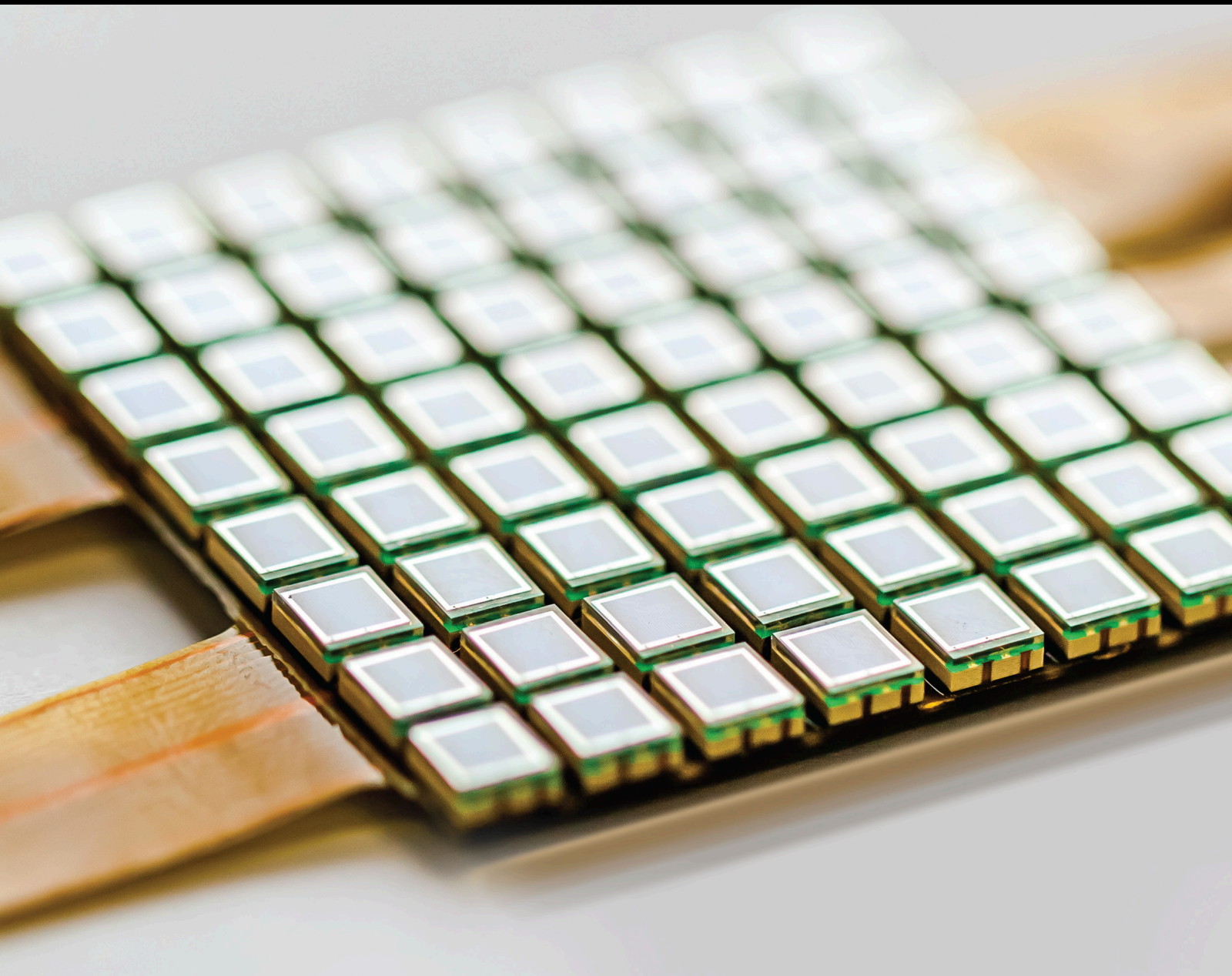


Advanced Sensor Technologies in Geospatial Sciences and Engineering 2020

Lead Guest Editor: Hyung-Sup Jung

Guest Editors: Sang-Hoon Hong and Lei Zhang





**Advanced Sensor Technologies in Geospatial
Sciences and Engineering 2020**

Journal of Sensors

**Advanced Sensor Technologies in
Geospatial Sciences and Engineering
2020**

Lead Guest Editor: Hyung-Sup Jung

Guest Editors: Sang-Hoon Hong and Lei Zhang






Copyright © 2023 Hindawi Limited. All rights reserved.

This is a special issue published in "Journal of Sensors." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chief Editor

Harith Ahmad , Malaysia

Associate Editors

Duo Lin , China
Fanli Meng , China
Pietro Siciliano , Italy
Guiyun Tian, United Kingdom

Academic Editors

Ghufran Ahmed , Pakistan
Manuel Alexandre , Spain
Bruno Andò , Italy
Constantin Apetrei, Romania
Fernando Benito-Lopez , Spain
Romeo Bernini , Italy
Shekhar Bhansali, USA
Matthew Brodie, Australia
Ravikumar CV, India
Belén Calvo, Spain
Stefania Campopiano , Italy
Binghua Cao , China
Domenico Caputo, Italy
Sara Casciati, Italy
Gabriele Cazzulani , Italy
Chi Chiu Chan, Singapore
Sushank Chaudhary, Thailand
Edmon Chehura, United Kingdom
Marvin H Cheng , USA
Lei Chu , USA
Mario Collotta , Italy
Marco Consales , Italy
Jesus Corres , Spain
Andrea Cusano, Italy
Egidio De Benedetto , Italy
Luca De Stefano , Italy
Manel Del Valle , Spain
Franz L. Dickert, Austria
Giovanni Diraco, Italy
Maria de Fátima Domingues , Portugal
Nicola Donato , Italy
Sheng Du , China
Mauro Epifani , Italy
Congbin Fan, China
Lihang Feng, China
Vittorio Ferrari , Italy
Luca Francioso, Italy

