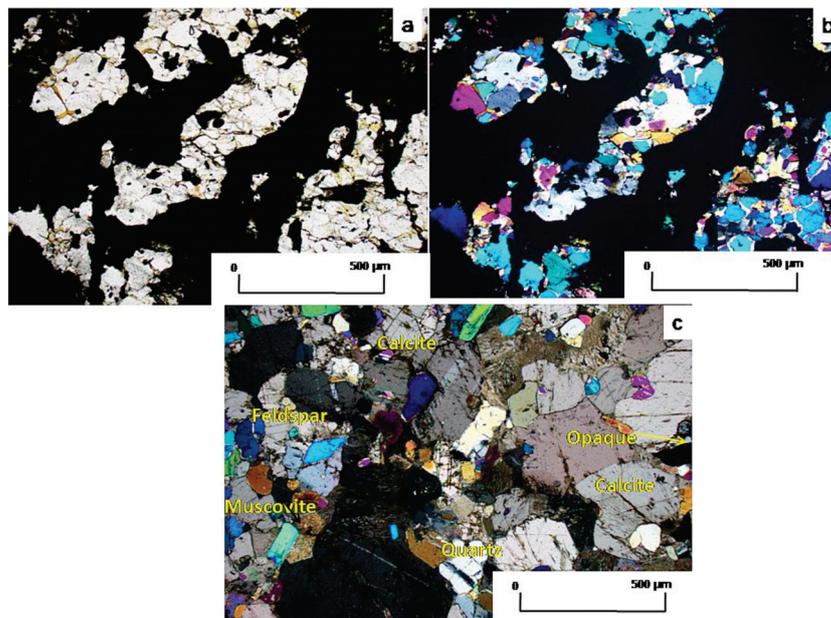
Figure 5. (**a**) Plain-polarized optical thin section of banded magnetite quartzite (BMQ) under microscope. (**b**) Cross-polarized optical thin section of BMQ. In this thin section, magnetite appears dark. (**c**) Cross-polarized optical thin section of calcareous silicate rocks.

We also analyzed the spectra of major minerals (e.g., calcite, sericite, limonite, goethite, and kaolinite) from the United States Geological Survey (USGS) spectral library after resampling to the bandwidth of AVIRIS-NG sensor bands [53]. Delineation of these minerals is important as metal mineralization in the study area has the surface imprints of Gossans (i.e., presence of goethite and limonite, seriicitized zone, calcite rich veining, etc.). We analyzed spectra of calcite, sercitie, and goethite and used the diagnostic spectral feature of these minerals (Figure 6) to propose a few relative band depth (RBD) images. The RBD images were used to delineate surface minerals of the study area, as documented in Table 2 [54].
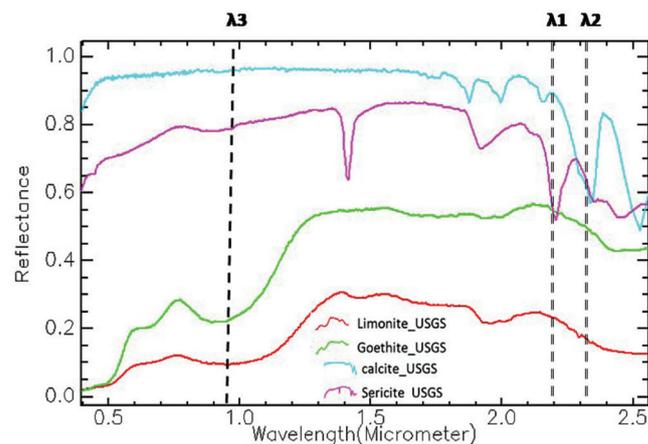
Figure 6. AVIRIS-NG spectral band resampled spectral profiles of the minerals associated with surface mineralization signatures. λ1 to λ4 refers to the absorption features of different minerals. λ1 = wavelength minima of diagnostic spectral feature of kaolinite. λ2 = wavelength minima of diagnostic spectral feature of sericite. λ3 = Wavelength minima of diagnostic spectral feature of calcite. λ4 = wavelength minima of diagnostic spectral feature of goethite and limonite.

Table 2. Spectral bands used for deriving AVIRIS-NG-based relative band depth image for different minerals associated with alteration and supergene enrichment.

S. No.	Mineral	Spectrometric Parameter		
		Wavelength of Shoulder 1 (nm)	Wavelength of Shoulder 2 (nm)	Wavelength of Abs-min (nm)
1	Calcite	2184	2389	2339
3	Illite/sericite	2144	2284	2204
4	Goethite	767	1222	937

4.2. AVIRIS-NG Data Processing and Spectral Mapping for Mapping Host Rock and Surface Signatures of Mineralization

We identified spectrometric parameters of each diagnostic feature of rock. The width and depth of the spectral features were calculated from continuum removed spectra within the spectral domain of spectral feature (Figure 7). We investigated different MRSFF image composites and RBD imagesfor detecting host rocks and proxy-minerals indicative of surface mineralization, respectively, in different parts of the study area. In the Gurla area, we found that the sericitized zone is associated with BMQ. It was ascertained based on the conjugate analysis of the RBD image compositeofcalcite, sericite, and iron oxideandthe MRSFF*fit* image of BMQ (Figure 8a,b). The sericitized zone wasidentified as a green-coloredunit within BMQ (Figure 8a,b). BMQ occurs as a structural hill in this area (Figure 8c). In this area, we identified pegmatitic intrusionwithin BMQand the highly weathered sericitized zone, which were validated in the field (Figure 9a,b).The surface imprint of the geological structure, i.e., a linear brittle structure, is identified in the FCC image ofAVIRIS-NG spectral bands. These linear brittle structures could be the conduits along which the late phase liquid had migrated, and ore had preferentially concentrated (Figure 8c).
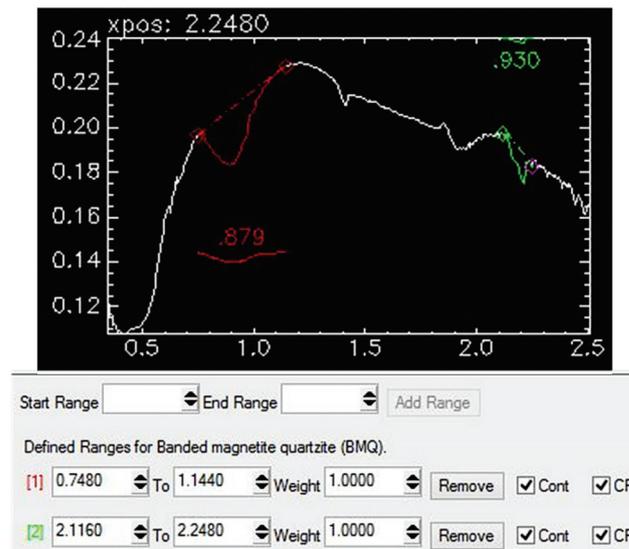
Figure 7. Graphical representation of how spectral range and other spectrometric parameters such as the width and depth of the diagnostic absorption features of each rock type are used to implement multi-range spectral feature fitting algorithm.
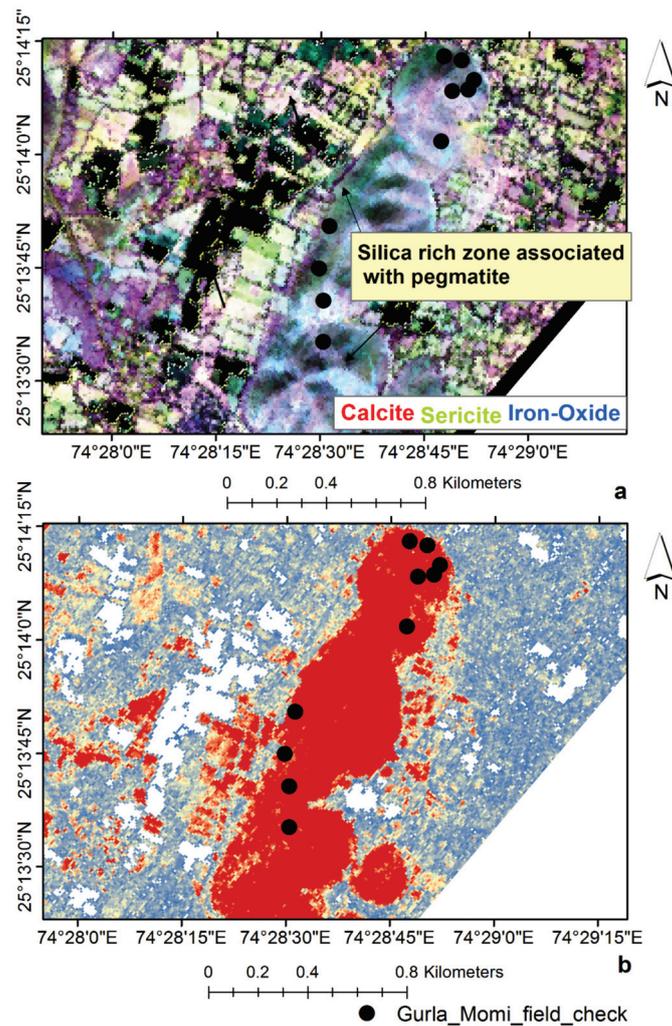

Figure 8. *Cont.*
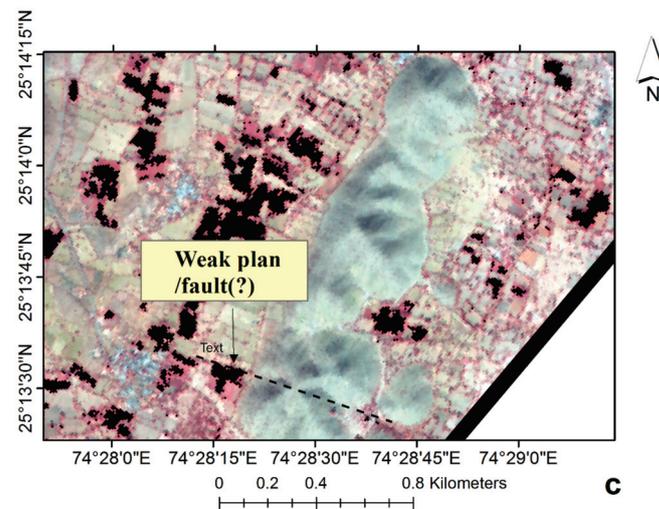
Figure 8. (**a**) FCC image of calcite, sericite, and Gossan relative band depth (RBD) to delineate the presence of pegmatite and silica rich, sericitized zones associated with BMQ in GurlaMomi area. (**b**) Surface exposures of Banded Magnetite Quartzite (BMQ)-rich zones identified in thresholded*fit* image (derived as the ratio of scale and RMS error image of MRSFF method). (**c**) FCC image composite of AVIRIS-NG bands of the same area are also shown to show. In this image, Red = 950 nm, Green = 650 nm, and Blue = 560 nm.
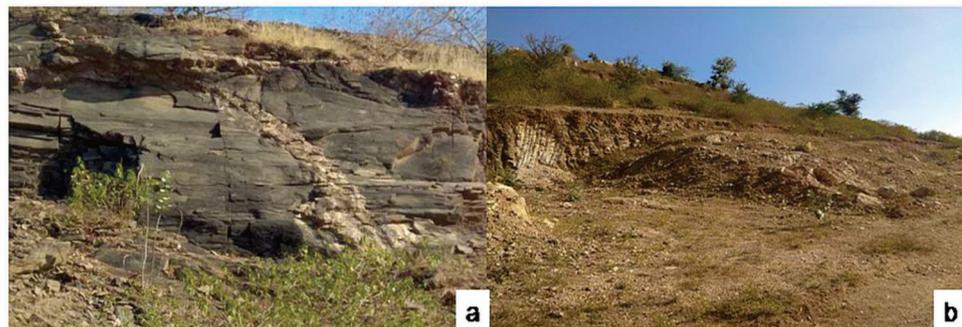

Figure 9. (**a**) Pegmatitic or silicic intrusion within the BMQ thin section near Gurla. (**b**) Surface exposures of sericitized altered rock associated with the pegmatite intrusion in the Gurla area as shown in the Figure 8a.

Another evidence of surface imprints of base metal mineralization in the study areas is thepresence of Gossan. A Gossanized zone was identified in the Baran area by the RBD image composite of calcite, sericite, and iron. The Gossanized zone appears blue in the image composite (Figure 10a). Gossanized zones were identified over the calcareous silicates (Figure 10b). We also identified the evidence of structural control around the area on which Gossans were exposed (Figure 10c). We detected the prevalence of goethite and limonite with in calcareous silicates (Figure 11a,b). Some brecciated BMQ were also present (Figure 11c).

We identified the presence of sericitised zone and calcite veining in the Devpura area within the BMQ host rock. We spotteda yellow-coloredsericitized zone within the iron-rich BMQ in the RBD image composite. The spatial extent of BMQ was delineated using the *fit* image (Figure 12a,b).We also inferredthe traces of axial planar cleavage on the area; this indicates the ductile deformation and mobilization, suitable for the concentration of metals (Figure 12c).In this area, the presence of silicified intrusion was identified with the presence of Bornite-stained BMQ during field validation (Figure 13a–c).
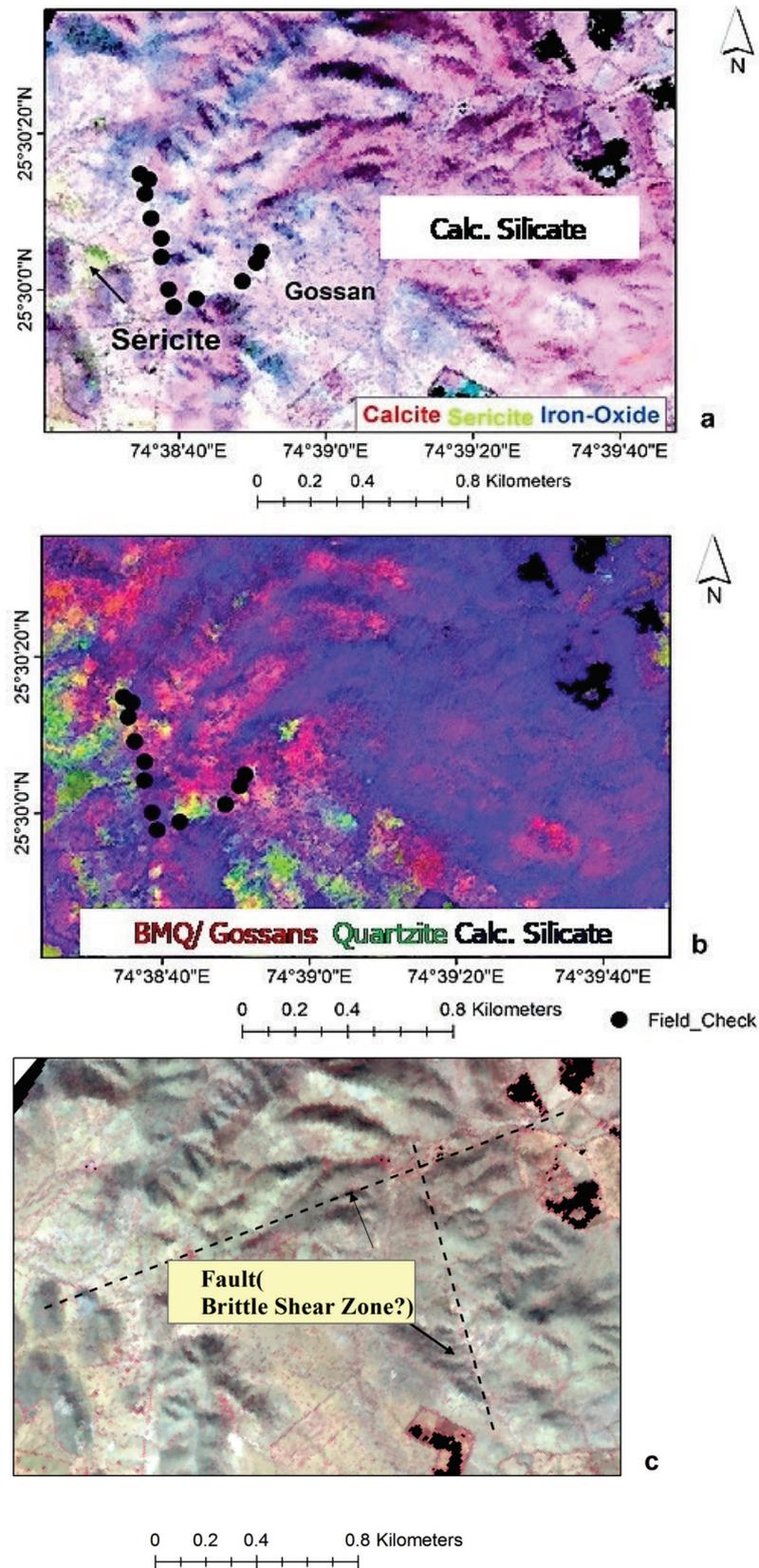
Figure 10. (**a**) FCC image of calcite, sericite, and Gossan RBDs prepared for Baran area. (**b**) MRSFF*fit* image composite developed from the *fit* images of BMQ, quartzite, and unclassified calcareoussilicates; (**c**) Some of the linear geological fabric was traced around Gossanized zone in the AVIRIS-NG FCC image. In this image, Red = 950 nm, Green = 650 nm, and Blue = 560 nm.
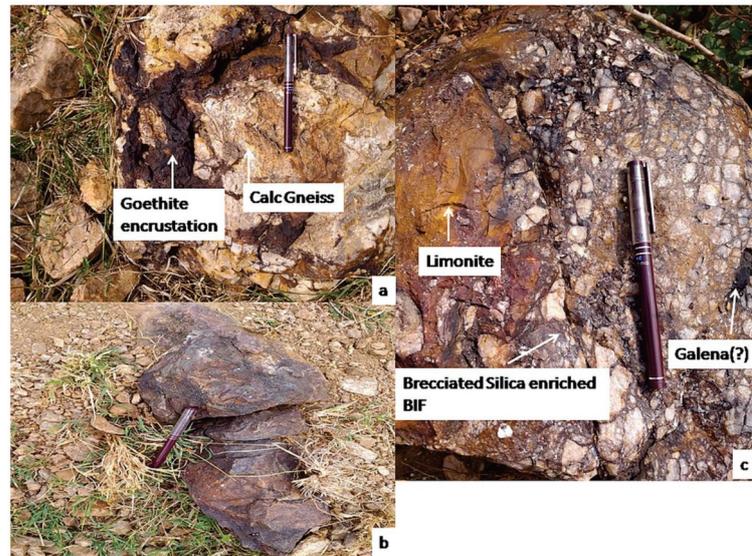
Figure 11. (**a**) Gossanized (goethite intruded) unclassified calcareous silicates near Baran area. (**b**) Surface exposures of BMQ, near Baran village. (**c**) Brecciated and altered BMQ hosting mineralization (presence of Galena, ore of lead), near Baran village.
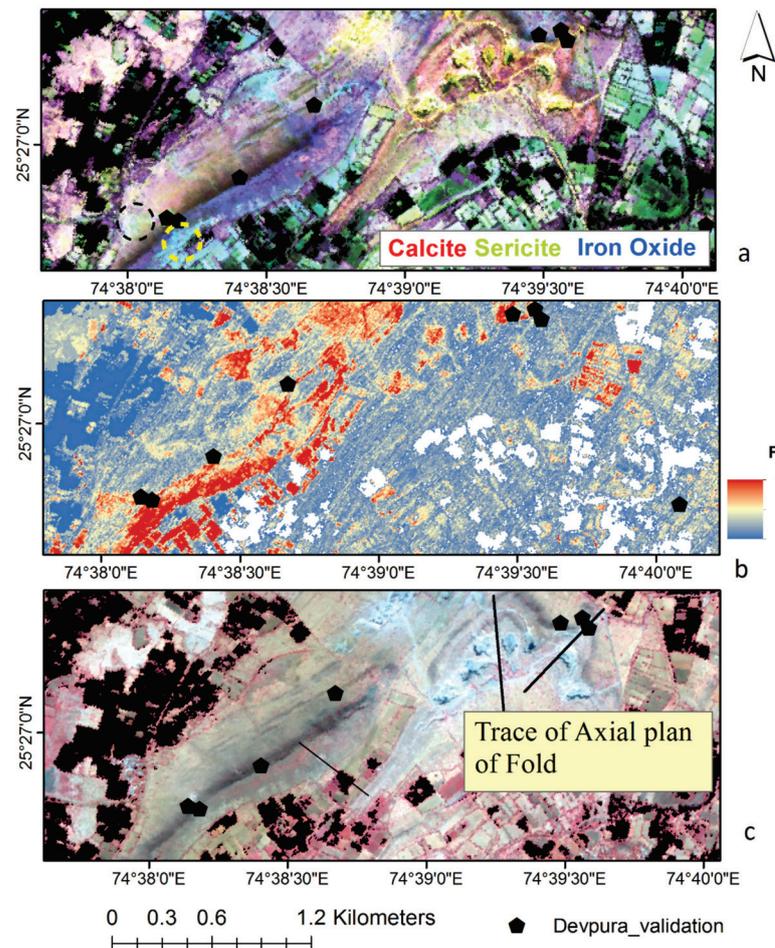

Figure 12. (**a**) FCC image of calcite, sericite, and Gossan RBDs for Devpura area. Circles are shown to highlight the pegmatite intrusion. (**b**) Surface exposures of Banded Magnetite Quartzite (BMQ)-rich zones delineated using the thresholded*fit* image. (**c**) Deformational imprints are shown in the FCC image composite of AVIRIS-NG bands. In the AVIRIS image, Red = 850 nm, Green = 650 nm, and Blue = 560 nm.