Bin Gao, China
Libo Gao , China
Carmine Granata , Italy
Pramod Kumar Gupta, USA
Mohammad Haider , USA
Agustin Herrera-May , Mexico
María del Carmen Horrillo, Spain
Evangelos Hristoforou , Greece
Shahid Hussain, China
Grazia Iadarola, Italy
Syed K. Islam , USA
Stephen James , United Kingdom
Sana Ullah Jan, United Kingdom
Bruno C. Janegitz , Brazil
Hai-Feng Ji , USA
Shouyong Jiang, United Kingdom
Roshan Prakash Joseph, USA
Niravkumar Joshi, USA
Rajesh Kaluri, India
Sang Sub Kim , Republic of Korea
Dr. Rajkishor Kumar, India
Rahul Kumar, India
Nageswara Lalam, USA
Antonio Lazaro , Spain
Chengkuo Lee, Singapore
Chenzong Li , USA
Yuan Li , China
Yuxing Li , China
Zhi Lian, Australia
Rosalba Liguori , Italy
Sangsoon Lim, Republic of Korea
Huan Liu , China
Jin Liu , China
Eduard Llobet , Spain
Jaime Lloret , Spain
Mohamed Louzazni, Morocco
Jesús Lozano , Spain
Oleg Lupan , Moldova
Frederick Maily , France
Leandro Maio, Italy
Pawel Malinowski , Poland
Carlos Marques , Portugal
Eugenio Martinelli , Italy
Antonio Martinez-Olmos , Spain
Giuseppe Maruccio , Italy

Yasuko Y. Maruo, Japan
Zahid Mehmood , Pakistan
Carlos Michel , Mexico
Stephen. J. Mihailov , Canada
Ehsan Namaziandost , Iran
Heinz C. Neitzert , Italy
Sing Kiong Nguang , New Zealand
Calogero M. Oddo , Italy
Tinghui Ouyang, Japan
Marimuthu Palaniswami , Australia
Alberto J. Palma , Spain
Davide Palumbo , Italy
Roberto Paolesse , Italy
Akhilesh Pathak , Thailand
Giovanni Pau , Italy
Giorgio Pennazza , Italy
Michele Penza , Italy
Salvatore Pirozzi , Italy
Antonina Pirrotta , Italy
Sivakumar Poruran, India
Stelios Potirakis , Greece
Biswajeet Pradhan , Malaysia
Giuseppe Quero , Italy
Linesh Raja, India
Maheswar Rajagopal, India
Valerie Renaudin , France
Armando Ricciardi , Italy
Christos Riziotis , Greece
Maria Luz Rodriguez-Mendez , Spain
Jerome Rossignol , France
Ylias Sabri , Australia
Sourabh Sahu, India
José P. Santos , Spain
Sina Sareh, United Kingdom
Isabel Sayago , Spain
Andreas Schütze , Germany
Praveen K. Sekhar , USA
Sandra Sendra, Spain
Sandeep Sharma, India
Sunil Kumar Singh Singh , India
Vincenzo Spagnolo, Italy
Kathiravan Srinivasan , India
Sachin K. Srivastava , India
Grigore Stamatescu , Romania

Stefano Stassi , Italy
Danfeng Sun, China
Ashok Sundramoorthy, India
Salvatore Surdo , Italy
Yunchao Tang, China
Roshan Thotagamuge , Sri Lanka
Guiyun Tian , United Kingdom
Vijay Tomer , USA
Sri Ramulu Torati, USA
Abdellah Touhafi , Belgium
Hoang Vinh Tran , Vietnam
Aitor Urrutia , Spain
Hana Vaisocherova - Lisalova , Czech Republic
Everardo Vargas-Rodriguez , Mexico
Xavier Vilanova , Spain
Stanislav Vitek , Czech Republic
Luca Vollero , Italy
Tomasz Wandowski , Poland
He Wen , China
Qihao Weng, USA
Penghai Wu , China
Qiang Wu, United Kingdom
Yuedong Xie, China
Chen Yang , China
Jiachen Yang, China
Aijun Yin, China
Chouki Zerrouki, France

Contents

Hyung-Sup Jung , Sang-Hoon Hong , and Lei Zhang 



Editorial (2 pages), Article ID , ()

Advantage of Combining OBIA and Classifier Ensemble Method for Very High-Resolution Satellite Imagery Classification

Ruimei Han, Pei Liu , Guangyan Wang, Hanwei Zhang, and Xilong Wu

Research Article (15 pages), Article ID 8855509, Volume 2020 (2020)

On Time-Series InSAR by SA-SVR Algorithm: Prediction and Analysis of Mining Subsidence

Yun Shi , Qianwen Li, Xin Meng, Tongkang Zhang, and Jingjian Shi 





Research Article (17 pages), Article ID 8860225, Volume 2020 (2020)

Hybrid PSO and Evolutionary Game Theory Protocol for Clustering and Routing in Wireless Sensor Network

Yu Song , Zhigui Liu , and Xiaoli He

Research Article (20 pages), Article ID 8817815, Volume 2020 (2020)

Spatiotemporal Characterization of Land Subsidence in Guandu (China) Revealed by Multisensor InSAR Observations

Wu Zhu , Xue-qi Zhang , Zhan-ke Liu , and Qian Zhu 

Research Article (13 pages), Article ID 8855364, Volume 2020 (2020)

Modelling Reservoir Chlorophyll-*a*, TSS, and Turbidity Using Sentinel-2A MSI and Landsat-8 OLI Satellite Sensors with Empirical Multivariate Regression

Yashon O. Ouma , Kimutai Noor, and Kipkemoi Herbert

Research Article (21 pages), Article ID 8858408, Volume 2020 (2020)



Integrating Image and Network-Based Topological Data through Spatial Data Fusion for Indoor Location-Based Services

Dasol Ahn , Alexis Richard C. Claridades , and Jiyeong Lee 

Research Article (12 pages), Article ID 8877739, Volume 2020 (2020)


Characteristics of the Reanalysis and Satellite-Based Surface Net Radiation Data in the Arctic

Minji Seo , Hyun-Cheol Kim , Kyeong-Sang Lee , Noh-Hun Seong , Eunkyung Lee , Jinsoo

Kim , and Kyung-Soo Han 



Research Article (13 pages), Article ID 8825870, Volume 2020 (2020)

Land Cover Classification Using SegNet with Slope, Aspect, and Multidirectional Shaded Relief Images Derived from Digital Surface Model

Dae Geon Lee, Young Ha Shin, and Dong-Cheon Lee 


Research Article (21 pages), Article ID 8825509, Volume 2020 (2020)

Pre- and Postcollapse Ground Deformation Revealed by SAR Interferometry: A Case Study of Foshan (China) Ground Collapse

Wu Zhu , Yan Zhang, Zhanke Liu , and Qian Zhu


Research Article (17 pages), Article ID 8899054, Volume 2020 (2020)

Search Space Reduction for Determination of Earthquake Source Parameters Using PCA and k -Means Clustering

Seongjae Lee and Taehyoun Kim 


Research Article (12 pages), Article ID 8826634, Volume 2020 (2020)

Adaptive Magnetic Anomaly Detection Method with Ensemble Empirical Mode Decomposition and Minimum Entropy Feature

Liming Fan , Chong Kang, Huigang Wang, Hao Hu, and Mingliang Zou

Research Article (10 pages), Article ID 8856577, Volume 2020 (2020)

Measurement of Cotton Canopy Temperature Using Radiometric Thermal Sensor Mounted on the Unmanned Aerial Vehicle (UAV)

Anjin Chang, Jinha Jung, Murilo M. Maeda, Juan A. Landivar, Henrique D. R. Carvalho, and Junho Yeom 



Research Article (7 pages), Article ID 8899325, Volume 2020 (2020)

Prediction Model of Thermal Thawing Sensibility and Thaw Depth for Permafrost Embankment along the Qinghai-Tibet Engineering Corridor Using MODIS Data

Fuqing Cui , Jianbing Chen , Zhiyun Liu , Wu Zhu , Wei Wang , and Wei Zhang 


Research Article (12 pages), Article ID 8819476, Volume 2020 (2020)

Developing a Data Model of Indoor Points of Interest to Support Location-Based Services

Alexis Richard C. Claridades  and Jiyeong Lee 


Research Article (16 pages), Article ID 8885384, Volume 2020 (2020)

Priority Analysis of Remote Sensing and Geospatial Information Techniques to Water-Related Disaster Damage Reduction for Inter-Korean Cooperation

Sunmin Lee, Sung-Hwan Park, Mounng-Jin Lee , and Taejung Song


Review Article (10 pages), Article ID 8878888, Volume 2020 (2020)

An Improved Full-Aperture ScanSAR Imaging Method Integrating the MIAA Based Aperture Interpolation

Jiaqi Ning, Dacheng Liu, Kaiyu Liu, Heng Zhang , and Yingjie Wang

Research Article (15 pages), Article ID 8196081, Volume 2020 (2020)

RS and GIS Supported Urban LULC and UHI Change Simulation and Assessment

Pei Liu, Shoujun Jia, Ruimei Han , Yuanping Liu, Xiaofeng Lu, and Hanwei Zhang




Research Article (17 pages), Article ID 5863164, Volume 2020 (2020)

Spatiotemporal Fusion of Remote Sensing Image Based on Deep Learning

Xiaofei Wang  and Xiaoyi Wang

Research Article (11 pages), Article ID 8873079, Volume 2020 (2020)

Regional Patch Detection of Road Traffic Network

Xia Zhu , Weidong Song , and Lin Gao 

Research Article (6 pages), Article ID 6836091, Volume 2020 (2020)

Editorial

Advanced Sensor Technologies in Geospatial Sciences and Engineering 2020

Hyung-Sup Jung ¹, **Sang-Hoon Hong** ², and **Lei Zhang** ³

¹Department of Geoinformatics, The University of Seoul, Seoul, Republic of Korea

²Department of Geological Sciences, Pusan National University, Busan, Republic of Korea

³College of Surveying and Geo-Informatics, Tongji University, Shanghai, China

Correspondence should be addressed to Sang-Hoon Hong; geoshong@pusan.ac.kr and Lei Zhang; lei.zhang@polyu.edu.hk

Received 19 July 2021; Accepted 19 July 2021; Published 19 July 2023

Copyright © 2023 Hyung-Sup Jung et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Advanced sensor technologies in geospatial sciences and engineering research have contributed to (1) geographic mapping and (2) understanding the geological, ecological, hydrological, and environmental characteristics of Earth surfaces. Various studies on recent advances in sensor technology in geospatial sciences and engineering have been found in several special issues. In this special issue, a total of 19 original research articles addressing the advanced sensor technologies in the geospatial sciences and engineering have been published. The papers published in this special issue can be summarized as follows:

R. Han et al. in their paper entitled “Advantage of Combining OBIA and Classifier Ensemble Method for Very High-Resolution Satellite Imagery Classification” tested the feasibility of GF-1 and GF-2 very high-resolution imagery in urban land use and land cover classification using the random forest ensemble classifier with the object-based image analysis (OBIA) method.

Y. Shi et al. in their paper entitled “On Time-Series InSAR by SA-SVR Algorithm: Prediction and Analysis of Mining Subsidence” integrated the small baseline subset SAR interferometry (SBAS-InSAR) method into the simulated annealing support vector regression (SA-SVR) algorithm for mining subsidence prediction.

Y. Song et al. in their paper entitled “Hybrid PSO and Evolutionary Game Theory Protocol for Clustering and Routing in Wireless Sensor Network” proposed an efficient

method to optimize the selection and formation of clusters at the initial stage by using an improved protocol.

W. Zhu et al. in their paper entitled “Spatiotemporal Characterization of Land Subsidence in Guandu (China) Revealed by Multisensor InSAR Observations” investigated the time-series two-dimensional ground deformation in Guandu using the multisensor InSAR observation.

Y. O. Ouma et al. in their paper entitled “Modelling Reservoir Chlorophyll-*a*, TSS, and Turbidity Using Sentinel-2A MSI and Landsat-8 OLI Satellite Sensors with Empirical Multivariate Regression” demonstrated the effectiveness of Sentinel-2A/MSI (S2A) and Landsat-8/OLI (L8) satellite sensors for the estimation of Chl-*a*, turbidity, and TSS water quality parameters in a large reservoir (case 2 water body).

D. Ahn et al. in their paper entitled “Integrating Image and Network-Based Topological Data through Spatial Data Fusion for Indoor Location-Based Services” proposed a methodology for the data fusion of image and network-based topological data without undertaking data conversion, using a separate data model, or reference data.

M. Seo et al. in their paper entitled “Characteristics of the Reanalysis and Satellite-Based Surface Net Radiation Data in the Arctic” assessed the net radiation products over the Arctic using validation and intercomparison analyses.

D. G. Lee et al. in their paper entitled “Land Cover Classification Using SegNet with Slope, Aspect, and Multidirectional Shaded Relief Images Derived from Digital Surface

Model” proposed an efficient way to classify terrain features using a convolutional neural network- (CNN-) based SegNet model by utilizing 3D geospatial data including infrared (IR) orthoimages, digital surface model (DSM), and derived information.