Figure 13. (**a**) The BMQ hill extending along NE-SW, near Devpura village (see Figure 12 as the reference). (**b**) The late phase silica intrusions are often associated with mineralization (Cu-sulfide) and is also identified within BMQ. See Figure 12b for further reference (**c**) The surface indications in the form of bornite stains within BMQ, at Devpura village.

In the Lanpriya area, the distribution of major rocks was well delineated using the FCC image composite of the *fit* image of quartzite, amphibolite, and silicates. The presence of sericitization and Gossanization were identified using FCC prepared using theRBD of surface proxy minerals. We identified the sericitizedzone and Gossans above the quartzite, and this was evident while comparing the *fit*image composite of three rock types and the RBD image composite of surface proxy minerals (Figure 14a,b). We also inferred imprints of some linear brittle structures in the AVIRIS-NG composite. In the field, presences of the red oxidized zone of iron oxide mineralswere identified in the sheared quartzite (Figure 15a,b).
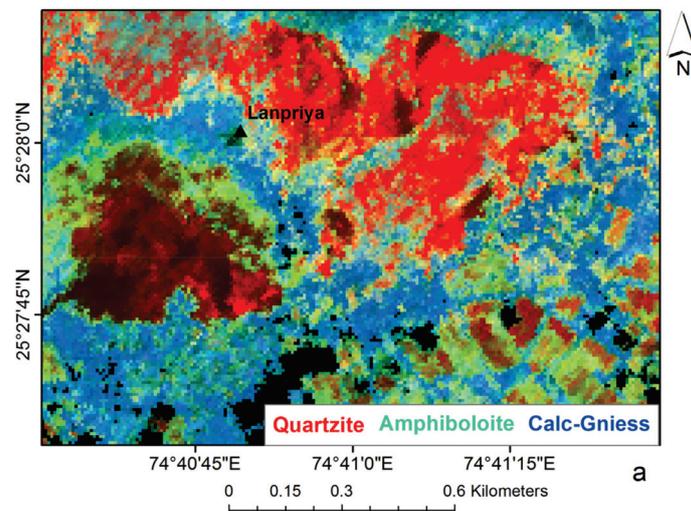

Figure 14. *Cont.*
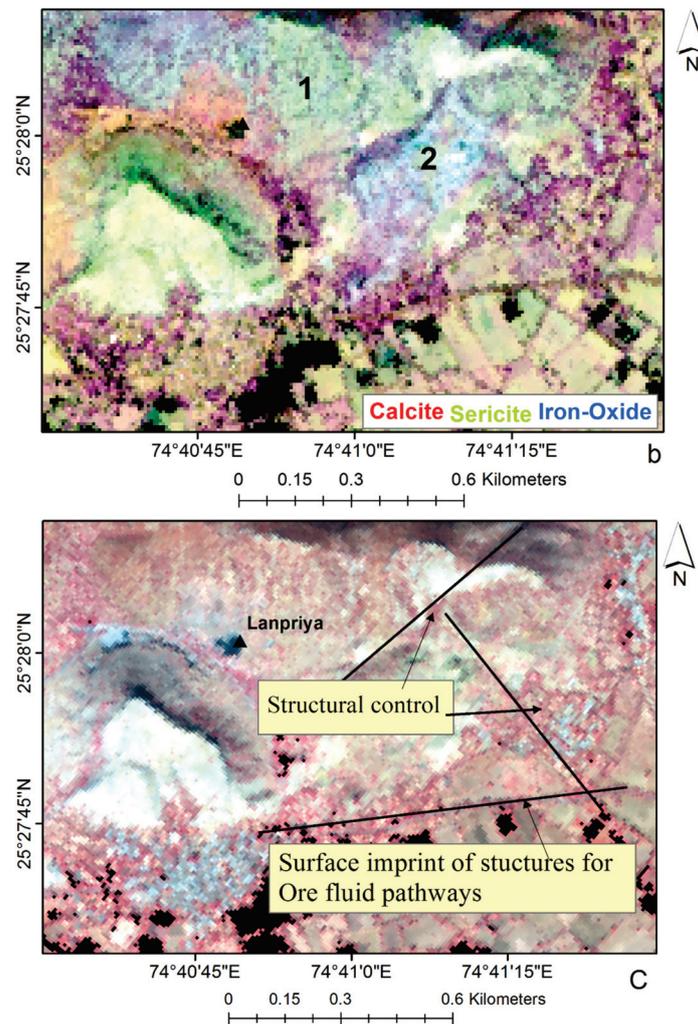
Figure 14. (**a**) FCC of *fit* images of multirange SFF algorithms detecting quartzite, amphibolite, and Calc-silicates. (**b**) Surface imprints of sericitization (1) and Gossanization (2) are also associated with the quartzite. (**c**) Structural control or the trace of the fault plan which might have facilitated fluid migration in the AVIRIS NG. In the AVIRIS image, Red = 850 nm, Green = 650 nm and Blue = 560 nm.
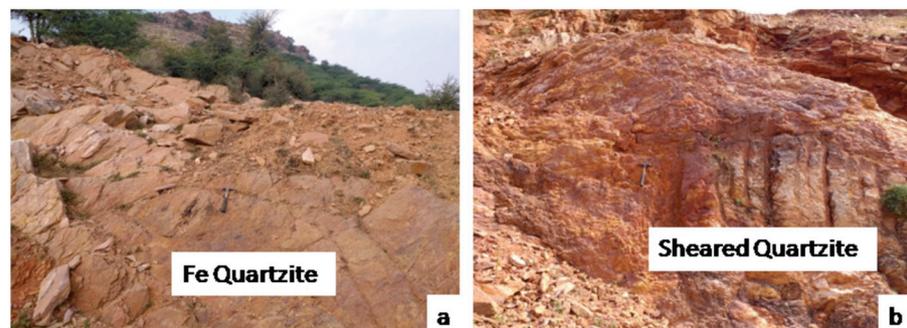
Figure 15. (**a**) The exposures of foliated and altered Fe rich quartzite near Lanpriya village. (**b**) The surface exposures of sheared and sericitized Fe-rich quartzite near Lanpriya village.

Surface exposure of calcite-rich veins and sericitized zones were also identified within BMQ at the Mangalpura area (Figure 16a). Calcite-rich zones are pinkish-red, and sericitised zones are green-colored patches within BMQ, which is in turn enhanced with blue color in the RBD image of alteration minerals (Figure 15a). Evidence of ductile deformation was also prominent in this area (Figure 16b).
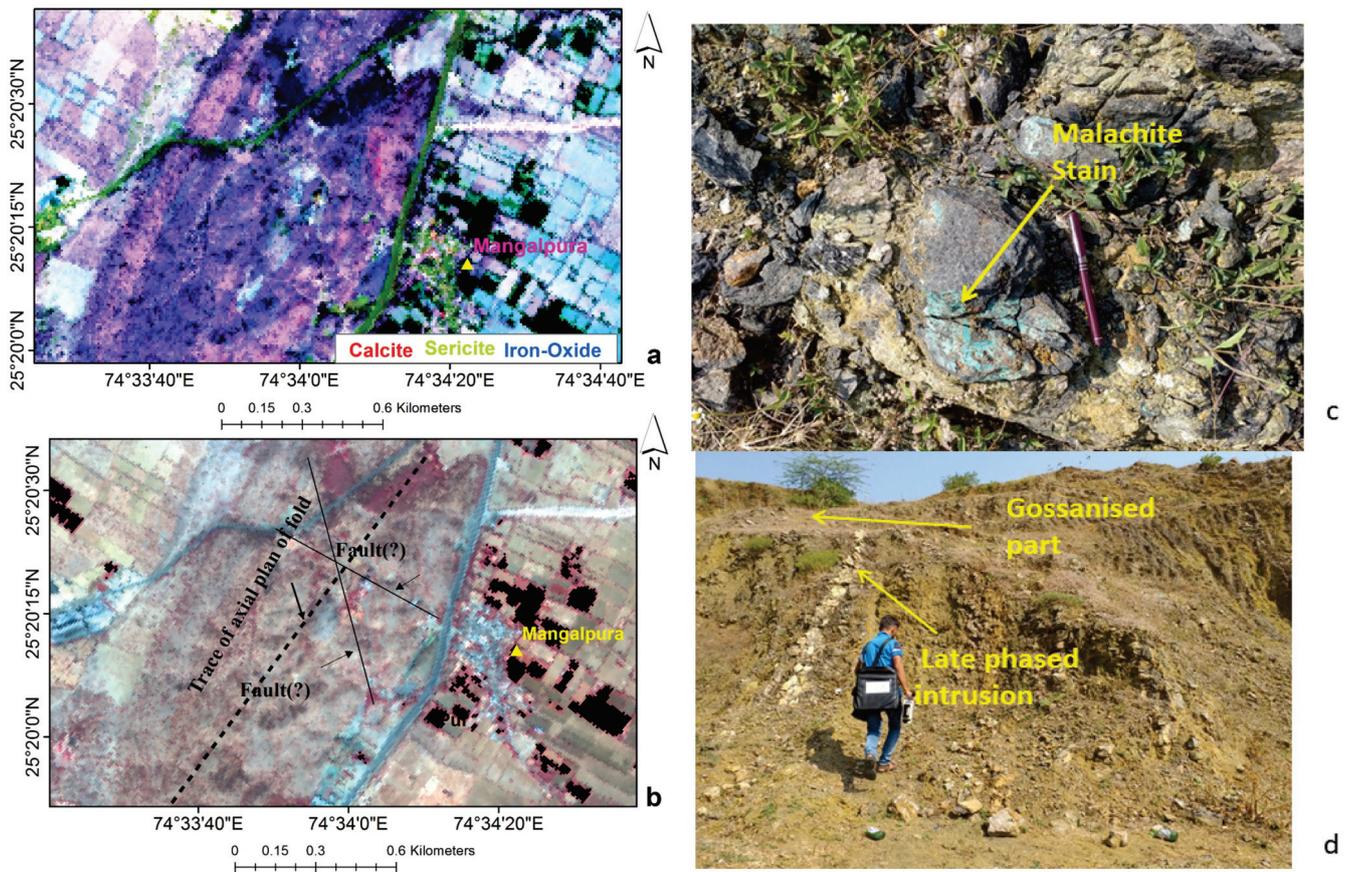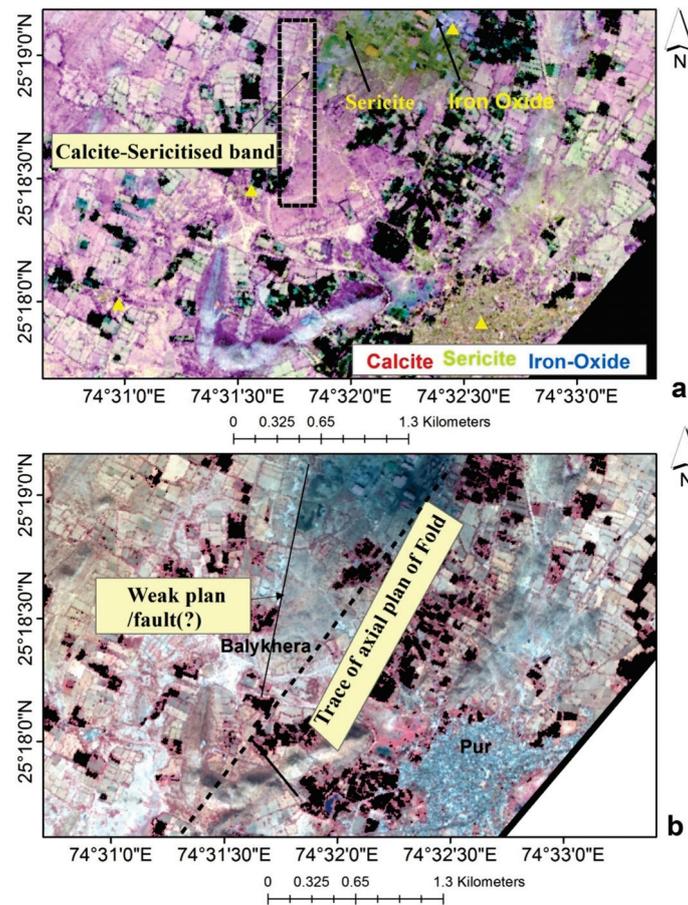
Figure 16. (**a**) False color image composite map of RBDs prepared for detecting calcite, sericite, and iron oxide are used to these minerals within BMQ around Mangalpur area. (**b**) FCC image composite of AVIRIS-NG bands are used to show localized geological structures. In the AVIRIS image, Red = 850 nm, Green = 650 nm, and Blue = 560 nm. (**c**) Malachite stains within BMQ were also identified during field validation. (**d**) Carbonate-rich sericitized veins were also identified in Mangalpura area.

Evidence of sericitization in quartzite and extensional, ductile structures is coexistent in the Balykhera area (Figures 16b and 17a). We also identified the presence of bornite on the exposed rock in the field within the Gossanized zone, associated predominantly with quartzite (Figures 16b and 18a).

4.3. Synergistic Analysis of Ground Magnetic Data and Surface Distribution of Alteration and Supergene Minerals

Most of these surface imprints (the locations of altered minerals and Gossanized zones) are aligned along the trend of the large linear structure, which was inferred from the residual magnetic anomaly map (Figure 19). This observation justifies the importance of the identified Gossan and alteration minerals as the possible indicators of localized enrichment of mineralization. This linear fabric is regarded as the suitable pathway for the movement of ore-bearing fluid, responsible for concentrating the mineral deposit. It has been believed that the base metal deposits were formed by convective seawater circulation in zones of crustal extension that traversed the Aravalli fold belt [24]. Surface proxies that occur along with the structures either indicate the imprints of the reaction processes of ore-bearing fluid or indicate the zones of mineral enrichments due to supergene processes. Both of these processes are expected along the linear fabric. Therefore, the spatial proximity of surface proxies of mineralization and linear structural controls of the ore forming process endorses the exploration utility of these surface proxies for finding new deposits.

Figure 17. (**a**) False color composite map of calcite, sericite, and iron oxide RBDs to detect calcite-sercite veins within BMQ around Balykhera area (**b**) FCC image composite of AVIRIS-NG bands showing important structural controls that might have facilitated ore-bearing fluid migration, alteration zone development, and mineral enrichment. In the AVIRIS image, Red = 850 nm, Green = 650 nm, and Blue = 560 nm.
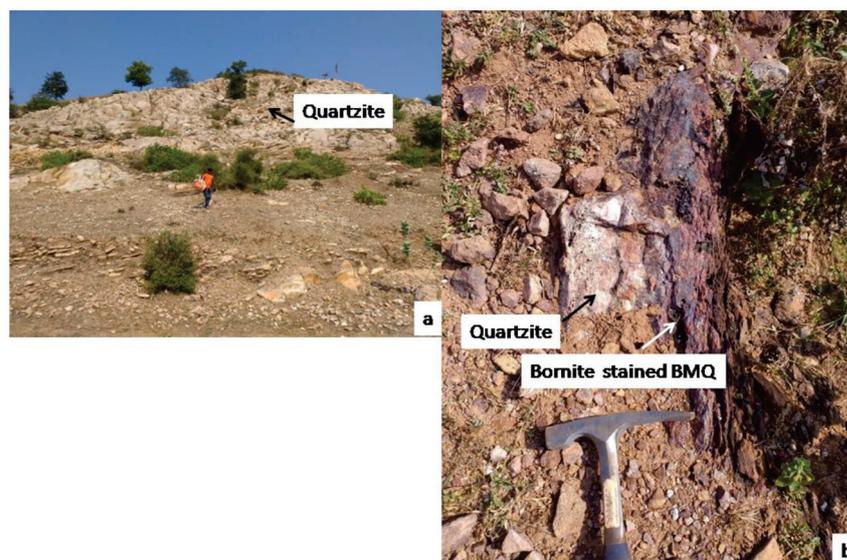
Figure 18. (**a**) The barrensurface exposures of quartzite hill trending NE-SW, near Balykhera area, Bhilwara district, Rajasthan. (**b**) The contact between sheared magnetite rich quartzite and BMQ shows surface indications of mineralization near Pur village near Balykhera, Bhilwara district, Rajasthan.
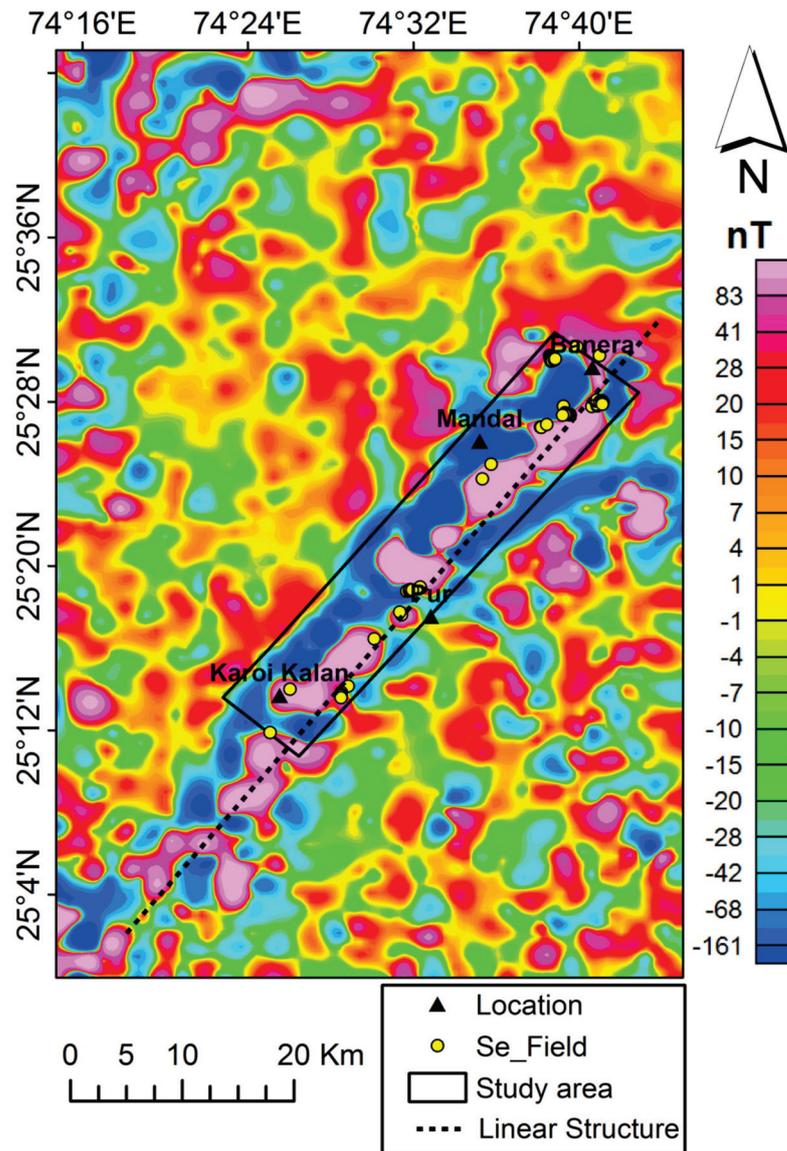
Figure 19. Residual magnetic anomaly map inferring linear structures along the field validated areas with the presence of surface proxies of mineralization (either the presence of Gossan or alteration minerals).

4.4. Results of XRF and Petrographic Analysis

It is important to identify the Cu-Pb-Zn enrichment in the rock samples associated with the surface proxies of mineralization. We analyzed the reflected property of rock samples collected from the areas with the evidence of Gossans and alteration minerals using ore microscopic to identify the concentration of sulfides.

We identifiedsulfides in the reflected polished section of the sample in Barran, Balykhera, Mangalpur, and Gurla areas (Figure 20). The XRF analysis of the samples collected from Gurla, Mangalpura, Suras, and Lanpriya showed the enriched Pb, Zn, Cu metal concentrations (Table 3). Mineralogical validation using reflected microscopy and XRF values of Pb-Zn-Cu concentrations confirm the importance of detected surface proxies of mineralization as indicators. Therefore, utilization of high spatial resolution spectral bands of AVIRIS-NG data was essential to delineate the scanty but important mineralization imprints. Some of the surface imprints identified in the areas such as Gurla, Mangalpura, and Lanpriya are associated with localized enrichment of metal, and these areas could be important for identifying pocket loads of base metal.
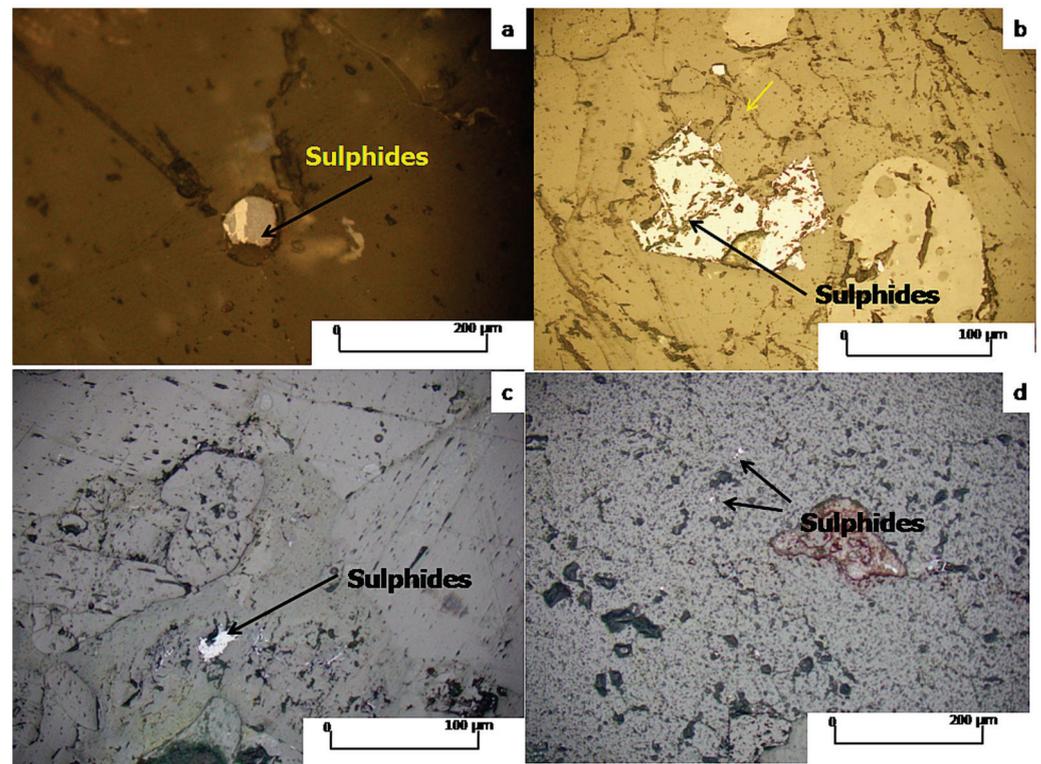
Figure 20. (**a**) Photomicrograph of sulfide (collected under reflected light) occurring within Quartzite in Mangalpura area. (**b**) Photomicrograph of sulfide (collected under reflected light) occurring within Balykhera area. (**c**) Photomicrograph of sulfide (collected under reflected light) occurring within Calcareous Silicates in Baran area. (**d**) Photomicrograph of Gossanized magnetite rich quartzite (under reflected light) near Gurla village.

Table 3. XRF values of representative samples collected from few important areas around surface.

Study Area	Mean Values of XRF Anomalies for Base Metals (in ppm)		
	Cu	**Pb**	**Zn**
Lanpriya	6286	5623	2915
Gurla-Momi	2623	12	11
Mangalpura	8438	7	22
AV 234	346	685	3026

5. Discussion

The identified structurally controlled Pb-Zn-Cu deposits from this study are the result of fluid migration associated with the cogenetic fault system [55]. Sometimes, the metal-bearing fluid source was hydrothermal, but the nature and geochemistry of fluids often get modified with the influx of meteoric water {54]. In the literature, convective seawater circulation in the zones of the crustal extension was regarded as responsible for the concentration of metal. The metal content of exhalative brines was precipitated in troughs, where biologic activity was prolific along the zone of crustal extension [56]. The base metal content of exhalative brines was precipitated from convective seawater, which evolved due to marine volcanic exhalation [56].

Most of Proterozoic crustal-scale faults extensional zones arebelieved to bethe important controls of ore mobilization and precipitation [56]. It is understood that the Pb, Zn, and Cu are preferentially partitioned into the late phase fluid. The degree of partitioning depends on the chlorine, the water content in the melt, the relationship between the vapor phase separation, and the crystallization of magma [57]. The acidic solutions often have pH within the range of 1.4 to 2.2; such solutions react with a variety of wall rock

and move towards super-saturation in terms of base metal concentration [57]. Therefore, the surface indication alterations caused by late phase fluid are also expected with base metal mineralization.

In the Pur-Banerastudy area, surface signatures related to the late phase fluid migration are common, and these signatures are identified in places such as the Gurla and Devpura areas. In general, the deposit grade is low and few isolated (Cu-Pb-Zn) deposits are reported in the basal sequence of the Aravalli Group [58].

However, there are possibilities to identify the additional pockets of deposits based on the results of this study. AVIRIS-NG data used here provide an excellent high-resolution map of surface proxies of mineralization by delineating the surface alteration zone and cap rock of mineralization. A few localized anomalies of alteration and supergene enrichment were identified from this study. Some of these localized and patchy surface proxies of mineralization are difficult to be detected using coarse spatial and spectral resolution spaceborne hyperspectral data.

This study identified the surface signatures of Cu-Pb-Zn mineralization by identifying the Gossans above calcareous silicate rocks and quartzite. These Gossans are indicative of base metal deposits, as calcareoussilicates and quartzites are both iron-poor rock. These signatures are prominent in the Baran, Lanpriya, and Mangalpura areas. On the other hand, the presence of malachite, bornite, and sericitized pegmatitic veins identified in Gurla, Devpura, indicate base metal enrichment in the BMQ. Most of the identified areas of surface proxies (Fe-oxidation, Gossans, and sericitization) are along the regional extensional zone, which is an important locale for late-phase fluid activity and ore deposition [59]. The ground magnetic study confirmed the presence of the regional extensional zone as linear discontinuity.

These surface anomalies are important to identify the pocket deposits in this mineralized province, provided these surface alterations are verified in light ofthe ore genetic model of base metal formation in the present geological province. In this study, the evidence of mineral enrichment was confirmed using several supporting data, including ore microscopyand XRF. A detailed account of the magnetic anomalies associated with the different base metal deposit and theirutilization are discussed in the literature [60]. In general, massive sulfide bodies show an increasingmagnetic response with the presence of an increasing amount of magnetic minerals [61,62]. Recent research discussed the importance of ground magnetic data to identify the structural control of base metal deposits [54].

In the present study, we also found that all the surface imprints of mineralization and their associated high Pb-Zn-Cu anomaly are associated with a NE-SW trending structure. We identified the imprints of shearing along this structure at the places such as Lanpriya and Baran, as host rocks had imprinted shearing and brecciation (Figures 11c and 15c). We also identified ductile structures in the Devpura area (Figures 12b and 15b). Most of these localized structures are the result of shearing operated along the linear structural fabric. The shearing might have resulted in localized brittle and ductile structures based on the depth of influence of shearing and the related changes in the rheological properties of the rock.

The identification of surface imprints of mineralization along the structural control confirmsthat structural controls can be potential indicators of base metal enrichment. In most of these sites, surface imprints of altered rocks are associated with the localized fault. The potential of the identified surface signatures as the mineralization indicator is further confirmed by the high concentration of Cu, Pb Zn in the samples collected from the study area (Table 3).

6. Conclusions

The present study has brought out the following aspects of airbornehyperspectral imaging utilization for base metal exploration:

1.   The identified spectral anomalies in a 1:10000 scale would provide a valuable exploration guide to explore discrete mineralized areas, which are extended along a structure trending NW-SE direction. These isolated, patchy surface proxies are importantfor detecting the localized enrichment of metals.

2. Conjugate use of MRSFF image productsfor delineating rock types and RBD image products for identifying surface mineralizationproved suitable to establish the relation between rock types and associated surface proxies of mineralization. For example, Gossans are formed above the calcareous silicates and quartzite, whereas serictization and carbonation are prominent over BMQ and quartzite.

3. Some ofthe surface mineral proxies had high XRF values indicative ofpromising high metal concentration. The spatial alignment of thesealteration zones or surface proxies along the structural fabricmakesthe occurrences of these surface proxies suitable for detailed exploration, as previous studies attributed the role of longitudinal structure in metal concentration in this area.

Author Contributions: Conceptualization, A.G.; methodology, A.G.; software, A.G., N.K.B.; validation, U.K.G., J.S., R.B. and A.G.; investigation, A.G., U.K.G., J.S. and R.B.; resources, A.G., R.B., K.V.K., P.V.N.R. and A.B.P.; formal analysis, A.G., U.K.G., J.S. and N.R.; data curation, A.G.; writing, A.G.; writing—review and editing, S.C., R.B., U.K.G. and A.B.P.; visualization, A.G., A.B.P. and S.C.; supervision, P.V.N.R. and N.R.; project administration, K.V.K. and P.V.N.R. All authors have read and agreed to the published version of the manuscript.

References

1. Guha, A.; Chatterjee, S.; Ooman, T.; Kumar, K.V. Synergistic use of ASTER, L-band ALOS PALSAR and hyperspectral AVIRIS-NG data for geoexploration of lode type gold deposit—A study in HuttiMaski Schist Belt, India. *Ore Geol. Rev.* **2020**, in press. [CrossRef]
2. Guha, A.; Mondal, S.; Chatterjee, S.; Kumar, K.V. Airborne imaging spectroscopy of igneous layered complex and their mapping using different spectral enhancement conjugated support vector machine models. *Geocarto Int.* **2020**, 1–17. [CrossRef]
3. Bhattacharya, S.; Kumar, H.; Guha, A.; Dagar, A.K.; Pathak, S.; Pasricha, K.R.; Mondal, S.; Kumar, K.V.; Farrand, W.; Chatterjee, S.; et al. Potential of airborne hyperspectral data for geo-exploration over parts of different geological/metallogenic provinces in India based on AVIRIS-NG observations. *Curr. Sci.* **2019**, *116*, 1143–1156. [CrossRef]
4. Kumar, C.; Chatterjee, S.; Oommen, T. Mapping hydrothermal alteration minerals using high-resolution AVIRIS-NG hyperspectral data in the Hutti-Maski gold deposit area, India. *Int. J. Remote Sens.* **2020**, *41*, 794–812. [CrossRef]
5. Rani, K.; Guha, A.; Kumar, K.V.; Bhattacharya, B.K.; Pradeep, B. Potential use of airborne hyperspectral AVIRIS-NG data for mapping proterozoicmetasediments in Banswara, India. *J. Geol. Soc. India* **2020**, *95*, 152–158. [CrossRef]
6. Bhattacharya, B.K.; Green, R.O.; Rao, S.; Saxena, M.; Sharma, S.; Kumar, K.A.; Srinivasulu, P.; Sharma, S.; Dhar, D.; Bandyopadhyay, S.; et al. An overview of AVIRIS-NG airborne hyperspectral science campaign over India. *Curr. Sci.* **2019**, *116*, 1082–1088. [CrossRef]
7. Rockwell, B.W. *Comparison of ASTER- and AVIRIS-Derived Mineral and Vegetation Maps of the White Horse Replacement Alunite Deposit and Surrounding Area, Marysvale Volcanic Field, Utah*; Scientific Investigations Report 2009-5117; U.S. Geological Survey: Reston, VA, USA, 2009; 31p.
8. Kruse, F.A.; Boardman, J.W.; Huntington, J.F. Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1388–1400. [CrossRef]
9. Kruse, F.A. Mineral mapping with AVIRIS and EO-1 Hyperion. In Proceedings of the 12th JPL Airborne Geoscience Workshop, Pasadena, CA, USA, 12–16 January 1998; pp. 24–28.
10. Sara, S.; Christian, M.; Christian, R. Mapping ultramafic complexes using airborne imaging spectroscopy and space-borne data in Arctic regions with abundant lichen cover, a case study from the Niaqornarssuit complex in South West Greenland. *Eur. J. Remote Sens.* **2020**, *53*, 156–175.
11. Kumar, C.; Chatterjee, S.; Oommen, T.; Guha, A. Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using AVIRIS-NG hyperspectral data in Gold-bearing granite-greenstone rocks in Hutti, India. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *86*, 102006. [CrossRef]
12. Sekandari, M.; Masoumi, I.; Pour, A.B.; Muslim, A.M.; Hossain, M.S.; Misra, A. ASTER and WorldView-3 satellite data for mapping lithology and alteration minerals associated with Pb-Zn mineralization. *Geocarto Int.* **2020**, 1–31. [CrossRef]
13. Vedas. 2021. Available online: https://vedas.sac.gov.in/aviris_web/ (accessed on 24 February 2020).

14. Govil, H.; Gill, N.; Rajendran, S.; Santosh, M.; Kumar, S. Identification of new base metal mineralization in Kumaon Himalaya, India, using hyperspectral remote sensing and hydrothermal alteration. *Ore Geol. Rev.* **2018**, *92*, 271–283. [CrossRef]

15. Mielke, C.; Muedi, T.; Papenfuß, A.; Boesche, N.K.; Rogass, C.; Gauert, C.D.K.; Altenberger, U.; de Wit, M.J. Multi-and hyperspectral spaceborne remote sensing of the Aggeneys base metal sulphide mineral deposit sites in the Lower Orange River region, South Africa. *S. Afr. J. Geol.* **2016**, *119*, 63–76. [CrossRef]

16. Zabcic, N.; Rivard, B.; Ong, C.; Mueller, A. Using airborne hyperspectral data to characterize the surface pH and mineralogy of pyrite mine tailings. Int. *J. Appl. Earth Obs. Geoinf.* **2014**, *32*, 152–162. [CrossRef]

17. Thornber, M.; Wildman, J. Supergene alteration of sulphides, VI. The binding of Cu, Ni, Zn, Co and Pb with gossan (iron-bearing) minerals. *Chem. Geol.* **1984**, *44*, 399–434. [CrossRef]

18. Suárez, S.; Prichard, H.M.; Velasco, F.; Fisher, P.C.; McDonald, I. Alteration of platinum-group minerals and dispersion of platinum-group elements during progressive weathering of the Aguablanca Ni-Cu deposit, SW Spain. *Miner. Depos.* **2010**, *45*, 331–350. [CrossRef]

19. Zhao, X.F.; Zhou, M.F. Fe-Cu deposits in the Kangdian region, SW China: A Proterozoic IOCG (iron-oxide-copper-gold) metallogenic province. *Miner. Depos.* **2011**, *46*, 731–747. [CrossRef]

20. Bharktya, D.; Gupta, R. Regional tectonics and sulphide ore localisation in Delhi-Aravalli belt, Rajasthan, India—Use of LANDSAT imagery. *Adv. Space Res.* **1981**, *1*, 299–302. [CrossRef]

21. Basu, K.K. Base metal mineralisation along the Pur-Banera belt, Bhilwara district, Rajasthan. *Misc. Publ. Geol. Surv. India* **1971**, *16*, 153–159.

22. D'Souza, J.; Prabhakar, N.; Xu, Y.; Sharma, K.K.; Sheth, H. Mesoarchaean to Neoproterozoic (3.2–0.8 Ga) crustal growth and reworking in the Aravalli Craton, northwestern India: Insights from the Pur-Banera supracrustal belt. *Precambrian Res.* **2019**, *332*, 105383. [CrossRef]

23. Shaji, T.S.; Desapati, T.; Fahmi, S.; Yadav, G.S.; Pande, A.K. Occurrence of uraninite and brannerite in the Samarkiya area, Bhilwara district, Rajasthan. *Curr. Sci.* **2007**, *92*, 592–594.

24. Deb, M.; Sarkar, S. Proterozoic tectonic evolution and metallogenesis in the Aravalli-Delhi orogenic complex, northwestern India. *Precambrian Res.* **1990**, *46*, 115–137. [CrossRef]

25. Banerjee, P.K.; Ghosh, S. Copper, lead and zinc. *Rec. Geol. Surv. India* **1972**, *102*, 41–50.

26. GSI. *Geology and Mineral Resources of Rajasthan*, 3rd ed.; No. 30, Part 12; Miscellaneous Publication: Albuquerque, NM, USA, 2011.

27. Bue, B.D.; Thompson, D.R.; Eastwood, M.; Green, R.O.; Gao, B.-C.; Keymeulen, D.; Sarture, C.M.; Mazer, A.S.; Luong, H.H. Real-time atmospheric correction of AVIRIS-NG imagery. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6419–6428. [CrossRef]

28. Guha, A.; Chakraborty, D.; Ekka, A.B.; Pramanik, K.; Kumar, K.V.; Chatterjee, S.; Subramanium, S.; Rao, D.A. Spectroscopic study of rocks of Hutti-Maski schist belt, Karnataka. *J. Geol. Soc. India* **2012**, *79*, 335–344. [CrossRef]

29. Milton, E.J.; Schaepman, M.E.; Anderson, K.; Kneubahler, M.; Fox, N. Progress in field spectroscopy. *Remote Sens. Environ.* **2009**, *113*, 92–109. [CrossRef]

30. Baldridge, A.M.; Hook, S.J.; Grove, C.I.; Rivera, G. The ASTER spectral library version 2.0. *Remote Sens. Environ.* **2009**, *113*, 711–715. [CrossRef]

31. Biggar, S.F.; Labed, J.; Santer, R.P.; Slater, P.N.; Jackson, R.D.; Moran, M.S. Laboratory Calibration of Field Reflectance Panels. In *Proceedings of the Recent Advances in Sensors, Radiometry, and Data Processing for Remote Sensing, Orlando, FL, USA, 6–8 April 1988*; SPIE: Bellingham, WA, USA, 1988.

32. Nicodemus, F.E. Reflectance nomenclature and directional reflectance and emissivity. *Appl. Opt.* **1970**, *9*, 1474–1475. [CrossRef]

33. GSM. GSM-Proton Precision Magnetometer. 2021. Available online: http://www.gemsys.ca/pdf/GSM_19T.pdf (accessed on 12 January 2021).

34. Pan Analytic. 2021. Available online: https://www.malvernpanalytical.com/ (accessed on 6 February 2021).

35. Guha, A.; Rao, D.A.; Ravi, S.; Kumar, K.V.; Rao, E.D. Analysis of the potential of kimberlite rock spectra as spectral end member using samples from Narayanpet Kimberlite Field, Andhra Pradesh. *Curr. Sci.* **2012**, *103*, 1096–1104.

36. Guha, A.; Kumar, K.V.; Porwal, A.; Rani, K.; Sahoo, K.; Kumar, S.A.; Singaraju, V.; Singh, R.; Khandelwal, M.; Raju, P.; et al. Reflectance spectroscopy and ASTER based mapping of rock-phosphate in parts of Paleoproterozoic sequences of Aravalli group of rocks, Rajasthan, India. *Ore Geol. Rev.* **2019**, *108*, 73–87. [CrossRef]

37. Shanmugam, S.; Srinivasaperumal, P. Spectral matching approaches in hyperspectral image processing. *Int. J. Remote Sens.* **2014**, *35*, 8217–8251. [CrossRef]

38. Clark, R.N.; Swayze, G.A.; Livo, K.E.; Kokaly, R.; Sutley, S.J.; Dalton, J.B.; McDougal, R.R.; Gent, C.A. Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems. *J. Geophys. Res. Space Phys.* **2003**, *108*, 1–44. [CrossRef]

39. Clark, R.N. Spectroscopy of rocks and minerals, and principles of spectroscopy. In *Manual of Remote Sensing: Remote Sensing for the Earth Sciences*; John Wiley & Sons: New York, NY, USA, 1999; Volume 3, pp. 2–58.