W. Zhu et al. in their paper entitled “Pre- and Postcollapse Ground Deformation Revealed by SAR Interferometry: A Case Study of Foshan (China) Ground Collapse” investigated the ground deformation in the vicinity of the collapsed sinkhole in the Foshan collapse event using multitemporal SAR interferometry. A total of 55 C-band Sentinel-1A images were used for the study.

S. Lee and T. Kim in their paper entitled “Search Space Reduction for Determination of Earthquake Source Parameters Using PCA and k -Means Clustering” presented a new search space reduction algorithm based on machine learning techniques to determine the earthquake source parameters.

L. Fan et al. in their paper entitled “Adaptive Magnetic Anomaly Detection Method with Ensemble Empirical Mode Decomposition and Minimum Entropy Feature” proposed an adaptive method of magnetic anomaly detection (MAD) with ensemble empirical mode decomposition and minimum entropy (EEMD-ME) feature to improve the detection performance in the case of low SNR and the complex magnetic environment.

A. Chang et al. in their paper entitled “Measurement of Cotton Canopy Temperature Using Radiometric Thermal Sensor Mounted on the Unmanned Aerial Vehicle (UAV)” developed a UAV-based thermal system using a quadcopter platform and a radiometric thermal sensor.

F. Cui et al. in their paper entitled “Prediction Model of Thermal Thawing Sensibility and Thaw Depth for Permafrost Embankment along the Qinghai-Tibet Engineering Corridor Using MODIS Data” developed prediction models of permafrost thermal thawing sensibility and thaw depth by incorporating the MODIS and in situ soil temperature observation data to reveal the distribution law of permafrost thermal thawing sensibility and thaw depth caused by road construction in Qinghai-Tibet engineering corridor (QTEC).

A. R. Claridades and J. Lee in their paper entitled “Developing a Data Model of Indoor Points of Interest to Support Location-Based Services” proposed a spatial-temporal Indoor POI data model to provide direction for the establishment of indoor POI data and to address limitations in currently available data specifications.

S. Lee et al. in their paper entitled “Priority Analysis of Remote Sensing and Geospatial Information Techniques to Water-Related Disaster Damage Reduction for Inter-Korean Cooperation” identified remote sensing and GIS techniques that could be useful in reducing the damage caused by water-related disasters while considering inter-Korean relations and the disasters that occur in Democratic People's Republic of Korea.

J. Ning et al. in their paper entitled “An Improved Full-Aperture ScanSAR Imaging Method Integrating the MIAA Based Aperture Interpolation” proposed an improved algorithm based on the missing-data iterative adaptive approach (MIAA) effectively to suppress the spikes, which adversely

affect the ScanSAR-related applications, such as target detection and interferometry.

P. Liu et al. in their paper entitled “RS and GIS Supported Urban LULC and UHI Change Simulation and Assessment” designed a remote sensing-based framework that investigates and analyzes how the land use and land cover changes (LULCC) in the process of urbanization affected the thermal environment.

X. Wang and X. Wang in their paper entitled “Spatiotemporal Fusion of Remote Sensing Image Based on Deep Learning” proposed a residual convolution neural network to improve the accuracy of spatiotemporal fusion.

X. Zhu et al. in their paper entitled “Regional Patch Detection of Road Traffic Network” proposed a regional patch detection (RPD) analysis for the road traffic network (RTN) structure.

Conflicts of Interest

The editors declare that they have no conflicts of interest regarding the publication of this special issue.

Acknowledgments

We would like to thank all the authors for their valuable contributions in this special issue as well as all the reviewers who are all experts on the theme. We also would like to thank all members of the editorial board for approving this special issue. This study was funded by (1) the National Research Foundation of Korea funded by the Korea government (NRF-2018M1A3A3A02066008) for H.S. Jung, (2) the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2020R1A2C1003451) for S.H. Hong, and (3) the National Nature Science Foundation of China under Grant 41774023 for L. Zhang.

*Hyung-Sup Jung
Sang-Hoon Hong
Lei Zhang*

Research Article

Advantage of Combining OBIA and Classifier Ensemble Method for Very High-Resolution Satellite Imagery Classification

Ruimei Han,^{1,2} Pei Liu ,^{1,2,3} Guangyan Wang,⁴ Hanwei Zhang,^{1,2} and Xilong Wu¹

¹Key Laboratory of Spatio-Temporal Information and Ecological Restoration of Mines(MNR), Henan Polytechnic University, Jiaozuo, Henan 454003, China

²School of Surveying and Mapping Land Information Engineering, Henan Polytechnic University, Jiaozuo, 454003 Henan, China

³Collaborative Innovation Center of Aerospace Remote Sensing Information Processing and Application of Hebei Province, 065000 Langfang, China

⁴Jiangsu Province Engineering Investigation and Research Institute Co. Ltd., Yangzhou, Jiangsu 225001, China

Correspondence should be addressed to Pei Liu; cumtlp@qq.com

Received 21 April 2020; Revised 30 September 2020; Accepted 28 October 2020; Published 25 November 2020

Academic Editor: Sang-Hoon Hong

Copyright © 2020 Ruimei Han et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate and timely collection of urban land use and land cover information is crucial for many aspects of urban development and environment protection. Very high-resolution (VHR) remote sensing images have made it possible to detect and distinguish detailed information on the ground. While abundant texture information and limited spectral channels of VHR images will lead to the increase of intraclass variance and the decrease of the interclass variance. Substantial studies on pixel-based classification algorithms revealed that there were some limitations on land cover information extraction with VHR remote sensing imagery when applying the conventional pixel-based classifiers. Aiming at evaluating the advantages of classifier ensemble strategies and object-based image analysis (OBIA) method for VHR satellite data classification under complex urban area, we present an approach-integrated multiscale segmentation OBIA and a mature classifier ensemble method named random forest. The framework was tested on Chinese GaoFen-1 (GF-1), and GF-2 VHR remotely sensed data over the central business district (CBD) of Zhengzhou metropolitan. Process flow of the proposed framework including data fusion, multiscale image segmentation, best optimal segmentation scale evaluation, multivariate texture feature extraction, random forest ensemble learning classifier construction, accuracy assessment, and time consumption. Advantages of the proposed framework were compared and discussed with several mature state-of-art machine learning algorithms such as the k -nearest neighbor (KNN), support vector machine (SVM), and decision tree classifier (DTC). Experimental results showed that the OA of the proposed method is up to 99.29% and 98.98% for the GF-1 dataset and GF-2 dataset, respectively. And the OA is increased by 26.89%, 11.79%, 11.89%, and 4.26% compared with the traditional machine learning algorithms such as the decision tree classifier (DTC), support vector machine (SVM), k -nearest neighbor (KNN), and random forest (RF) on the test of the GF-1 dataset; OA increased by 32.31%, 13.48%, 9.77%, and 7.72% for the GF-2 dataset. In terms of time consuming, by rough statistic, OBIA-RF spends 223.55 s, SVM spends 403.57 s, KNN spends 86.93 s, and DT spends 0.61 s on average of the GF-1 and GF-2 datasets. Taking the account classification accuracy and running time, the proposed method has good ability of generalization and robustness for complex urban surface classification with high-resolution remotely sensed data.