40. Guha, A.; Ghosh, B.; Chaudhury, S.; Rani, K.; Kumar, K.V. Spectral response of few important textural variants of chromitite and its potential in estimating relative grades of chromitite—A case study for chromitite of Nuggihalli Schist Belt, India. *Curr. Sci.* **2018**, *114*, 1721. [CrossRef]

41. Smith, R.E.; Singh, B. Recognizing, in lateritic cover, detritus shed from the Archaean Gossan Hill Cu-Zn-Au volcanic-hosted massive sulphide deposit, Western Australia. *Geochem. Explor. Environ. Anal.* **2007**, *7*, 71–86. [CrossRef]

42. Thébault, E.; Finlay, C.C.; Beggan, C.D.; Alken, P.; Aubert, J.; Barrois, O.; Bertrand, F.; Bondar, T.; Boness, A.; Brocco, L.; et al. International geomagnetic reference field: The 12th generation. *Earth Planets Space* **2015**, *67*, 79. [CrossRef]
43. Baranov, V. A new method for interpretation of aeromagnetic maps: Pseudo-gravimetric anomalies. *Geophysics* **1957**, *22*, 359–382. [CrossRef]
44. Roest, W.R.; Pilkington, M. Identifying remanent magnetization effect in magnetic data. *Geophysics* **1993**, *58*, 653–659. [CrossRef]
45. Spector, A.; Grant, F.S. Statistical methods for interpreting aeromagnetic data. *Geophysics* **1970**, *35*, 293–302. [CrossRef]
46. Roy, A.; Aina, A.O. Some new magnetic transformations. *Geophys. Prospect.* **1986**, *34*, 1219–1232. [CrossRef]
47. Blakely, R.J. *Potential Theory in Gravity and Magnetic Applications*; Cambridge University Press Science: London, UK, 1995.
48. Guo, L.; Meng, X.; Chen, Z.; Li, S.; Zheng, Y. Preferential filtering for gravity anomaly separation. *Comput. Geosci.* **2013**, *51*, 247–251. [CrossRef]
49. Taggart, J.E., Jr.; Siems, D.F. *Major Element Analysis by Wavelength Dispersive X-ray Fluorescence Spectrometry*; U.S. Geological Survey Open-File Report 02-223-T; U.S. Geological Survey: Reston, VA, USA, 2002; pp. T1–T9.
50. Taggart, J.E., Jr.; Lindsey, J.R.; Scott, B.A.; Vivit, D.V.; Bartel, A.J.; Stewart, K.C. Analysis of geologic materials by wavelength-dispersive X-ray fluorescence spectrometry. In *Methods for Geochemical Analyses*; Baedecker, P.A., Ed.; U.S. Geological Survey Professional Paper 1770; U.S. Geological Survey: Reston, VA, USA, 1987; pp. E1–E19.
51. Longhi, I.; Sgavetti, M.; Chiari, R.; Mazzoli, C. Spectral analysis and classification of metamorphic rocks from laboratory reflectance spectra in the 0.4–2.5 μ m interval: A tool for hyperspectral data interpretation. *Int. J. Remote Sens.* **2001**, *22*, 3763–3782. [CrossRef]
52. Kokaly, R.F.; Clark, R.N.; Swayze, G.A.; Livo, K.E.; Hoefen, T.M.; Pearson, N.C.; Wise, R.A.; Benzel, W.M.; Lowers, H.A.; Driscoll, R.L.; et al. *USGS Spectral Library Version 7: Data Series 1035*; U.S. Geological Survey: Reston, VA, USA, 2017; p. 61.
53. Crowley, J.K.; Brickey, D.W.; Rowan, L.C. Airborne imaging spectrometer data of the Ruby Mountains, Montana: Mineral discrimination using relative absorption band-depth images. *Remote Sens. Environ.* **1989**, *29*, 121–134. [CrossRef]
54. Sugden, T.J.; Deb, M.; Windley, B.F. The tectonic setting of mineralisation in the Proterozoic Aravalli Delhi Orogenic belt, NW India. *Dev. Precambrian Geol.* **1990**, *8*, 367–390.
55. Zhai, D.-G.; Liu, J.-J.; Wang, J.-P.; Yao, M.-J.; Wu, S.-H.; Fu, C.; Liu, Z.-J.; Wang, S.-G.; Li, Y.-X. Fluid evolution of the Jiawula Ag-Pb-Zn deposit, Inner Mongolia: Mineralogical, fluid inclusion, and stable isotopic evidence. *Int. Geol. Rev.* **2013**, *55*, 204–224. [CrossRef]
56. Holland, H.D. Granites, solutions, and base metal deposits. *Econ. Geol.* **1972**, *67*, 281–301. [CrossRef]
57. Porwal, A.; Carranza, E.J.M.; Hale, M. Tectonostratigraphy and base-metal mineralization controls, Aravalli province (Western India): New interpretations from geophysical data analysis. *Ore Geol. Rev.* **2006**, *29*, 287–306. [CrossRef]
58. Yadav, S.; Porwal, A. Controls on SEDEX mineralization in the Aravalli-Delhi Fold Belt: Insights from integrated 3D geological and geophysical modelling. In Proceedings of the EGU General Assembly Conference, Vienna, Austria, 23–28 April 2017; Volume 19, p. 1529.
59. Hood, P.J.; Holroyd, M.T.; McGrath, P.H. Magnetic methods applied to base metal exploration. In *Geophysics and Geochemistry in the Search for Metallic Ores*; Hood, P.J., Ed.; Geological Survey of Canada, Economic Geology: Montreal, Canada, 1979; pp. 77–104.
60. Whiteley, R.J. (Ed.) *Geophysical Case History of the Woodlawn Orebody*; Pergamon Press: New South Wales, Australia, 1981; p. 588.
61. Clark, D.; Tonkin, C. Magnetic anomalies due to pyrrhotite: Examples from the Cobar area, N.S.W., Australia. *J. Appl. Geophys.* **1994**, *32*, 11–32. [CrossRef]
62. Kumar, K.S.; Begum, S.P.; Srinivas, G.S.; Rao, M.S.H.; Prasad, P.P.; Seshunarayna, T.; Tiwari, R.K. Spectral analysis of magnetic data in the Ajmer-Sambar Sector, North-Central Rajasthan. *J. Geol. Soc. India.* **2018**, *92*, 368–372. [CrossRef]

Airborne Hyperspectral Data Acquisition and Processing in the Arctic: A Pilot Study Using the Hyspex Imaging Spectrometer for Wetland Mapping

Jordi Cristóbal [1,2,*], Patrick Graham [1], Anupma Prakash [1], Marcel Buchhorn [3], Rudi Gens [4], Nikki Guldager [5] and Mark Bertram [5]

1   Geophysical Institute, University of Alaska Fairbanks, 903 Koyukuk Dr., Fairbanks, AK 99775-7320, USA; pgraham@alaska.edu (P.G.); aprakash@alaska.edu (A.P.)
2   Efficient Use of Water in Agriculture Program, Institute of Agrifood Research and Technology, Fruitcentre, Parc Científic i Tecnològic Agroalimentari de Lleida 23, 25003 Lleida, Spain
3   Remote Sensing Unit, Flemish Institute for Technological Research (VITO), Boeretang 200, 2400 Mol, Belgium; marcel.buchhorn@vito.be
4   Alaska Satellite Facility, Geophysical Institute, University of Alaska Fairbanks, 903 Koyukuk Dr., Fairbanks, AK 99775-7320, USA; rgens@alaska.edu
5   Yukon Flats National Wildlife Refuge, U.S. Fish and Wildlife Service, 101 12th Avenue, Fairbanks, AK 99701, USA; Nikki_Guldager@fws.gov (N.G.); Mark_Bertram@fws.gov (M.B.)
*   Correspondence: j.cristobal@alaska.edu

Abstract: A pilot study for mapping the Arctic wetlands was conducted in the Yukon Flats National Wildlife Refuge (Refuge), Alaska. It included commissioning the HySpex VNIR-1800 and the HySpex SWIR-384 imaging spectrometers in a single-engine Found Bush Hawk aircraft, planning the flight times, direction, and speed to minimize the strong bidirectional reflectance distribution function (BRDF) effects present at high latitudes and establishing improved data processing workflows for the high-latitude environments. Hyperspectral images were acquired on two clear-sky days in early September, 2018, over three pilot study areas that together represented a wide variety of vegetation and wetland environments. Steps to further minimize BRDF effects and achieve a higher geometric accuracy were added to adapt and improve the Hyspex data processing workflow, developed by the German Aerospace Center (DLR), for high-latitude environments. One-meter spatial resolution hyperspectral images, that included a subset of only 120 selected spectral bands, were used for wetland mapping. A six-category legend was established based on previous U.S. Geological Survey (USGS) and U.S. Fish and Wildlife Service (USFWS) information and maps, and three different classification methods—hybrid classification, spectral angle mapper, and maximum likelihood—were used at two selected sites. The best classification performance occurred when using the maximum likelihood classifier with an averaged Kappa index of 0.95; followed by the spectral angle mapper (SAM) classifier with a Kappa index of 0.62; and, lastly, by the hybrid classifier showing lower performance with a Kappa index of 0.51. Recommendations for improvements of future work include the concurrent acquisition of LiDAR or RGB photo-derived digital surface models as well as detailed spectra collection for Alaska wetland cover to improve classification efforts.

Keywords: HySpex; hyperspectral image processing; classification; wetlands mapping; Arctic

1. Introduction

In Alaska, wetlands cover twenty-two percent of the state's area, according to the most recent survey carried out by [1]. However, over the past 200 years Alaska has lost less than one percent of its wetland area compared to an estimated fifty-three percent loss in other states in the US [2]. Climate change is likely to alter the historic stability of wetland conditions, particularly in the Arctic [1], as the hydrological inputs necessary for wetland formation change in response. The many climate destabilizers that Interior Alaska

is experiencing may cause wetland degradation specifically through thermal erosion, permafrost thaw, changing snow cover amounts and duration, shifting precipitation patterns, and paludification [3–6]. In addition, the degradation or loss of wetlands may result in significant changes in weather systems and alter precipitation patterns themselves [7]. Wetlands in Alaska represent an important habitat that provide essential ecosystem functions and many benefits to humans, plants, and animals from local to continental scales. These benefits include food and habitat for vegetation, wildlife, fish and shellfish species, food and habitat for human subsistence gathering, flood storage and stormflow modification, ground-water recharge and discharge, and the maintenance of water quality [7,8].

Monitoring and describing changes to all physical, chemical, and biological parameters for the Alaska wetlands is unrealistic due to their scope and complexity. However, using identified and interpreted proxies to assess the condition of an environment is a proven ecological monitoring method [9]. Complex interactions between geology; topography; climate; and physical, biological, and chemical systems result in various hydrological regimes and wetland types, each with their own plant and animal species assemblages. With the increase in the extremity of hydrological regime, the degree of specialization and fidelity of plant species increases [10]. Therefore, vegetation species assemblages may be used to infer wetland type and is the basis for the widely utilized Cowardin wetland classification system [11,12]. This ecological classification system was developed by the U.S. Fish and Wildlife Service to establish consistent terms and definitions used in inventory of wetlands and to provide standard measurements for wetland mapping. It is based on vegetation cover interpreted from aerial photos and is also widely used in Canada for wetland mapping [11,12]. Traditionally, vegetation mapping and wetland inventories require labor-intensive, costly, and time-consuming field work, including taxonomical information, collateral and ancillary data analysis, and the visual estimation of percentage cover for each species [13]. This workload is exacerbated in Alaska from the added difficulty of accessing wetland areas that are far from population centers, road systems, or aircraft landing strips, and the need to traverse difficult terrain. The national wetland inventory for Alaska is based almost entirely on 1978–1986, 1:60,000-scale, color-infrared imagery collected as part of the Alaska High Altitude Photography Acquisition Program (AHAP), with only 42 percent of the state having been mapped as of 2019 [14]. More recent inventories and studies of Alaskan wetlands [1,3,5,15,16] are often based on medium-resolution multispectral earth observation imagery, such as Landsat. Though orbital hyperspectral sensors such as Hyperion on board the EO-1 satellite have acquired high latitude imagery, there are several drawbacks to their use, especially in wetland studies, including inadequate spatial resolution for upland wetland delineation and mapping [17] and substantial periods of cloud cover during summer in high latitudes that reduce the chance of quality data acquisition during satellite overpass.

Prior to 2015, there was no direct access to a research-grade hyperspectral imaging system in Alaska that could be deployed for airborne hyperspectral remote sensing [18]. It was only in 2017–2018 that National Aeronautics and Space Administration (NASA) acquired Airborne Visible Infrared Imaging Spectrometer Next Generation (AVIRIS NG) data over Alaska's boreal forests [19]. The HySpex airborne data set, such as presented in this pilot study, complements the NASA AVIRIS NG dataset and has the advantage that it can be collected locally at targeted sites with a greater and controlled frequency, and at lower costs than the NASA AVIRIS NG dataset [20]. The campaign also has other general advantages and disadvantages of airborne remote sensing for wetland mapping [21].

Compared to field-based wetland mapping, remote sensing techniques offer an economical and practical alternative to discriminate and estimate biochemical and biophysical parameters of wetland species [13] and to assist academic researchers and government agencies in mapping and monitoring wetlands [17,22]. In particular, hyperspectral imaging, also called imaging spectroscopy, has the advantage of capturing the distinct spectral signatures of land covers associated with wetlands within the 400 to 2500 nm spectral region [23]. However, hyperspectral imaging in high latitudes does come with challenges

from low sun angles that introduce pronounced anisotropic effects to images [18]. The goal of this study was to demonstrate the capability of airborne hyperspectral imaging for wetland mapping using the Yukon Flats as a test case by means of (a) commissioning a hyperspectral imaging system, HySpex, in a small aircraft for airborne data acquisition in high-latitude environments; (b) correcting for geometric and radiometric distortions that are uniquely inherent in a high-latitude environment; (c) developing image processing protocols to generate prototypes of seamless mosaics, hypercubes, and thematically classified image products; and (d) mapping major wetland types in selected sites in the Yukon Flats.

2. Study Area

The Yukon Flats (Figure 1 top panel) is a 25,900 km^2 wetland located within the Yukon Flats National Wildlife Refuge in eastern Interior Alaska. It is of international significance as a breeding area for migratory birds who use all the major flyways in North America. Three pilot study areas were selected within the Yukon Flats representing the wide variety of available wetland habitats.

The Yukon Flats area is bisected by the Yukon River and the Arctic Circle, situated between the Brooks Range to the north and the White and Crazy mountains to the south, and extends 360 km east to west (65°45′ and 67°30′ north latitude and 142°30′ and 150°00′ west longitude). The area has a low relief and is situated upon a broad plain of active and abandoned poorly drained alluvial floodplain deposits with a high water table [24]. The Yukon Flats experience a cold continental subarctic climate, with extreme temperatures and solar radiation between summer and winter. Although the Yukon Flats lack extensive meteorological data records, the Fairbanks International Airport located roughly 150 km south of the refuge provides National Climatic Data Center values from 1951 to 2009 [3] with a mean annual temperature of −3 °C, mean January temperature of −23 °C, and mean July temperature of 17 °C. The annual precipitation at Fairbanks International Airport is 26.7 cm water equivalent with snow cover from October through to April [3].

Much of the Yukon Flats is underlain by discontinuous permafrost [16], which substantially affects the surface and subsurface hydrology. Approximately 40,000 shallow lakes cover an area of 1147 km^2, and are mainly oxbow or thermokarst-type lakes [25]. Although the area receives little annual precipitation, permafrost layers inhibit water movement and drainage and permafrost thaw leads to the development of shallow thermokarst lakes [3,24–26]. In addition, the short ice-free season in the region limits water loss due to evapotranspiration [27]. There are closed basin lakes (no outlet) and open basin lakes (with outlets) found within the Yukon Flats, with some lakes having direct contact with the groundwater table [24]. Due to the shallowness of the lakes and diurnal mixing there is little stratification within the water column during summer. In the winter, many lakes freezse to the bottom leaving these waterbodies devoid of permanent fish stock. The shallow lakes, developed in the complex carbonate-rich alluvial sediments of the basin, have highly variable nutrient concentrations, but most water bodies within the Yukon Flats are rich in nutrients and are either eutrophic or hypereutrophic. The high nutrient levels of many lakes allow for high populations of phytoplankton and invertebrates. Moreover, many lakes (~25%) are slightly brackish [24], with the potential of increased salinity concentrations and eutrophication as annual temperatures continue to rise and evaporation and permafrost degradation increase [28].

In addition to the abundant lakes and waterways, the Yukon Flats host a variety of habitats, with mixed boreal forest dominated by black spruce (*Picea mariana*) and white spruce (*Picea glauca*) covering much of the area. Stands of Alaska birch (*Betula neoalaskana*), and quaking aspen (*Populus tremuloides*) are common, along with willow (*Salix* ssp.) and alder (*Alnus* ssp.) thickets. Graminoid and sedge (*Carex* spp.) grasslands occupy many areas, while emergent plants (*Equisetum* spp., *Typha* spp., *Scirpus* spp.) are found within the wetlands and peripheries of lakes. The abundance and diversity of land cover types, along with the large amount of shallow, nutrient-rich water bodies has made the Yukon Flats an important breeding habitat for waterfowl [28], with over 100 different species and an

estimated 0.5 to 1.5 million ducks, geese, and swans nesting there annually, including lesser scaup (*Aythya affinis*), white-winged scoters (*Melanitta fusca*), and horned grebes (*Podiceps auratus*) [29].
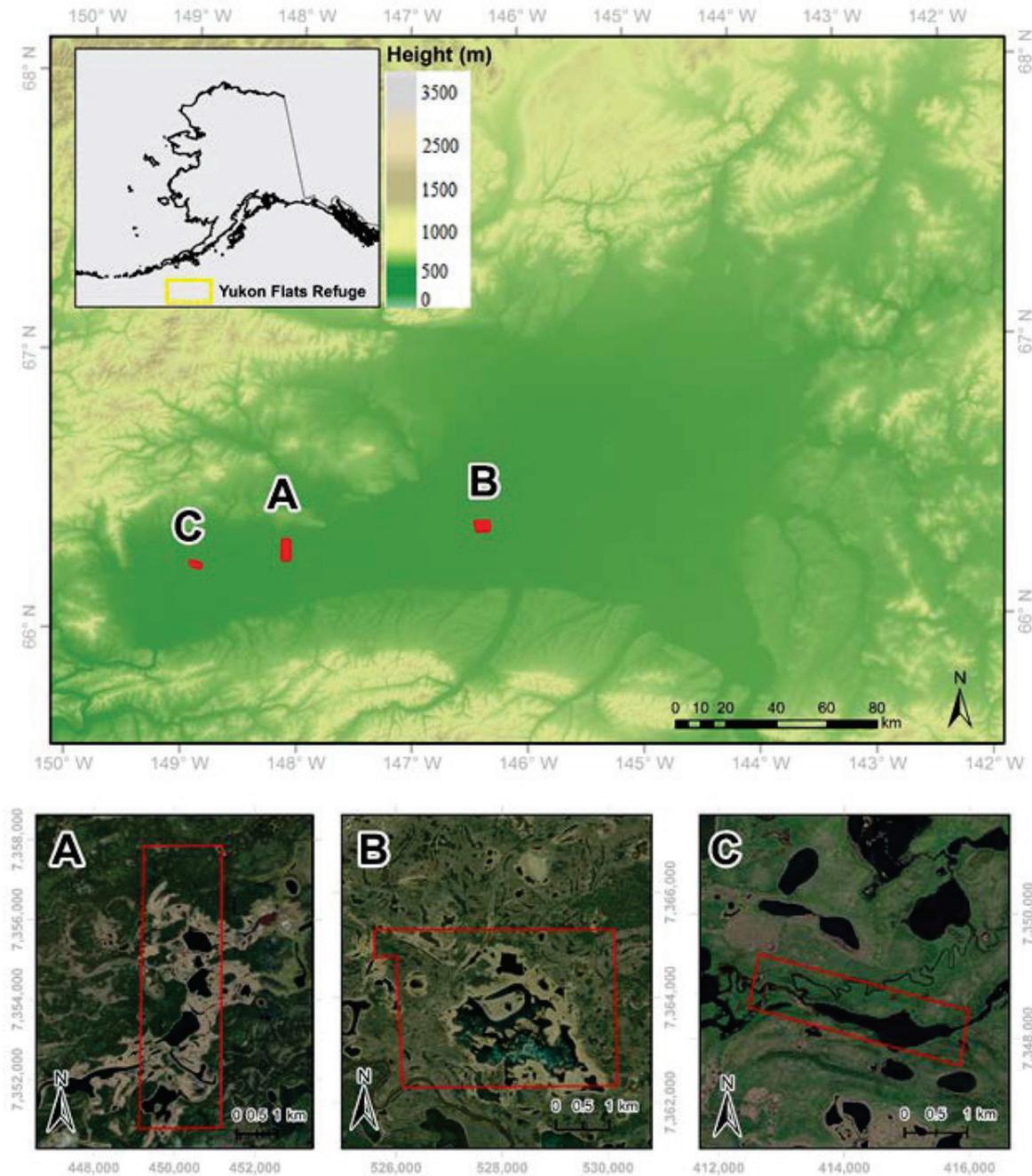
Figure 1. Map of the Yukon Flats National Wildlife Refuge, Alaska, showing flight areas labeled A, B, and C. Upper panel in geographic projection system and lower panels in UTM-6N projection system. All panels are in datum WGS84.

3. HySpex System Commissioning and Data Acquisition

3.1. HySpex Hyperspectral Imaging System

The HySpex hyperspectral imaging system, manufactured by Norsk Elektro Optikk (NEO), is configured for airborne acquisition using two sensors connected to a data acquisition unit (DAU) and an inertial measurement unit (IMU). The sensors are pushbroom

(or along-track) scanners. Both sensors have low stray light levels, low polarization sensitivity, and high signal to noise ratios that allow for imaging highly dynamic scenes. The VNIR-1800 sensor samples spectra in the 400 to 1000 nm region, with 182 spectral channels sampling 3.26 nm per channel and a 16-bit radiometric resolution. It has a 17° field of view, with 1800 spatial pixels of 0.16 mrad across-track by 0.32 mrad along-track. The SWIR-384 sensor samples spectra in the 950 to 2500 nm region, with 288 spectral channels sampling 5.45 nm per channel, with a 16-bit radiometric resolution. It has 384 spatial pixels with a resolution of 0.73 mrad along-track by 0.73 mrad cross-track, and a 16-degree field of view. Both sensors were equipped with a field expander optics that doubled the pixel field of view (FOV) for the VNIR and SWIR-sensors to 34 degrees and 32 degrees, respectively.

The DAU is a purpose-built Windows 7 machine that runs the HySpex acquisition and control software, HySpex AIR, and provides power to the HySpex sensors. The DAU software is monitored and controlled by the system operator in air via a retro-reflective touch screen. The IMU is an IMAR iTrace RT-F400 IMU/Global Positioning System (GPS). Coupled with a GPS receiver antenna, the IMU provided all kinematic measurements during acquisition, such as the acceleration, angular rate and roll, pitch, and yaw of the aircraft, as well as position and velocity.

3.2. Integration of HySpex into Aircraft

A Found Bush Hawk, owned and operated by the U.S. Fish and Wildlife Service, served as the acquisition platform for the HySpex system (Figure 2a). The VNIR-1800 and SWIR-384 sensors, along with an IMAR iTrace RT-f400 IMU/GPS, were attached to a passive vibration dampening mount (Figure 2b) and secured to the aircraft with the across-track sensor line perpendicular to the aircraft's flight direction (Figure 2c). A GPS antenna connected to the IMU by coaxial cable was mounted to the roof of the aircraft and X and Y offsets between the antenna and IMU were measured and input into the IMU for the georectification of images during post processing. Field of view expander lenses were attached to each sensor (Figure 2d), increasing the VNIR-1800 and SWIR-384 across track field-of-views of 17° and 16° to 34° and 32°, respectively. This allowed for maintaining a flight altitude that would capture an approximately one-meter ground pixel resolution in the SWIR image while reducing the number of flight lines needed to cover a selected flight area with the required 40 to 75% side-lap for the later mosaicking of the scene. Finally, the system was controlled in flight by the compact, high-performance DAU, connected with a 1 terabyte solid state hard drive and a compact, touch screen flat-panel monitor for in-flight operation and system-monitoring (Figure 2e).

A QUINT-UPS 24-volt/3.4 amp-hour uninterruptible power supply provided power to the HySpex system in flight and was fed by a 12-volt to 24-volt DC converter plugged into the aircraft's electrical system. The system was powered and tested prior to flight to ensure reliable operation and an unobstructed view through the aircraft's floor aperture. While performing a ground test acquisition of the HySpex system, a visible and near-infrared high contrast checkboard was placed below the aircraft aperture and moved perpendicular to the detector lines of the HySpex sensors to simulate aircraft movement over ground. The VNIR and SWIR images were then inspected visually to determine if the view angles of both sensors would be unobstructed by the aircraft fuselage during flight acquisition (Figure 2e).

During camera integration in the aircraft, the offset between the IMU and GPS antenna receiver and the offset between the two HySpex cameras were measured with sub-centimeter precision. This information was important to incorporate in the processing stage to achieve the precise co-registration of the images acquired by both cameras.

3.3. Flight Planning and Data Acquisition

On 2 September and 3 September 2015, three areas at the Yukon Flats Refuge were flown (see Figure 1). The targeted spatial resolution was 0.5 m for the VNIR-1800 and 1 m for the SWIR-384 at 2451 feet above ground level at 165 km·hr^{-1} flight speed. These areas

were chosen based on the water chemistry and ecological habitat. Two flight areas, A and B (Figure 1), were identified by Refuge personnel as "high priority" targets due to their historical knowledge of the plant and waterfowl communities. The camera time frame period was provided by the system manufacturer NEO and the maximum frame period, to prevent under-sampling the ground scene, was determined by flight altitude and ground speed. A 40% side-lap between flight lines was used to adjust for aircraft banking and to aid in the georectification process. A test acquisition flight line was flown over the target area with increasing integration times, until at least one spectral band was saturated for the relevant ground cover. The integration time was then reduced by roughly 10% of the saturation value to optimize the data quality for the ground cover type, while still having a margin for variability in reflectance levels throughout the spectral bands.

Figure 2. Commissioning the HySpex system for airborne acquisition. (**a**) Refuge Found Bush Hawk; (**b**) VNIR-1800, SWIR-384, and IMU/GPS units mounted to vibration dampening plate; (**c**) field of view expander optic attached to VNIR-1800 and SWIR-384 sensors; (**d**) detail of HySpex sensors in aircraft; (**e**) overhead view of the system secured and operating in aircraft; (**f**) testing system visibility through floor aperture of the aircraft.

Images were acquired within two hours before and after solar noon. Study areas A and B were flown either east to west or north to south to account for and reduce the BRDF effects which are pronounced in areas of high latitude. A total of 7, 7, and 2 flight lines were taken for areas A, B, and C, respectively, with a total area coverage of 14.2, 12.5, and

3 km², respectively. Raw image files for the VNIR and SWIR sensors and navigational data files from the IMU/GPS were recorded and stored on the DAU for processing.

4. Hyperspectral Data Processing

The HySpex data processing chain, originally developed by the German Aerospace Center (DLR) [30], was adapted for this study to address the challenges of hyperspectral imaging at high latitudes (see Appendix A for a graphic depiction of the image processing flow chain). This processing method uses raw imagery from the two HySpex cameras and positional and navigational data to produce robust, geo-registered surface radiance and reflectance products. In addition to the standard geometric and radiometric corrections, processing workflows were also added to minimize the effects of BRDF.

4.1. Raw Images to At-Sensor Radiance Images

Both the VNIR and SWIR cameras were radiometrically calibrated by the manufacturing company, NEO, using an integration sphere. This same calibration was applied prior to aerial surveys to ensure that the acquired images had high radiometric quality. The 16-bit raw integer value data acquired per flight line were converted to at-sensor radiance value imagery according to [31], as summarized below. The first step is image acquisition in raw VNIR and SWIR file format using image metadata. For each flight line, data were recorded on the DAU during image acquisition in 16-bit digital number (DN) format as following:

$$DN[i,j] = Ni[i,j] \cdot QE[i] \cdot RE[i,j] \cdot BG[i,j], \tag{1}$$

where i is the spectral band number; j is the spatial pixel number; DN are the digital numbers (0 to 65,535); Ni is the number of incoming photons corresponding to spatial pixel j and band number i during the integration time t; QE is the quantum efficiency (photoelectron to photon ratio) of the total system, including optics and detector for band I; SF is a scaling factor expressing DN per photoelectron (scaling factor is determined during the radiometric calibration of the instrument); and BG is a background matrix.

Raw data were then converted to real time calibrated data, CN, using the relationship between real-time calibrated DN values and the incoming light, expressed as:

$$CN[i,j] = \frac{DN[i,j] - BG[i,j]}{RE[i,j]} \cdot dw. \tag{2}$$

Finally, the at-sensor absolute radiance in W·m⁻²·sr⁻¹·nm⁻¹ was computed as follows:

$$LN[,j] = \frac{CN[i,j] \cdot h \cdot c \cdot SL}{QE[i] \cdot SF \cdot dw \cdot t \cdot A \cdot W \cdot Dl[i] \cdot 1[i]}, \tag{3}$$

where t is the integration time, h is the Planck constant, c is the speed of light, A is the area of the light entrance aperture, Ω is the solid angle of a single pixel, $\Delta\lambda$ is the spectral sampling of the camera in nanometers, λ is the wavelength in nanometers, and SL is the Global Land Survey Digital Elevation Model (GLSDEM) scaling factor determined during calibration.

4.2. Image Orthorectification

The orthorectification model and geocoding of VNIR and SWIR at-sensor radiance flight line was performed with PARGE (ReSe Applications software) using the navigation files generated by the IMU/GPS for a flight area. A subset and resampled portion of the 90 m resolution GLSDEM was used for terrain elevation data input and resampled to the final spatial resolution of the hyperspectral imagery, which was 1 m for both VNIR and SWIR bands.

To ensure a proper alignment between both the VNIR and SWIR cameras, a boresight calibration was performed before the orthorectification process by flying over the University of Alaska Fairbanks campus prior to the image acquisition at the study area. Ground control points with a high GPS accuracy were identified and then used for aerotriangulation to

determine the image attitude with respect to a local mapping frame and compared to the IMU-derived attitude matrix to derive a boresight matrix. This resulted in a boresight calibration offset that was applied in the georectification process to correct for the angular misalignment between the frames of reference of the IMU and the VNIR and SWIR cameras.

Orthorectification accuracy was evaluated by performing an automatic image to image registration between VNIR band 171 (954 nm λ) and SWIR band 2 (954 nm λ) flight line orthoimages. One hundred tie points were generated between corresponding pixels of the two images and their root mean square error (RMSE) was evaluated. An RMSE of less than 0.5 m, indicating sub-pixel accuracy between tie points, was deemed satisfactory and indicated that both cameras were successfully aligned. Flight lines where then layer stacked in a single hypercube (see Figure 3).
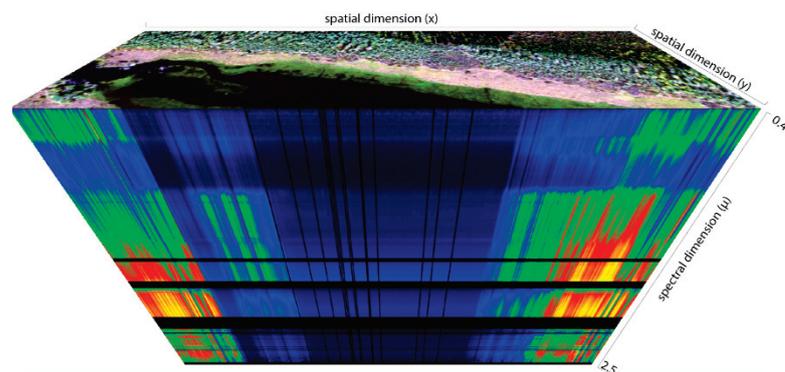
Figure 3. Example of a HySpex hypercube for the flight area C. A flight line hypercube with spatial dimensions, x and y, and spectral dimension, μ. Each spatial pixel in the flight line has a corresponding spectral signal from 0.4 to 2.5 μm. Dark horizontal regions in the spectral dimension were masked out due to water absorption bands.

Although the VNIR images were collected at a nominal spatial resolution of 0.5 m, they were resampled using the nearest-neighbor method to a 1m resolution to match the spatial resolution of the SWIR images.

4.3. Radiometric Correction

After generating an orthorectified hypercube, a radiometric correction was applied using ATCOR-4 (ReSe Applications), which is specially designed for HySpex hyperspectral data. ATCOR-4 performs both atmospheric and topographic corrections for airborne sensors for optical regions (0.4 to 2.5 μm) using topography, image geometry, aerosols, water vapor information, and sensor model information [32]. This procedure was necessary to correct for the effects of atmospheric water vapor, optical thickness of the atmosphere, position of the sun, and differences in illumination caused by topography. A brief explanation about ATCOR can be found in [33]. The DEM used in the orthorectification process also provides information on the elevation, slope, aspect, and scan angle for ATCOR-4 corrections. Aqua MODIS water vapor product (MYD_05), taken about the same time as the airborne imagery, was used as a proxy for water vapor information. This product has previously shown good agreement when compared with ground data [34]. Finally, the flight line navigation files and the sensor model generated during calibration were also used by ATCOR-4. Images were collected on days with clear skies and with a default visibility value of 23 km, which was used to account for aerosol load.

An additional consideration when correcting radiometry is to minimize the BRDF impact from the variation in viewing and solar illumination geometry. Anisotropic reflectance can significantly alter the measured radiometry and surface reflectance of the same land cover type depending on the solar illumination, wavelength, surficial properties, and viewing angle of the sensor and must be corrected to ensure consistent radiometry across a scene. High solar zenith angles that are exhibited in high-latitude regions such as

Alaska [35] further exacerbate this effect. BRDF effects, such as those in our study area, are apparent in scenes where the view and/or solar zenith angles vary over a large angular range, causing the normally processed reflectance value for vegetated surfaces to deviate up to 30% [32]. The influence of the BRDF effects can be split into topographical effects dependent on the slope and aspect of the terrain and the structural and optical properties of the vegetation.

The topographic effects were partly accounted for in the topographic correction (image orthorectification), thus the BRDF effect correction method (BREFCOR) was applied to the radiometrically corrected hypercube imagery in ATCOR-4 accounting for the vegetation structure effects. The BREFCOR method corrects observer BRDF effects by applying a fuzzy classification index that covers all surface types as a unified continuous index, and then fits the Ross-Li-sparse reciprocal BRDF model to all flight line images within a flight area and land cover classes to obtain a generic BRDF correction function [32,36].

4.4. Spectral Binning and Final Mosaic

The hypercube output by BREFCOR contained 457 discrete spectral bands. Spectral binning was applied to the dataset to reduce the significant noise generated by high instrument sensitivity. After applying spectral binning, the hypercube was reduced to a total of 229 bands (see Table 1 and Figure 4). Additionally, atmospheric water vapor bands (from 1340 to 1411 nm and from 1786 to 1829 nm) were masked out of the data, as they contained no valuable information for the purposes of ground cover classification.
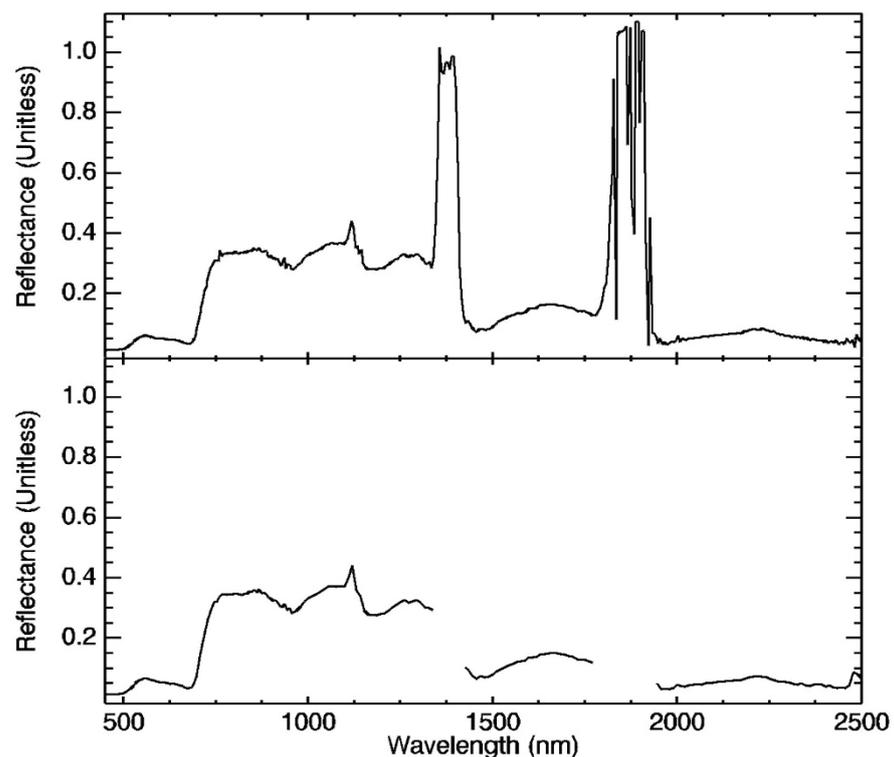
Figure 4. An example of the effect of spectral binning on the spectral profile of a single pixel of vegetation. The top panel shows the unbinned profile with 457-bands and the bottom panel shows the profile from 2x-binned 229-bands. Note the increased noise in the unbinned profile (top panel).

Once all flight lines were orthorectified, radiometrically corrected, and spectrally binned, they were mosaicked together to produce a single georeferenced hypercube image of the flight area. It was ensured that at least 25 tie points were found between two adjacent flight lines, leading to an RMSE for each flight area of less than 0.5 pixels.

Table 1. Number of bands, spectral range, and spectral resolution before and after the binning process.

	Sensor	Bands Per Hypercube	Spectral Range (nm)	Spectral Resolution Per Band (nm)
Without Spectral Binning	VNIR-1800	1–171	416–955	3.26
	SWIR-384	172–457	960–2509	5.45
2x Spectral Binning	VNIR-1800	1–85	418–950	6.33
	SWIR-384	86–229	957–2508	10.86

5. Wetland Mapping

5.1. Category Definition

Flight areas A and B were classified following a six-category legend (see Table 2). These categories included the most relevant and common vegetation found in the study area and were those that could be detected in at least 1 m resolution pixels. The accessibility and constraints posed by the remote study sites made it impossible collect field spectra.

Table 2. Wetland mapping classes and description.

Class Attribute	Class Description
Water	-Areas of open water lacking emergent vegetation.
Equisetum spp. and emergent vegetation	-Areas where perennial herbaceous vegetation accounts for 75–100% of the cover and the soil or substrate is periodically saturated with or covered with water.
Bog, grasses, and sedge	-Areas characterized by natural herbaceous vegetation including grasses and forbs; herbaceous vegetation accounts for 75–100% of the cover.
White/black spruce	-Areas of open or closed evergreen forest dominated by tree species (primarily *Picea mariana* and *Picea glauca*) that maintain their leaves all year, with a canopy that is never without green foliage.
Deciduous vegetation (including shrubs)	-Areas dominated by trees tree species (primarily *Betula neoalaskana* and *Populus tremuloides*) and shrubs characterized by natural or semi-natural woody vegetation with aerial stems, generally less than 6 m tall, with individuals or clumps not touching to interlocking (including *Salix* spp., and *Alnus* spp.) that shed foliage simultaneously in response to seasonal change.
Bare ground	-Areas characterized by bare rock, gravel, sand, silt, clay, or other earthen material, with little or no "green" vegetation present regardless of its inherent ability to support life. Vegetation, if present, was more widely spaced and scrubby than that in the "green" vegetated categories.