1. Introduction

The classification accuracy of remotely sensed data and its sensitivity to classification algorithms have a critical importance for the geospatial community, as classified images provide the base layers for many applications and models [1].

Recent availability of submeter resolution imagery from advanced satellite sensors, such as WorldView-3 and Chinese GaoFen series, can provide new opportunities for detailed urban land cover mapping at the object level [2]. Applications such as environmental monitoring, natural resource management, and change detection require more accurate,

detailed, and constantly updated land cover-type mapping [3]. Detailed urban land cover information is not only essential for understanding the urban environment changes and monitoring and managing the urban ecological environment but also for supporting the government to make a decision on urban expansion, urban planning, and management [4–6]. In the last few decades, it has become an effective and convenient mean to obtain this information from remotely sensed imagery, because of its unique advantages of frequent and wide coverage, by machine learning classification technology [7].

Most of land use and land cover classification research are traditionally based on low- and medium-resolution remotely sensed imagery, such as MODIS [6, 8, 9], Landsat [10–12], and SPOT1/4 [13]. However, urban surface coverage presents high-frequency heterogeneity, resulting in a large number of mixed pixels in medium- and low-resolution images. With the rapid development of sensor technology, a large number of high-resolution remotely sensed imagery (IKONOS, Quickbird, GeoEye-1, WorldView-1-4, GF-1/2, etc.) in meters or submeters are becoming more and more popular [14]. With the characteristics of high definition and abundant spatial information, high-resolution satellite image can compensate the shortcomings of mixing pixels in low- and medium-resolution images in urban land cover classification [15, 16]. And high spatial resolution images, where spatial resolution is equal or a little equal to 4 meters, could make it possible to map complex urban surface. A major challenge in using high spatial resolution for detailed urban mapping comes from the high level of intraclass spectral variability, such as building roof and road, and low level of interclass spectral variability, such as water body and shadow. In this condition, traditional pixel-based classification algorithms such as the maximum likelihood classification (MLC) can easily make missclass error and generate the salt-and-pepper effect which may reduce classification accuracy for very high-resolution imagery.

There are currently various classification algorithms, each with its own advantages and limitations [17]. And that, common mature statistical-based machine learning algorithms, such as MLC, requires hypothesis that training data follows a normal distribution, but high-resolution images cannot meet this requirement. Many previous studies revealed a bunch of machine learning algorithms such as support vector machines (SVM) [1], artificial neural networks (ANN) [18], and decision tree [19] have been popular for land cover classification. These classifiers always have limitations in practical applications in areas such as volatile and complex urban area, due to the enhanced complex of spatial relationship between pixels and the complex earth's surface phenomenon [21, 22]. Recently, ensemble methods have been introduced to integrate multiple single classifiers to improve classification performances. The combination of multisource remote sensing and geographic data is believed to offer improved accuracies in land cover classification [23]. In general, there are two steps to build the ensemble, namely, generating base learners and combining base learners. In order to obtain a good ensemble, the base learner should be as accurate as possible and as diverse as possible. Due to its high potential and superior performance, ensemble methods have been employed in a

remote sensing community. Existing theoretical and empirical studies have reported that ensemble classifiers can obtain more accuracy prediction and outperform individual classifiers [3, 17, 23–25]. The random forest (RF) classifier, as one of the more popular ensemble learning algorithms in recent years, is composed of multiple decision trees in that each tree is trained using bootstrap sampling and employing the majority vote for the final prediction [26, 27]. It has received increasing attention due to its excellent classification result, the ability to avoid overfitting, and the rapid speed to process [28–31].

In order to overcome limitations of pixel-based classification, object-based image analysis (OBIA) or geospatial object-based image analysis (GEOBIA) has been introduced to improve the quality of information extraction from high-resolution imagery. Image segmentation is a critical and important step in (geographic) object-based image analysis (GEOBIA or OBIA). The final feature extraction and classification in OBIA are highly dependent on the quality of image segmentation [32]. There are two main steps, containing segmentation and classification in OBIA. The processed object of OBIA is not a pixel, but an object composed of multiple adjacent homogenous pixels through segmentation, which containing not only the spectral information but also the textual and contextual information from imagery [32, 33]. Due to its advantages, OBIA has been more popular in the remote sensing community and successfully applied in land cover classification [34–36]. Many previous studies have showed that the OBIA method had outperformed the pixel-based classification [32, 37–40]. However, the number of input features used for classification has grown exponentially, some of which are irrelevant and redundant features, affecting the performance of the classifier, especially, when the purpose of segmentation has been changed from helping pixel labeling to object identification at present era [32].

In this paper, we verify the ability of GF-1 and GF-2 very high-resolution imagery in urban land use and land cover classification. For this purpose, we combined the random forest ensemble classifier with the OBIA method. We test the method on two selected complex urban areas of a metropolis city. And the proposed strategy was also compared with the pixel-based random forest and the state-of-art mature machine learning algorithm including the SVM, KNN, and DT classifiers from classification accuracy and operational efficiency aspects.

The rest of this paper is organized as follows: a brief introduction about the study area and dataset and preprocessing are given in Section 2. The framework and details of the proposed methodology strategy based on object-oriented analysis and random forest are drawn in Section 3. The results and discussion are shown in Section 4. Finally, the conclusions are drawn in Section 5.