5.2. Training and Test Areas Selection and Band Selection

Training areas were photointerpreted and digitized using the 30 m spatial resolution "Vegetation Map and Classification: Northern, Western, and Interior Alaska" [37], which served as an auxiliary dataset. This map, developed in 2012, was derived from 18 previous regional maps and is the most recent and best vegetation map available for the study area. The visual interpretation of a HySpex natural color composite generated from bands 12 (λ = 487 nm), 41 (λ = 671 nm), and 55 (λ = 760 nm) and a HySpex false color composite generated from bands 41 (λ = 671 nm), 55 (λ = 760 nm), and 62 (λ = 804 nm) further aided the training area selection for vegetation classification. A total of 70% of total training areas were used as a training set for supervised classification, with the remaining 30% used in a post-classification accuracy assessment. The assignment of training and validation samples was performed randomly for each category.

Before image classification, a detailed spectral analysis of defined categories was applied to reduce data redundancy of hyperspectral imagery as well as to derive a subset of significant bands. For this purpose, the BandMax algorithm in ENVI was used. This algorithm increases classification accuracy by determining an optimal subset of bands to help a user separate the user's targets from known background materials [38]. Using the training areas in A and B as user-defined target and background materials, the BandMax algorithm identified a subset of 120 optimal bands for further wetland mapping.

5.3. Image Classification Methods: Hybrid Classification, Maximum Likelihood and Spectral Angle Mapper (SAM)

Three methods were used for classifying wetlands: (a) maximum likelihood supervised classification, (b) a hybrid classification (unsupervised + supervised), and (c) spectral angle mapper (SAM). All classifications were run with the ENVI software.

For the maximum likelihood classification, probability thresholds were set between 0 and 1 and a single probability threshold value of 0.5 was used for all classes. The hybrid classification consisted of an IsoData unsupervised classification combined with a supervised classification [39]. We ran IsoData several times with a minimum number of classes ranging from 7 to 48, a number of iterations from 20 to 30, a minimum pixels per class from 25 to 50, a minimum class distances from 5 to 15, and maximum merge pairs from 2 to 6. Finally, in the supervised step, depending on the fidelity and representativeness thresholds, resulting clusters from the unsupervised classification class were linked to a corresponding thematic class or left as unclassified. SAM is a physically based spectral classification that uses an n-D angle to match pixels to reference spectra [40]. The average spectra from the training areas were derived and different threshold configurations (ranging from 0.15 to 1.75) for each category were used to run SAM.

6. Results and Discussion

6.1. Commissioning and Data Acquisition

The HySpex system was successfully commissioned into a Found Bush Hawk aircraft. The designed flexibility and modular nature of the HySpex system made it relatively straight forward to test, transport, install, and remove multiple times before the acquisition flights. This allowed us to perform preflight measurements and test components of the system at a local hanger. The custom made aluminum mounting plate for the vibration dampening mount and wooden mounting board for the DAU and power system, with measured mounting hardpoints for the camera system and aircraft floor plates, ensured that the HySpex system could be easily mounted and unmounted as needed. This is a particularly useful feature for aircraft with multiple concurrent missions and for remote sensing in high latitude regions such as Alaska that witness rapidly changing weather conditions. This flexible setup also reduced costs by not committing aircraft use solely to data acquisition. Experience gained in this study will be helpful for commissioning similar airborne data acquisition systems.

6.2. Image Processing

6.2.1. Systematic VNIR Sensor Response Drop Correction and Systematic Stripping in VNIR and SWIR Spectral Bands

During initial geometric processing, a systematic drop in the VNIR sensor response was discovered along the edges of the acquired imagery (Figure 5). To evaluate the systematic drop, the VNIR camera was tested using a calibrated integration sphere with and without field of view expander optics attached. The drop was only present when the expander optics were attached. To maintain a high radiometric fidelity, the low sensor-response areas along the edges of the raw image files were masked out for all flight lines. However, this resulted in minor data gap of around 0.14 ha in the final mosaics for area B. This illustrates the importance of increasing the side-lap in the flight planning to account for HySpex sensor response drop areas when using the field expander on the VNIR camera.
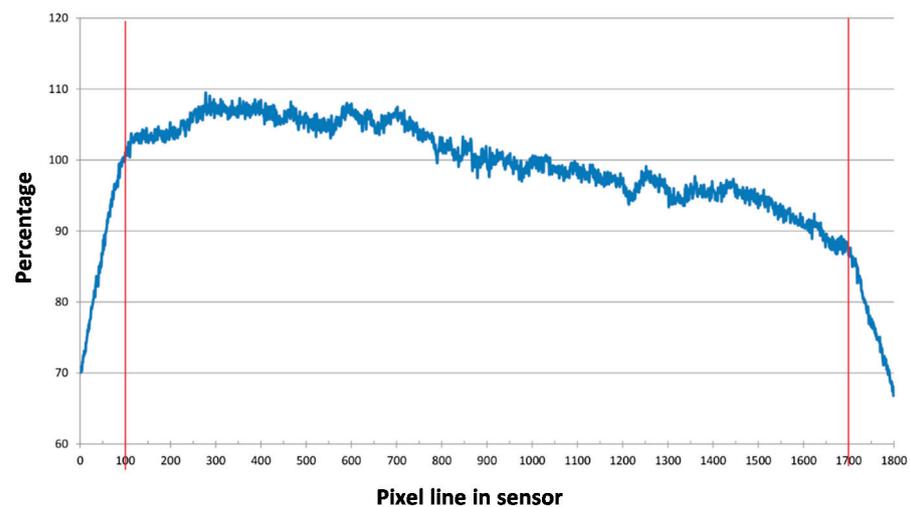
Figure 5. Sensor response drop on edge pixels of VNIR camera flight line, shown by red line.

6.2.2. Geometric and Radiometric Corrections

The Found Bush Hawk proved to be an excellent acquisition platform for HySpex imagery that led to a robust geometric correction, with an RMSE of less than 0.5 between flight lines. However, issues with geometric correction were detected due to the relative instability that are typical of flight in smaller single-engine aircraft. Image artifacts were due to the inherent turbulence (high roll, yaw, pitch) associated with small aircraft, which could not be automatically compensated by the IMU. These errors were corrected through data post-processing. This is a quality convenience tradeoff. Although larger aircraft platforms may provide more stability and may need fewer flight attitude corrections while collecting imagery in a flight line, smaller aircraft provide more flexibility when commissioning instrumentation and flying over remote areas. Overall, the low RMSE of less than 0.5 m in geometric correction obtained for the whole dataset collected in this study shows the configuration robustness for both the platform and geometric correction choice that yielded high-end hyperspectral orthomosaics for high latitudes (Figure 6).

Real-Time Kinematic (RTK) GPS measurements that provide more accurate GPS readings can be used for more accurate geometric correction. Subscription to satellite-based augmentation systems, such as OmniStar that utilizes L1/L2 carrier-phase correction signals with dual-frequency compatible receivers such as the IMAR iTRACE, or the use of post-processing differential GPS correction in regions with GPS receiver base stations may also be used to increase the accuracy of georectification.

Carefully planned flight lines reduced the differential illumination affects due to the position of the sun relative to flight path (Figure 6). The specific flight plan designed for high latitude areas for image acquisition (flying north to south and east to west) was successfully applied and BRDF effects were minimized in the dataset, thus allowing for seamless mosaics. However, flight lines in area C displayed strong BRDF effects as they were not collected using the same specific directions as the other areas. This was particularly visible in the final mosaic (Figure 6 panel C) where the lower flight line was brighter than the upper one, yielding approximately 32% reflectance difference between the two flight lines. This strongly suggests that for high latitude hyperspectral image acquisition, flight plans should be designed in advance flying north-to-south or east-to-west to minimize BRDF effects. Another factor that impacted the data quality in this study was the presence of a low cirrus cloud in an otherwise clear sky day that obstructed one of the flight lines in area B, altering reflectance values. During geometric and atmospheric processing, several bands within the VNIR and SWIR images were found to exhibit systematic striping across the image in the along-track (or flight) direction. To ensure high radiometric fidelity, bands exhibiting systematic striping and spectral noise from water vapor 1 to 3 (416 to 422 nm), 99 to 102 (726 to 736 nm), 122 to 128 (799 to 818 nm), 136 to 150 (843 to 888 nm), 156 (907 nm),

and 200 to 204 (1112 to 1133 nm) were omitted from the final data set. This was carried out before spectral binning and classification processing.
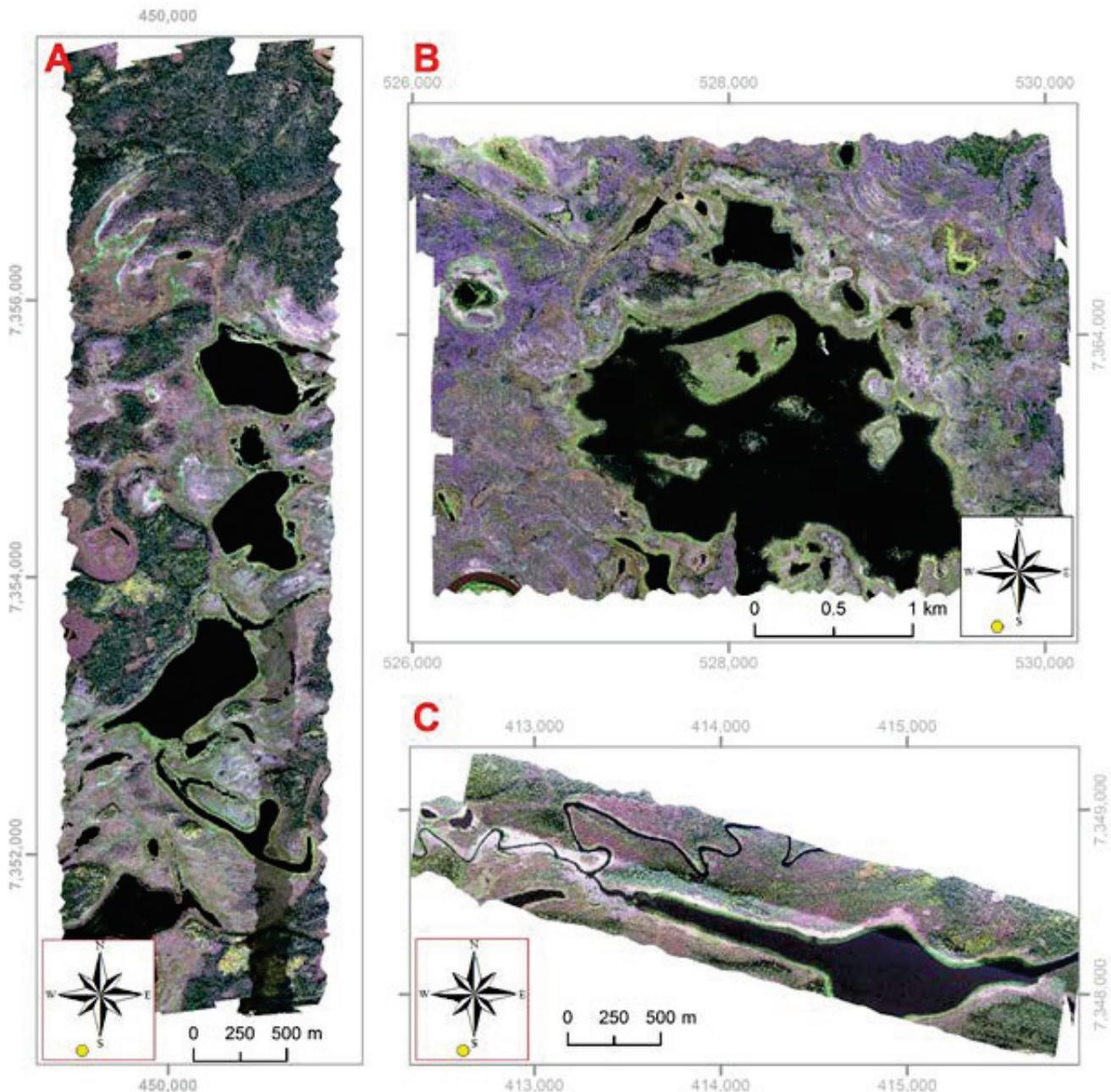
Figure 6. True color (red: band 38 (652 nm); green: band 22 (0.550 nm); blue: band 11 (481 nm)) orthomosaics for areas **A**, **B**, and **C** at a 1 m spatial resolution. A flight area shows a darkened flight line due to a low cirrus undetected when acquiring the imagery. C flight area shows Bidirectional reflectance distribution function (BRDF) effect between two flight lines flown northeast to southwest where the flight line at the top of the image is darker for the same ground cover type. The position of the sun is shown below the north arrow for each image. Coordinates are in UTM-6N and datum is in WGS-84.

After removing these bands, averaged spectra (training and test spectra dataset) for all land categories displayed a good visual performance except for water, which showed negative values (Figure 7, negative values for water were excluded for a better spectral comparison). In the water bodies, ATCOR4 undercorrected the water spectra, giving negative values beyond 820 nm where water reflectance was supposed to be 0, similar to those reported by [41]. For image classification purposes, negative values beyond 820 nm

were set to 0. To compute absolute reflectance, in situ spectra as well as better water vapor estimates are needed for radiometric correction.
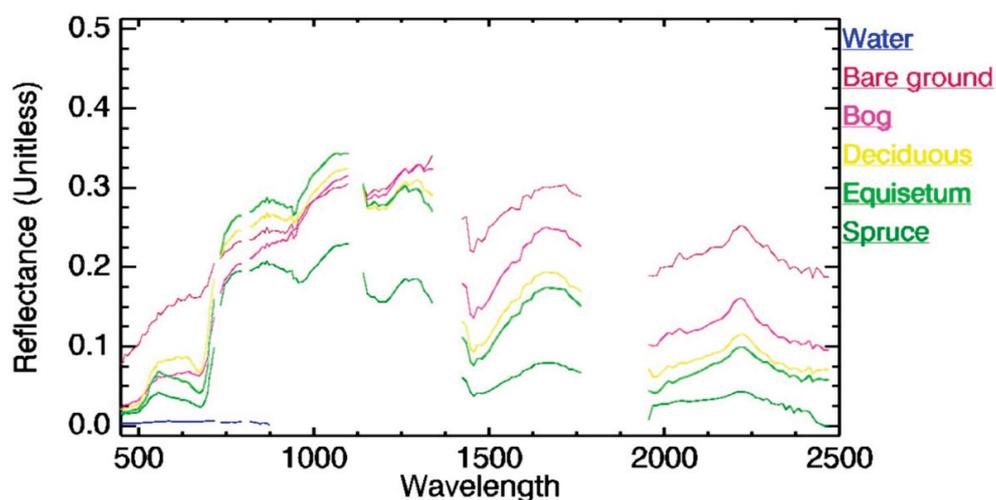
Figure 7. Averaged spectra extracted from training and test areas for all classification categories. Note that negative values for water were removed for better visual comparison of the spectra.

Image-derived spectra for spruce and deciduous vegetation were visually compared to similar spectra taken from available spectral libraries. For the spruce category, 12 samples of white spruce (*Picea glauca*) and black spruce (*Picea mariana*) spectra [42] were averaged. For the deciduous category, 12 samples of paper birch (*Betula neoalaskana*) and aspen (*Populus tremuloides*) spectra were also averaged [42]. A total of 100 pixels representative of both categories in study areas A and B were sampled, excluding shadowed areas for better spectra comparison.

HySpex spruce spectra were similar in shape to published spruce spectra (Figure 8 lower panel) but had a lower overall reflectance, likely due to the differences in phenology. The spectral libraries included spectra from the peak summer whereas the HySpex images were collected in early fall. Although more research is needed to fully compare field and HySpex spectra for wetland characterization, the radiometric correction applied to HySpex imagery produced appropriate spectra for wetland land categories.

6.3. Image Classification: Results

Best classification performance for A and B flight areas occurred when using the maximum likelihood classifier with a Kappa agreement coefficient ranging from 0.94 to 0.96, followed by the SAM classifier with a Kappa value ranging from 0.6 to 0.64 and, lastly, by the hybrid classifier, showing a lower performance with a Kappa value of ranging from 0.46 to 0.56 (Figures 9 and 10 and, Tables 3 and 4 and Appendix B Tables A1–A4). Moreover, the producer and user accuracy were also higher for the maximum likelihood classifier, with average values from 72% to 100%. SAM classification relies on external datasets, therefore, the lack of field spectra or a specific spectral library for high latitude areas for different types of vegetation, water and soils in this study may have caused this method to perform poorly. The training areas were used in place of a spectral reference dataset. However, this approach did not yield better results than the maximum likelihood classification. Pixel-based classifiers, such as SAM, perform best for hyperspectral imagery when extracting individual spectra for different plant species [43], which we lacked. Moreover, when these training areas were used with the SAM classifier, more than 35% of the final image was not classified due to the lack of proper spectra. To assign non-classified pixels to categories, a selective mode filter with a 3 by 3 convolution matrix was used. This removed non-classified pixels and increased image accuracy, but the performance was still poorer than the maximum likelihood method. In [44], the authors similarly found that maximum likelihood classification achieved a higher accuracy than SAM based on image-derived

spectral endmembers utilizing airborne hyperspectral CASI data in an inland wetland complex near the Grand River, a tributary of Lake Erie in Ontario, Canada. This study also grouped different wetland vegetation species into broader land cover classes.
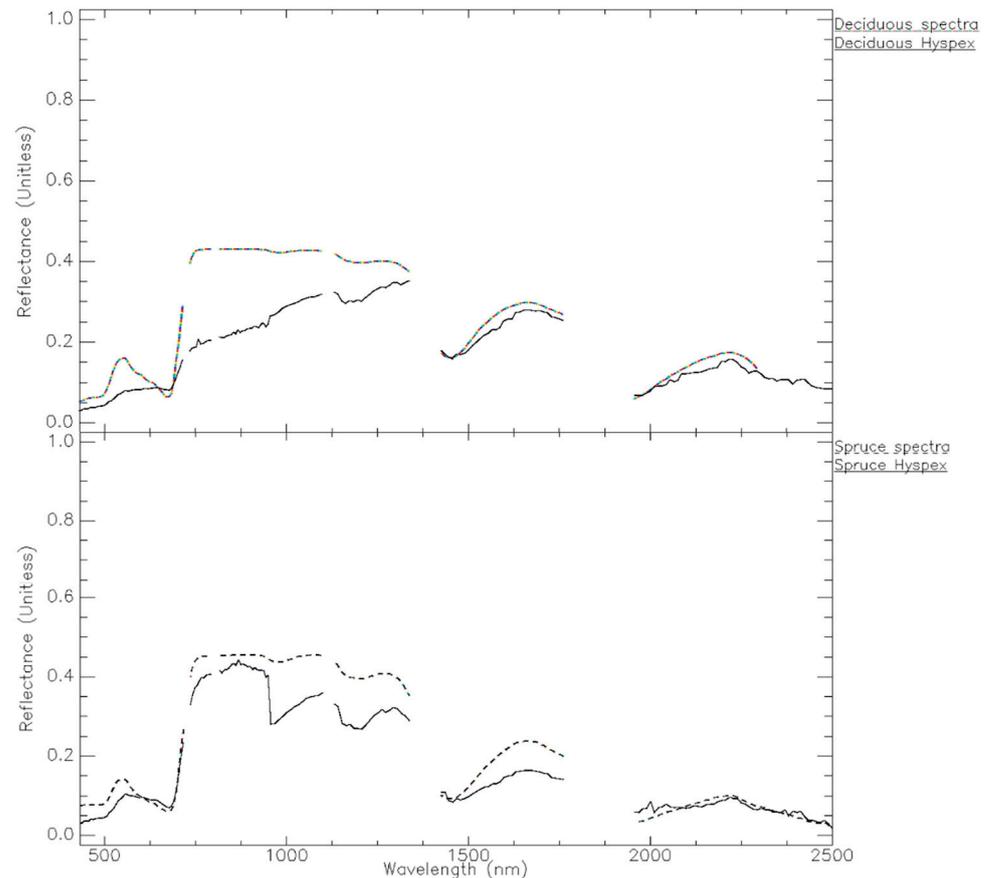
Figure 8. Comparison of HySpex spectra (solid line) and spectral library data (dashed line) for deciduous (upper panel) and spruce (lower panel) categories. Atmospheric water vapor and noisy bands were removed from the spectral library data for a good visual comparison with the HySpex spectra. Deciduous spectral library data were only available up to 2300 nm.