2. Study Area and Data Preprocess

The study site is a central business district (CBD) of Zhengdong new district which is located in the eastern part of Zhengzhou city, capital of Henan province, and is a new urban area invested and developed by Zhengzhou Municipal Committee, municipal government in accordance with the

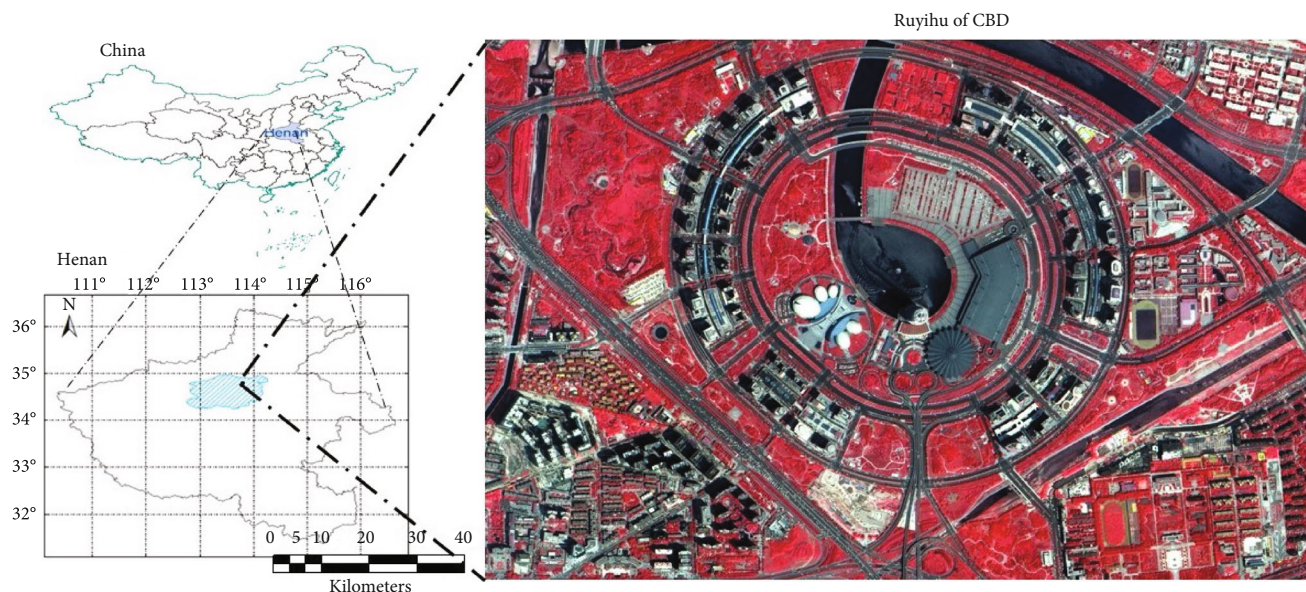


FIGURE 1: The study area location and corresponding GF-2 imagery.

TABLE 1: The classification scheme of land cover in the study area.

Category	Symbols	Description
Urban building areas (UB)	1	Residential area, campus, low density building
Urban commercial areas (UC)	2	Commercial area, high density building
Urban green areas (UG)	3	Grassland, tree group, oasis in the lake, artificial grass
Urban road areas (UR)	4	Cement and asphalt pavement
Urban water area (UW)	5	River, lake
High building shadow (HS)	6	Shadow of high buildings

State Council approved the City of Zhengzhou city master plan in order to implement the megacity framework, expand the size of the city, and accelerate urbanization and urban modernization strategy. Based on the National Economic and Technological Development Zone, the original area of CBD is about 25 km², west from 107 national road, east to Jingzhu Expressway, south to the airport highway, north to Lianhuo Expressway, and the long-term planning area of CBD is about 150 km². The study area is focused on Ruyihu, the center of CBD (see Figure 1), which is surrounded by three landmarks of the CBD-Zhengzhou International Convention and Exhibition Center, Henan Arts Center, Zhengzhou Convention and Exhibition Hotel. The land surface is dominated by human-made material, which is a challenging task to identify different land use and land cover types. According to the planning and construction situation, the types of surface cover are mainly divided into the urban building areas (UB), urban commercial area (UC), urban green areas (UG), urban road areas (UR), urban water area (UW), and high building shadow (HS) (see Table 1). Data availability statement stated that the very high-resolution remotely sensed data used in this research is provided by Henan Data and Application Center of the High Resolution Earth Observation System through signing a contract with National Defense Science and Technology Bureau of Henan.

Unfortunately, we do not have the priority to share the high-resolution satellite remote sensed data. Anyway, we can share our code used in this research. Researchers can test algorithms using these codes with their own datasets and repeat experiments to obtain similar research conclusions. Researchers who are interested in this code can download it from hyperlink <https://pan.baidu.com/s/19nXD7oHwq0FnpZJ7T6p5HQ>, using password 6xae, or contact with the corresponding author to obtain source data to conduct secondary analysis.

2.1. Remote Sensing Data and Preprocessing. Under the background of “Chinese high-resolution earth observation” major project, a series of high-resolution satellites have been launched, involving GF-1 (2 m res. panchromatic camera/8 m res., multispectral camera/16 m res., and wide-angle multispectral camera), GF-2 (1 m res., panchromatic camera/4 m res., and multispectral camera), GF-3 (1 m res., C-band synthetic aperture radar), GF-4 (50 m res., fixed-point camera in geostationary orbit), GF-5 (VNIR hyperspectral camera), GF-6 (2 m res., wide-angle multispectral camera), and GF-7 (stereographic cartography cameras). GF-1 satellite is the first low earth orbit remote sensing satellite of China’s high-resolution earth observation system, which breaks through the key technologies of optical remote sensing for

TABLE 2: Sensor parameters of GF-1 and GF-2 satellites.

Parameters		PAN/multispectral	
		GF1	GF2
Satellite	Orbit type	Sun synchronization	Sun synchronization
	Orbit altitude	646 km	631 km
	Repeat time	41 days	69 days
	Panchromatic	0.45–0.90 μm	0.45–0.90 μm
Spectral range	Multispectral	Blue	0.45–0.52 μm
		Green	0.52–0.59 μm
		Red	0.63–0.69 μm
		NIR	0.77–0.89 μm
Spatial resolution	Panchromatic	2 m	1 m
	Multispectral	8 m	4 m
Width		60 km	45 km
Receive time		2015-07-14	2015-7-3
Orbit ID		11937	5113
Product type		Standard	Standard
Product level		Level 1A	Level 1A
(Path, row)		(3, 97)	(3,152)
(Width, height)		(4548, 4544), (18192, 18164)	(29200, 27620)
Cloud		0%	5%
Solar(azimuth, zenith)		(141.278, 73.7496)	(134.537, 21.6698)
Satellite(azimuth, zenith)		(297.354, 88.3146)	(283.342, 83.754)

high spatial resolution and multispectral and wide coverage. It can meet the needs of research data support in the fields of resources and environment, precision agriculture, and disaster measurement, which has become an important means of information services and other aspects. It is of great strategic significance to improve the level of satellite engineering in China and the self-sufficiency rate of high-resolution data.