Hybrid classification showed the lowest performance, with an averaged Kappa value of 0.5 for both flight areas. The misclassification between equisetum, spruce (in area A), bog (in area B) and deciduous classes was the main reason for this lower performance. Finally, it is interesting to note that the low cirrus cloud that affected the area B bottom right section caused misclassification in the hybrid and maximum likelihood classifications, but it did not impact the final map for SAM classification (Figure 9). An advantage of SAM classification is that because it is dependent on the angle between two vectors in n-dimensional space, but not the vector magnitude, it is insensitive to varying magnitudes of illumination for the same cover types [40,45].

For all classifications, there was confusion between spruce and equisetum categories in areas with emergent vegetation, though this confusion was the least in maximum likelihood classifications. An increase in the training dataset on these areas a posteriori contributed to the better results using maximum likelihood. Expanded training datasets were also used in both hybrid and SAM classification, however the results did not improve. The mixing of plant and water signals in areas of flooded emergent vegetation results in a decrease in total reflected radiation, and the intensity of this effect is dependent on vegetation density, water depth, and canopy structure [46]. Due to the wide range of reflectance values represented in areas of emergent vegetation, the spectral signature of this class may overlap the signals from water, terrestrial vegetation, and soil. The lowered

reflectance values of the equisetum class in some areas were similar to the spruce class spectral signature, causing misclassification between the two when using spectrally based methods. Similar misclassifications between conifer species and emergent vegetation when using a SAM on hyperspectral imagery were also described by [43] and in other wetland studies [22,44,46]. To solve this issue, a buffer approximating 100 m along the lakes could be applied to mask out spruce pixels within these areas and reclassify them into equisetum category. If available, a LIDAR flight or an optical camera used to derive a digital surface model may also help to discriminate between both categories, as spruce trees are taller than equisetum.
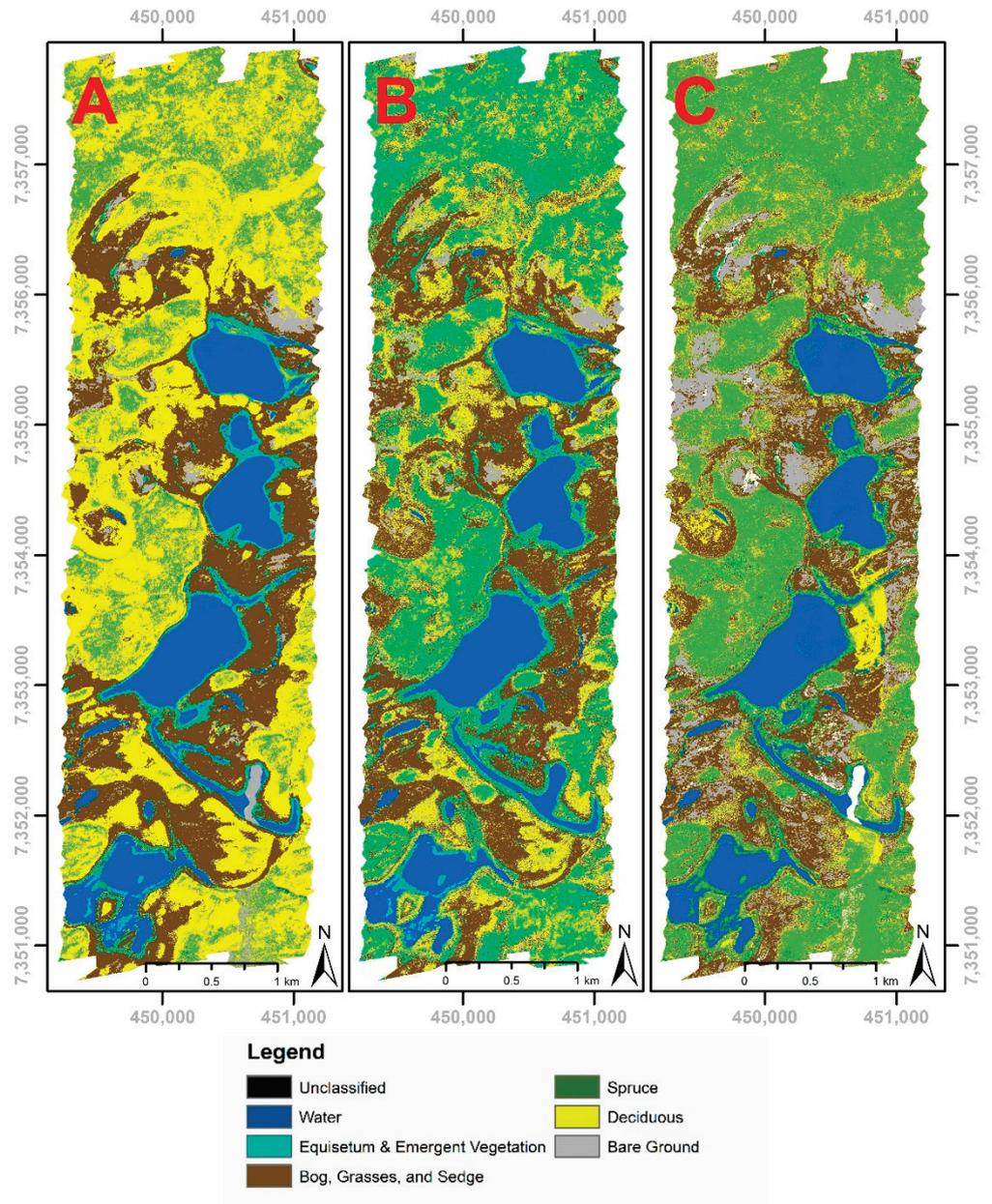
Figure 9. Wetland classification results for A area using a maximum likelihood classifier (right panel **A**), a Spectral Angle Mapper (SAM) classifier (middle panel **B**), and a hybrid classifier (left panel **C**).
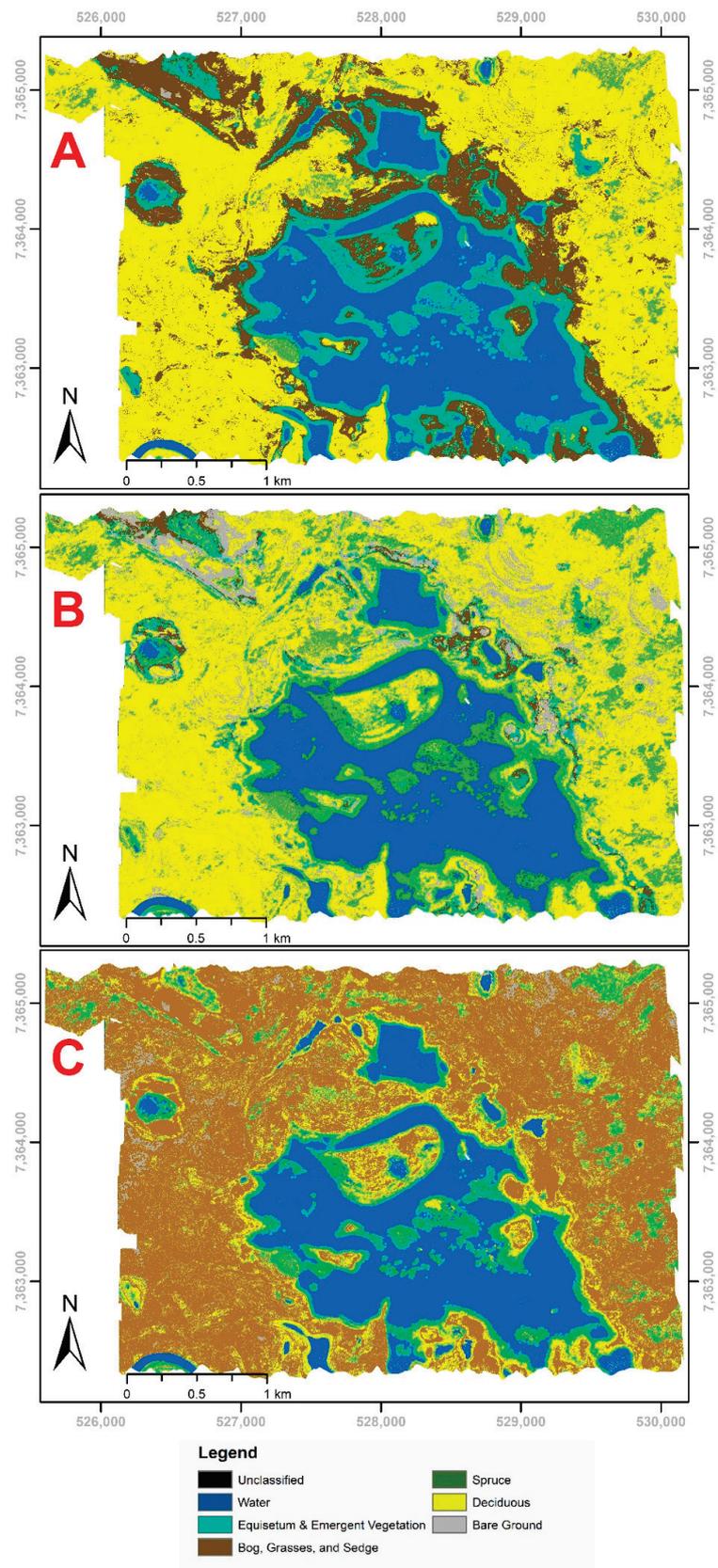
Figure 10. Wetland classification results for B area using a maximum likelihood classifier (top panel **A**), a Spectral Angle Mapper (SAM) classifier (middle panel **B**), and a hybrid classifier (bottom panel **C**).

Table 3. Confusion matrix for area A maximum likelihood classification. Results in %.

		Wetlands Map to Test							Commission error	User's accuracy
		a	b	c	d	e	f	Total		
Evaluation dataset	Water (a)	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0
	Equisetum (b)	0.0	96.0	0.0	4.0	0.0	0.0	100.0	4.0	96.0
	Bog (c)	0.0	0.0	97.8	0.3	0.3	1.7	100.0	2.2	97.8
	Spruce (d)	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0
	Deciduous (e)	0.0	0.0	5.5	6.1	88.3	0.0	100.0	11.7	88.3
	Bare ground (f)	0.0	0.0	55.6	0.0	0.0	44.4	100.0	55.6	44.4
	Total	100.0	100.0	100.0	100.0	100.0	100.0			
	Omission error	0.0	0.0	8.4	14.0	0.7	27.3		Kappa: 0.94	
	Producer's accuracy	100.0	100.0	91.6	86.0	99.3	72.7			

Table 4. Confusion matrix for area B maximum likelihood classification. Results in %.

		Wetlands Map to Test							Commission error	User's accuracy
		a	b	c	d	e	f	Total		
Evaluation dataset	Water (a)	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0
	Equisetum (b)	100.0	0.0	0.0	0.0	0.0	0.0	100.0	4.3	95.7
	Bog (c)	0.0	95.7	4.3	0.0	0.0	0.0	100.0	1.1	98.9
	Spruce (d)	0.0	0.0	98.9	0.0	0.0	1.1	100.0	0.0	100.0
	Deciduous (e)	0.0	0.0	0.0	100.0	0.0	0.0	100.0	7.9	92.1
	Bare ground (f)	0.0	0.0	0.0	6.8	92.1	1.1	100.0	0.0	100.0
	Total	100.0	100.0	100.0	100.0	100.0	100.0			
	Omission error	0.6	0.0	3.1	13.7	0.0	21.4		Kappa: 0.96	
	Producer's accuracy	99.4	100.0	96.9	86.3	100.0	78.6			

Timing image acquisition during periods of maximum spectral separability between plant communities or species of interest may further enhance classification efforts. In [47], the authors found that spectral reflectance differences were statistically significant between stands of hydrophytes, with maximum separation of species occurring during flowering or early seed stages. This, however, requires a detailed characterization of spectral responses over time between land cover types, which currently does not exist for Alaska in published spectral libraries.

Deciduous and spruce classes also showed some confusion. Deciduous and spruce usually form mixed forest, which is not easy to classify in this landscape. Moreover, due to the low angle of the sun above the horizon, the resulting imagery has self-cast shadows projected in the same vegetation that could lead to misclassification between both categories. Spruce and deciduous classes were notably misinterpreted as equisetum in these shadowed forested areas.

Therefore, resampling the whole imagery at a 5 m spatial resolution could be useful to increase the accuracy for both categories by integrating the shadowed regions within a pixel and averaging the spectral response between shadowed and non-shadowed areas of the same class. The addition of a shadow class could also be useful here, especially for any spatial statistical analyses that end users might perform on the thematic maps for land management or other purposes.

Bare ground category showed an intermediate accuracy in area A but a low accuracy in area B and had the least area representativeness of all other categories. Thus, it was not an easy category to classify and needs to be better defined. Bare ground showed spectra similar to dead vegetation, and the spectral endmember of this category may not be static.

Finally, water, bog, and equisetum categories yielded a high averaged accuracy of more than 90% in the user and producer accuracy for the maximum likelihood classifier. These categories each covered broad land cover types and incorporated different vegetation types with similar spectral, spatial, and canopy characteristics. The further segmentation of these classes into narrower categories would require more in situ knowledge of and data for the imaged areas, as well as detailed spectral characterization, and possibly alternate classification techniques to accurately identify and assign narrower classes.

7. Conclusions

In this study, the viability of airborne hyperspectral imaging for wetland mapping in the high latitudes of Alaska is demonstrated. High-resolution, orthorectified imagery with a good radiometry was produced for selected areas where only low-resolution and decades-old imagery existed. The classified vegetation maps derived from airborne hyperspectral images are an important contribution that help to further the understanding of how vegetation is responding to the rapidly changing climate in Interior Alaska.

A hyperspectral imaging system (HySpex) was configured in a Found Bush Hawk aircraft and used to acquire data over selected sites in the Yukon Flats. Custom designed mounting provided the flexibility needed for the system to be installed in different aircraft and ensured fast setup and break down before and after each flight, freeing the aircraft for other flights. This flexible yet robust mounting system helped with reducing commissioning costs and ensures the long-term viability of airborne data acquisition. Data were acquired with optimized flight configuration over three study areas that together represented the variety of wetland vegetation cover and water chemistry.

The hyperspectral data processing chain developed by the German Aerospace Center (DLR) was adapted for high latitudes by including a step for BRDF removal. Boresight calibration helped in geometrically correcting both the VNIR and SWIR images with an RMSE of less than 0.5 m at a 1 m spatial resolution. Data were successfully corrected radiometrically using ATCOR4 and ancillary water vapor data from the MODIS water vapor product. Although field spectra were not collected during image acquisition due to logistic constrains, visual comparison with available spectral libraries suggested a good radiometric correction.

For wetlands mapping, a 6-category legend was established based on previous USGS and USFWS information and previously available maps. Three different classification methods were applied in the two targeted areas using a spectral subset of selected 120 bands: hybrid classification, spectral angle mapper, and maximum likelihood. Final wetland maps were successfully classified using a maximum likelihood method with Kappa values higher than 0.94, and the average user and producer accuracy more than 90% for almost all categories. The best classification performance occurred when using the maximum likelihood classifier, followed by the SAM classifier and lastly the hybrid classifier, which showed a lower performance, with averaged Kappa indices of around 0.95, 0.62, and 0.5, respectively, for the two study areas. Although the SAM methodology is specifically suited for hyperspectral mapping, the lack of field spectra hampered the final outcome. It is important to note that the spruce and equisetum spectra in emergent areas were quite similar due to the decrease in reflectance caused by the integration of water in pixels covering areas of emergent vegetation. This led to the misclassification of pixels, especially when using the SAM and hybrid classifiers. Misclassification also occurred between spruce and deciduous categories, although it was minor.

In order to improve these results, future work should focus on the integration of LIDAR data or a digital surface model derived from standard RGB data to distinguish between equisetum and spruce categories. In addition, building a spectral library for Alaska wetlands vegetation will improve classification efforts and allow for the application of other hyperspectral classification techniques.

Appendix A. Hyperspectral Data Processing Workflow

The hyperspectral data processing workflow, adapted from the DLR data processing workflow (Habermeyer et al., 2012), consists of three steps: raw image processing and image orthorectification (Figure A1), radiometric and BRDF corrections (Figure A2), and mosaicking and spectral binning (Figure A3). The specific methods used in each step are described in Section 4. "Hyperspectral data processing".

Step 1: Raw image processing and image orthorectification: For quality control general sensor characteristics (e.g., spectral smile), sensor calibration and performance issues (e.g., striping, data drops), and external conditions during overflight (e.g., cloud cover) are assessed. Using the HySpex RAD module (software provided by the manufacturer), the laboratory derived calibration coefficients to convert brightness values (DN values) to at-sensor radiance are applied. The IMU/GPS data are then input in the HySpex NAV

module. Later, images are orthorectified in the PARGE software using the NAV data, the HySpex sensor model, and a digital elevation model (DEM). Additionally, offset angles from a boresight calibration are applied. The outputs from Step 1 are VNIR and SWIR orthorectified images, scan angle files, and DEMs stacked into a hyperspectral supercube (see Figure A1).
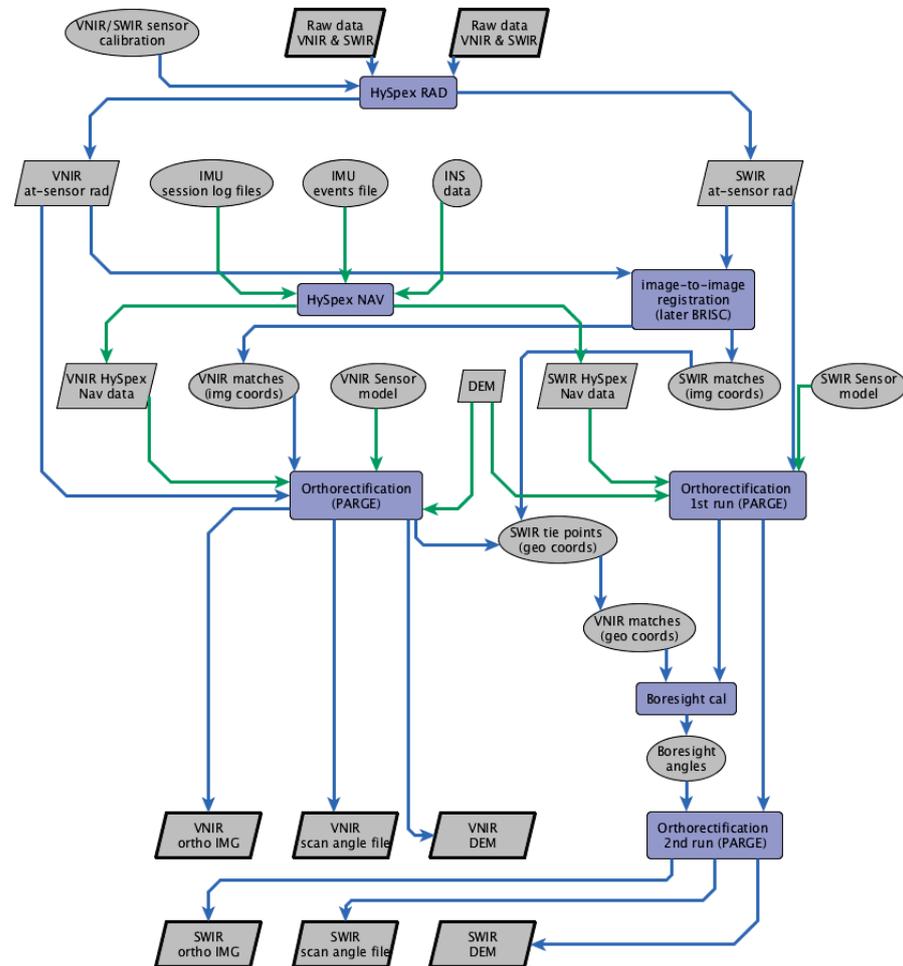
Figure A1. Raw image processing and image orthorectification workflow.

Step 2: Complete radiometric corrections including atmospheric, topographic, and BRDF corrections: This step brings together several input parameters measured or derived in the Step 1 into the SPECTRA module (Figure A2). Atmospheric correction is performed using a radiative transfer-based approach with ATCOR software to convert at-sensor radiance to surface reflectance values. Reliable atmospheric correction of the hyperspectral data requires a DEM and the robust parameterization of atmospheric column properties including atmospheric gases (water vapor and oxygen) and aerosol optical thickness (AOT). The DEM products are also used to apply BRDF corrections at this stage. Similarly, the user can provide atmospheric parameters based on atmospheric profiles of the study site (if available), or the processing chain will use a modeled standard atmosphere for the specific geographic region. The output from this step is a data cube corrected for geometric, atmospheric, and BRDF effects.