The selected remotely sensed data is GF-1 and GF-2 very high-resolution satellite images, which was acquired on July 14, 2015 and July 30, 2015, respectively. The specific parameters are shown in Table 2.

The preprocess of selected dataset includes radiation calibration, atmospheric correction, geometric registration, orthorectification, image fusion (NNDiffuse pan-sharpening algorithm), and image resize. Radiation calibration is the process of converting DN values of image data into apparent reflectivity using atmospheric correction techniques. Equation (1) can be used to convert the channel observation DN value to equivalent brightness value.

$$L_{\epsilon}(\lambda_{\epsilon}) = \text{Gain} * \text{DN} + \text{Bias}, \quad (1)$$

where Gain is the calibration slope, DN is satellite observation value, Bias is calibration intercept, and these parameters can be obtained from meta file with satellite data. Then, the apparent reflectivity can be calculated based on the brightness value using

$$\rho = \frac{\pi L_{\lambda} d^2}{\text{ESUN}_{\lambda} \cos \theta}, \quad (2)$$

where ESUN is solar spectral radiation, d is solar-earth distance, and $\cos \theta$ is the zenith angle of sun.

The selected atmospheric correction is based on a 6S radiative transfer model, which is a package included in Pixel Information Export (<http://www.piesat.cn/en/index.html>). The purpose of atmospheric correction is to eliminate the absorption and dispersion from the sun and target.

Geometric correction includes image registration and orthorectification. The purpose of geometric correction is to correct image deformation caused by system and nonsystemic factors. In this research, the image-to-image registration method was selected to correct multispectral data based on panchromatic data of GF-1 and GF-2 sensors, respectively.

Orthorectification is the process of correcting image space and geometric distortion to generate a multicenter projection plane orthographic image. In addition to correcting geometric distortions caused by general system factors, it can also eliminate geometric distortion caused by terrain.

Image fusion is the process of generating new images under the prescribed geographical coordinate system according to a certain algorithm. This study combines multispectral data with high spatial resolution and single-band images with high spatial resolution, making the fused images have both high spatial resolution and rich spectral resolution.

3. Methodology

The proposed methodology in this research (shown in Figure 2) includes three main stages: (1) multiscale segmentation and multifeature extraction; (2) construction of the state-of-art machine learning algorithms such as decision

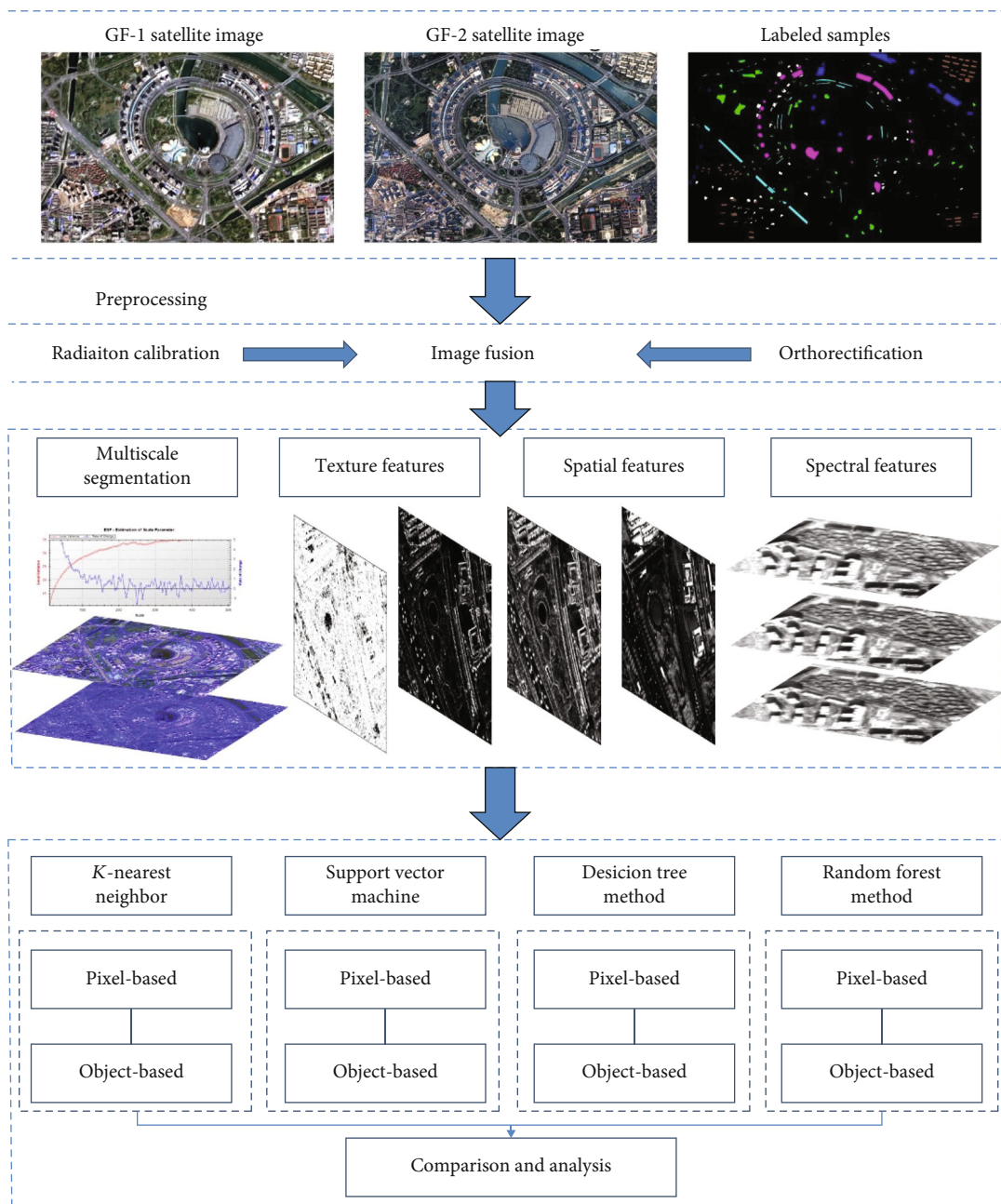


FIGURE 2: Framework of urban land cover classification using GF-1 and GF-2 remotely sensed data.

tree classifier (DTC), random forest (RF), support vector machine (SVM), *k*-nearest neighbor (KNN), and object-based image analysis random forest (OBIA-RF); and (3) accuracy assessment and comparison analysis and discussion; more details can be found in Figure 2.