Step 3: Mosaicking and spectral binning: After the radiometric correction step, the individual corrected flight lines are mosaicked. Finally, to further increase the signal to noise ratio, spectral binning is applied (Figure A3).
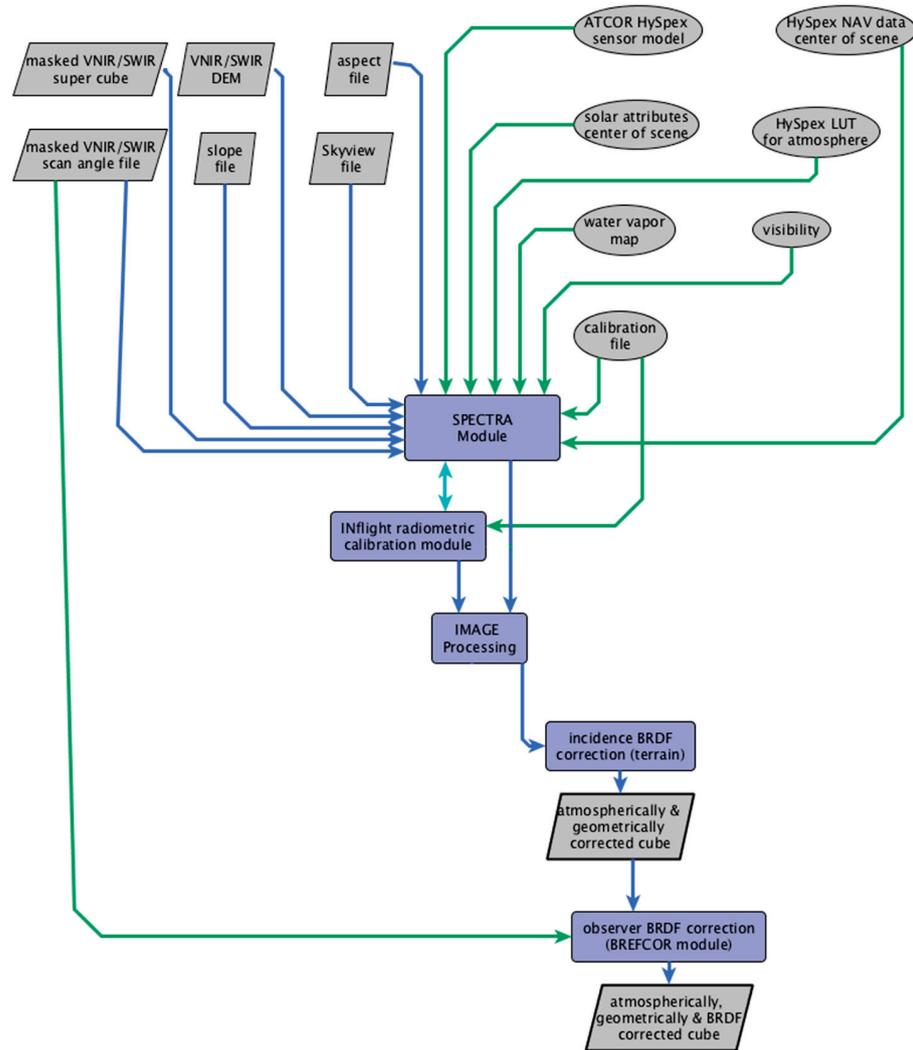
Figure A2. Radiometric correction workflow.
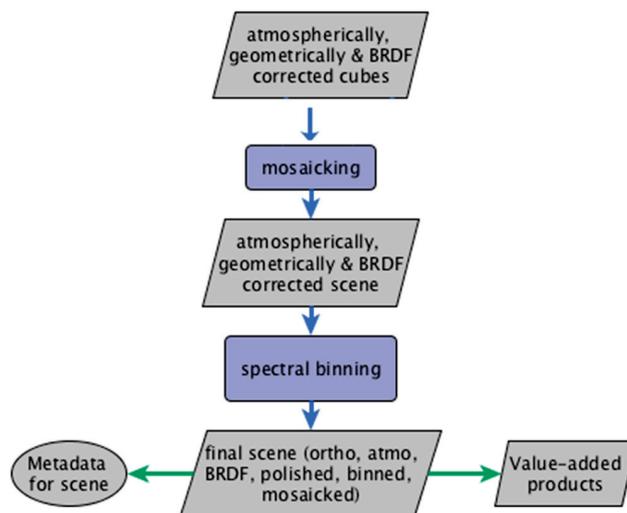
Figure A3. Mosaicking and spectral binning workflow.

Appendix B.

Classification results for the hybrid and SAM classifiers for both A and B flight areas.

Table A1. Confusion matrix for area A hybrid classification. Results in %.

		Wetlands Map to Test							
		a	b	c	d	e	f	Total	Commission error	User's accuracy
Evaluation dataset	Water (a)	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0
	Equisetum (b)	0.0	48.9	0.0	12.2	38.9	0.0	100.0	51.1	48.9
	Bog (c)	0.0	5.8	66.2	0.5	26.1	1.4	100.0	33.8	66.2
	Spruce (d)	0.0	30.7	18.4	44.7	6.1	0.0	100.0	55.3	44.7
	Deciduous (e)	0.0	16.5	45.6	2.5	35.4	0.0	100.0	64.6	35.4
	Bare ground (f)	0.0	0.0	82.9	0.0	0.6	16.6	100.0	83.4	16.6
	Total	100.0	100.0	100.0	100.0	100.0	100.0			
	Omission error	0.0	73.8	65.6	9.7	80.7	9.4		Kappa: 0.46	
	Producer's accuracy	100.0	26.2	34.4	90.3	19.3	90.6			

Table A2. Confusion matrix for area B hybrid classification. Results in %.

		Wetlands Map to Test							
		a	b	c	d	e	f	Total	Commission error	User's accuracy
Evaluation dataset	Water (a)	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0
	Equisetum (b)	0.0	53.4	0.0	5.4	41.2	0.0	100.0	46.6	53.4
	Bog (c)	0.0	0.0	85.9	0.0	7.7	6.4	100.0	14.1	85.9
	Spruce (d)	0.0	0.0	0.0	79.5	20.5	0.0	100.0	20.5	79.5
	Deciduous (e)	0.0	9.1	17.8	1.7	71.4	0.0	100.0	28.6	71.4
	Bare ground (f)	0.0	0.0	88.1	0.0	0.0	11.9	100.0	88.1	11.9
	Total	100.0	100.0	100.0	100.0	100.0	100.0			
	Omission error	0.6	8.2	82.6	12.8	59.0	11.9		Kappa: 0.56	
	Producer's accuracy	99.4	91.8	17.4	87.2	41.1	88.1			

Table A3. Confusion matrix for area A SAM classification. Results in %.

		Wetlands Map to Test							
		a	b	c	d	e	f	Total	Commission error	User's accuracy
Evaluation dataset	Water (a)	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0
	Equisetum (b)	0.0	40.5	16.9	24.9	17.6	0.0	100.0	59.5	40.5
	Bog (c)	0.0	2.0	90.7	1.7	5.6	0.0	100.0	9.3	90.7
	Spruce (d)	0.0	36.1	0.0	57.4	6.5	0.0	100.0	42.6	57.4
	Deciduous (e)	0.0	1.4	0.0	4.3	94.2	0.0	100.0	5.8	94.2
	Bare ground (f)	0.0	0.0	67.9	0.0	2.8	29.4	100.0	70.6	29.4
	Total	100.0	100.0	100.0	100.0	100.0	100.0			
	Omission error	0.0	27.4	31.4	57.2	55.2	0.0		Kappa: 0.64	
	Producer's accuracy	100.0	72.6	68.6	42.8	44.8	100.0			

Table A4. Confusion matrix for area B SAM classification. Results in %.

		Wetlands Map to Test							
		a	b	c	d	e	f	Total	Commission error	User's accuracy
Evaluation dataset	Water (a)	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0
	Equisetum (b)	0.0	48.8	0.0	18.8	32.4	0.0	100.0	59.5	40.5
	Bog (c)	0.0	0.0	73.5	3.1	19.7	3.7	100.0	9.3	90.7
	Spruce (d)	0.0	1.6	0.0	63.4	35.0	0.0	100.0	42.6	57.4
	Deciduous (e)	0.0	44.1	7.3	3.2	45.1	0.3	100.0	5.8	94.2
	Bare ground (f)	0.0	0.0	0.0	0.0	0.0	100.0	100.0	70.6	29.4
	Total	100.0	100.0	100.0	100.0	100.0	100.0			
	Omission error	0.0	27.4	31.4	57.2	55.2	0.0		Kappa: 0.64	
	Producer's accuracy	100.0	72.6	68.6	42.8	44.8	100.0			

References

1. Flagstad, L.; Steer, M.A.; Boucher, T.; Aisu, M.; Lema, P. *Wetlands across Alaska: Statewide Wetland Map and Assessment of Rare Wetland Ecosystems*; University of Alaska Anchorage: Anchorage, AK, USA, 2018; p. 151.
2. Dahl, T.E. *Wetlands Losses in the United States 1780s to 1980s*; U.S. Department of the Interior, Fish and Wildlife Service: Washington, DC, USA, 1990.
3. Chen, M.; Rowland, J.C.; Wilson, C.J.; Altmann, G.L.; Brumby, S.P. Temporal and spatial pattern of thermokarst lake area changes at yukon flats, alaska. *Hydrol. Process.* **2014**, *28*, 837–852. [CrossRef]
4. Jorgenson, M.T.; Racine, C.H.; Walters, J.C.; Osterkamp, T.E. Permafrost degradation and ecological changes associated with a warming climate in central alaska. *Clim. Chang.* **2001**, *48*, 551–579. [CrossRef]
5. Roach, J.K.; Griffith, B.; Verbyla, D. Landscape influences on climate-related lake shrinkage at high latitudes. *Glob. Chang. Biol.* **2013**, *19*, 2276–2284. [CrossRef] [PubMed]
6. Haynes, K.M.; Connon, R.F.; Quinton, W.L. Permafrost thaw induced drying of wetlands at scotty creek, nwt, canada. *Environ. Res. Lett.* **2018**, *13*, 114001. [CrossRef]
7. Carter, V. Wetland hydrology, water quality, and associated functions. In *United States Geological Survey Water Supply Paper*; United States Geological Survey: Reston, VA, USA, 1996; Volume 2425, pp. 35–48.
8. Hall, J.V.; Frayer, W.E.; Wilen, B.O.; Fish, U.S. Status of Alaska Wetlands. U.S. Fish & Wildlife Service: Alaska Region, AK, USA; Anchorage, AK, USA, 1994.
9. Niemi, G.J.; McDonald, M.E. Application of ecological indicators. *Annu. Rev. Ecol. Evol. Syst.* **2004**, *35*, 89–111. [CrossRef]
10. Tiner, R.W. *Defining Hydrophytes for Wetland Identification and Delineation*; Defense Technical Information Center: Fort Belvoir, VA, USA, 2012.
11. Cowardin, L.M.; Fish, U.S.; Service, W.; Program, B.S. *Classification of Wetlands and Deepwater Habitats of the United States*; Fish and Wildlife Service, U.S. Department of the Interior: Washington, DC, USA, 1979.
12. Finlayson, C.M.; van der Valk, A.G. Wetland classification and inventory: A summary. *Vegetatio* **1995**, *118*, 185–192. [CrossRef]
13. Adam, E.; Mutanga, O.; Rugege, D. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review. *Wetl. Ecol. Manag.* **2010**, *18*, 281–296. [CrossRef]
14. Group, A.W.T.W. *Statewide Wetland Inventory Ten Year Strategic Plan 2019–2029*; U.S. Fish & Wildlife Service: Anchorage, AK, USA, 2019; p. 22.
15. Jorgenson, M.T. *Landcover Mapping for Bering Land Bridge National Preserve and Cape Krusenstern National Monument, Northwestern Alaska*; U.S. Department of the Interior, National Park Service, Natural Resource Program Center: Washington, DC, USA, 2004.
16. Pastick, N.J.; Jorgenson, M.T.; Wylie, B.K.; Minsley, B.J.; Ji, L.; Walvoord, M.A.; Smith, B.D.; Abraham, J.D.; Rose, J.R. Extending airborne electromagnetic surveys for regional active layer and permafrost mapping with remote sensing and ancillary data, yukon flats ecoregion, central alaska. *Permafr. Periglac.* **2013**, *24*, 184–199. [CrossRef]
17. Klemas, V. Remote sensing of wetlands: Case studies comparing practical techniques. *J. Coast. Res.* **2011**, *27*, 418–427.
18. Cristóbal, J.; Graham, P.; Buchhorn, M.; Prakash, A. A new integrated high-latitude thermal laboratory for the characterization of land surface processes in alaska's arctic and boreal regions. *Data* **2016**, *1*, 13. [CrossRef]
19. Miller, C.E.; Green, R.O.; Thompson, D.R.; Thorpe, A.K.; Eastwood, M.; Mccubbin, I.B.; Olson-duvall, W.; Bernas, M.; Sarture, C.M.; Nolte, S.; et al. *Above: Hyperspectral Imagery from Aviris-Ng, Alaskan and Canadian Arctic, 2017–2018*; ORNL DAAC: Oak Ridge, TN, USA, 2019.
20. Petropoulos, G.P.; Manevski, K.; Carlson, T. Hyperspectral remote sensing with emphasis on land cover mapping: From ground to satellite observations. In *Scale Issues in Remote Sensing*; Weng, Q., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014; pp. 289–324.
21. Mahdavi, S.; Salehi, B.; Granger, J.; Amani, M.; Brisco, B.; Huang, W. Remote sensing for wetland classification: A comprehensive review. *Gisci. Remote Sens.* **2018**, *55*, 623–658. [CrossRef]
22. Ozesmi, S.L.; Bauer, M.E. Satellite remote sensing of wetlands. *Wetl. Ecol. Manag.* **2002**, *10*, 381–402. [CrossRef]
23. Kaufmann, H.; Förster, S.; Wulf, H.; Segl, K.; Guanter, L.; Bochow, M.; Heiden, U.; Müller, A.; Heldens, W.; Schneiderhan, T. *Science Plan of the Environmental Mapping and Analysis Program (Enmap)*; Deutsches GeoForschungsZentrum GFZ: Postdam, Germany, 2012.
24. Heglund, P.J.; Jones, J.R. Limnology of shallow lakes in the yukon flats national wildlife refuge, interior alaska. *Lake Reserv. Manag.* **2003**, *19*, 133–140. [CrossRef]
25. Arp, C.D.; Jones, B.M. *Geography of Alaska Lake Districts: Identification, Description, and Analysis of Lake-Rich Regions of a Diverse and Dynamic State*; 2008-5215; U. S. Geological Survey: Reston, VA, USA, 2009.
26. Shur, Y.; Kanevskiy, M.; Jorgenson, M.; Dillon, M.; Stephani, E.; Bray, M.; Fortier, D. Permafrost degradation and thaw settlement under lakes in yedoma environment. In *Proceedings of the Tenth International Conference on Permafrost*; International Contributions; Hinkel, K.E., Ed.; The Northern Publisher: Salekhard, Russia, 2012; Volume 10, p. 383.
27. Ford, J.; Bedford, B.L. The hydrology of alaskan wetlands, U.S.A.: A review. *Arct. Alp. Res.* **1987**, *19*, 209–229. [CrossRef]
28. Lewis, T.L.; Lindberg, M.S.; Schmutz, J.A.; Heglund, P.J.; Rover, J.; Koch, J.C.; Bertram, M.R. Pronounced chemical response of subarctic lakes to climate-driven losses in surface area. *Glob. Chang. Biol.* **2015**, *21*, 1140–1152. [CrossRef]
29. Heglund, P.J. *Patterns of Wetland Use among Aquatic Birds in the Interior Boreal Forest Region of Alaska*; University of Missouri: Columbia, MI, USA, 1994.

30. Habermeyer, M.; Bachmann, M.; Holzwarth, S.; Müller, R.; Richter, R. Incorporating a Push-Broom Scanner into a Generic Hyperspectral Processing Chain. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 5293–5296.

31. Norsk Elektro Optikk. *Imaging Spectrometer: Users Manual*; Norsk Elektro Optikk: Oslo, Norway, 2014.

32. Richter, R.; Schlapfer, D. *Atmospheric/Topographic Correction for Airborne Imagery*; ReSe Applications: Wil, Switzerland, 2015.

33. Davaadorj, A. *Evaluating Atmospheric Correction Methods Using Worldview-3 Image*; University of Twente: Enshede, The Netherlands, 2019.

34. Cristóbal, J.; Jiménez-Muñoz, J.; Prakash, A.; Mattar, C.; Skoković, D.; Sobrino, J. An improved single-channel method to retrieve land surface temperature from the landsat-8 thermal band. *Remote Sens.* **2018**, *10*, 431. [CrossRef]

35. Buchhorn, M. Ground-Based Hyperspectral and Spectro-Directional Reflectance Characterization of Arctic Tundra Vegetation Communities: Field Spectroscopy and Field Spectro-Goniometry of Siberian and Alaskan Tundra in Preparation of the Enmap Satellite Mission. Doctoral Dissertation, Universitaetsverlag Potsdam, Potsdam, Germany, 2014.

36. Schläpfer, D.; Richter, R. Evaluation of Brefcor Brdf Effects Correction for Hyspex, Casi, and Apex Imaging Spectroscopy Data. In Proceedings of the 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lausanne, Switzerland, 24–27 June 2014; pp. 1–4.

37. Boggs, K.; Flagstad, L.; Boucher, T.; Kuo, T.; Fehringer, D.; Guyer, S.; Megumi, A. *Vegetation Map and Classification: Northern, Western and Interior Alaska*, 2nd ed.; University of Alaska: Anchorage, AK, USA, 2016; p. 110.

38. Pu, R. *Hyperspectral remote sensing: Fundamentals and Practices*; CRC Press: Boca Raton, FL, USA, 2017; p. 466.

39. Serra, P.; Pons, X.; Saurí, D. Post-classification change detection with data from different sensors: Some accuracy considerations. *Int. J. Remote. Sens.* **2003**, *24*, 3311–3340. [CrossRef]

40. Kruse, F.A.; Lefkoff, A.B.; Boardman, J.W.; Heidebrecht, K.B.; Shapiro, A.T.; Barloon, P.J.; Goetz, A.F.H. The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data. *Remote. Sens. Environ.* **1993**, *44*, 145–163. [CrossRef]

41. Markelin, L.; Simis, S.; Hunter, P.; Spyrakos, E.; Tyler, A.; Clewley, D.; Groom, S. Atmospheric correction performance of hyperspectral airborne imagery over a small eutrophic lake under changing cloud cover. *Remote Sens.* **2016**, *9*, 2. [CrossRef]

42. Hovi, A.; Raitio, P.; Rautiainen, M. A spectral analysis of 25 boreal tree species. *Silva. Fenn.* **2017**, *51*, 4. [CrossRef]

43. Harken, J.; Sugumaran, R. Classification of iowa wetlands using an airborne hyperspectral image: A comparison of the spectral angle mapper classifier and an object-oriented approach. *Can. J. Remote Sens.* **2005**, *31*, 167–174. [CrossRef]

44. Jollineau, M.Y.; Howarth, P.J. Mapping an inland wetland complex using hyperspectral imagery. *Int. J. Remote Sens.* **2008**, *29*, 3609–3631. [CrossRef]

45. Leckie, D.G.; Cloney, E.; Jay, C.; Paradine, D. Automated mapping of stream features with high-resolution multispectral imagery: An example of the capabilities. *Photogramm. Eng. Remote Sens.* **2005**, *71*, 11. [CrossRef]

46. Silva, T.S.F.; Costa, M.P.F.; Melack, J.M.; Novo, E.M.L.M. Remote sensing of aquatic vegetation: Theory and applications. *Environ. Monit. Assess.* **2008**, *140*, 131–145. [CrossRef]

47. Best, R.G.; Wehde, M.E.; Linder, R.L. Spectral reflectance of hydrophytes. *Remote Sens. Environ.* **1981**, *11*, 27–35. [CrossRef]

MDPI