3.1. Multiscale Segmentation and Feature Extraction. Very high-resolution (VHR) remote sensing images have a limitation in spectral information which means there are 4 spectral bands including green, blue, red, and near infrared in general. While the VHR remote sensing images are always rich in detailed characters, more specific details of land surface will be presented on the images. In order to overcome this shortcoming, we conquer the disadvantage and make full use of

advantages of these data. After data preprocessing, we performed multiscale segmentation and employed feature extraction based on the segmented results to obtain multifeature image sets as inputs of image classification models.

Quality of segmentation has a direct effect on the performance of classification, which is related to the segmentation parameters selected by an analyst. Most of the segmentation algorithms are regarded as a subjective task with the trial-and-error strategy. The multiscale segmentation algorithm, the most popular method currently, employed in this experiment is merging pixels of the original image into small object patches from bottom to top, and then merging the small patches into large patches to complete the merging of the regional objects [41–43]. Three major parameters for the

TABLE 3: Overview of extracted features based on segmentation.

Feature type	Quantity	Feature name
Spectral feature	9	Spectral bands 1-4, brightness, ratio band, standard deviation band4, NDVI, NDWI
Texture feature	8	Entropy, mean, variance, homogeneity, contrast, dissimilarity, correlation, angular second moment
Spatial feature	7	Area, border index, compactness, density, length/width, length, shape index

multiscale segmentation algorithm are scale, shape, and compactness, defining within-object homogeneity. Here, we select the appropriate scale parameters and heterogeneity standard specifications to ensure the highest homogeneity within the generated object and the heterogeneity between adjacent objects and other objects. Scale parameter is considered the most effective parameter affecting the segmentation quality [44–46]. In this study, shape parameter was set as 0.1. The optimal segmentation scale parameter of GF-1 image is quantitatively evaluated by ESP-2 (estimation of scale parameter), of which the principle is to select the optimal scales based on the rate of change (ROC) curve for the local variance (LV) of object heterogeneity at a corresponding scale [44, 47]. Peaks value of ROC-LV curves were considered the most appropriate segmentation scales at which the image can be segmented in the most optimal levels. Methodology model of ROC can be described as

$$\text{ROC} = \frac{LV_L - LV_{L-1}}{LV_{L-1}} \times 100, \quad (3)$$

where LV_L is mean standard deviation of the object in the L layer and LV_{L-1} is mean standard deviation in the next lower layer. When ROC was obtained, the optimal segmentation scale parameter is selected by visually interpreting based on segmentation result and the boundary matching effect of the actual feature [47].

On the basis of the best segmentation result, a total of 24 spectral features, texture features, and spatial geometric features were extracted. The extracted spectral, textural, and spatial features from segmented VHR remotely sensed data can be summarized as shown in Table 3.

In detail, the selected spectral features are mean value of all four bands (which means average value of all image objects). The brightness feature is that the sum of the average values of the layers containing spectral information divided by the number of layers of the image object, which can be calculated by

$$B = \frac{1}{N_{\text{band}}} \sum_{i=1}^n \bar{C}_i, \quad (4)$$

where B is the brightness value of the object, N_{band} is the total number of bands contained in the object, and C_i is the average gray value of the object.

In the high-resolution image, since the reflectivity of the water body in the near-infrared band is significantly lower than that of other ground objects, the shadows are very similar in many features with water body so that they are difficult to separate. In order to highlight the water, the ratio

and standard deviation of the fourth band and NDWI were additionally extracted. The ratio of the fourth band is the average gray value of all pixels in the fourth band divided by the average gray value of all pixels in all 4 bands of the image. In addition, only layers containing spectral information can be used to obtain reasonable results. The standard deviation of band 4 is calculated from all the pixel values contained in an object in band 4.

The NDWI refers to the normalized ratio index between the green band and the near-infrared band in the image. Using NDWI can better distinguish the water in the image from other features. It can be calculated by

$$\text{NDWI} = \frac{B_{\text{green}} - B_{\text{nir}}}{B_{\text{green}} + B_{\text{nir}}}, \quad (5)$$

where B_{green} is the value of object in the green band and B_{nir} is the value of object in the near-infrared band.

NDVI is a vegetation index proposed based on the reflection characteristics of vegetation in the visible and infrared bands. It is the ratio of the difference between the reflection intensity value in the visible red band and the reflection intensity value in the near-infrared band to the sum of the two. The formula can be described as

$$\text{NDVI} = \frac{B_{\text{nir}} - B_{\text{red}}}{B_{\text{nir}} + B_{\text{red}}}, \quad (6)$$

where B_{nir} is the value of object in the near-infrared band and B_{red} is the value of object in the red band.

Although limited to spectral information, high-resolution remote sensing images contain rich geometric and structural information, which can reflect the spatial distribution and geometric forms of ground objects. The selected texture features contain eight features unit extracted based on the gray level cooccurrence. The selected eight textural features extracted from GLCM include entropy, mean, variance, homogeneity, contrast, dissimilarity, correlation, and angular second moment. Gray-level cooccurrence matrix (GLCM), which is calculated based on statistic method, is also known as gray-level spatial dependence matrix considered one of the most popular techniques used for texture analysis. GLCM has strong ability to assess texture features by considering spatial relationship of pixels and its surrounding. GLCM mean value is not simply the average of all original pixel values; pixel value is weighted by its frequency of its occurrence in combination with a certain neighbor pixel value. Variance in GLCM texture performs the same task as does the common descriptive statistic called variance.

Entropy measures the complexity of a given image, which reflects the sharpness of the image and the depth of the texture; entropy can be calculated using

$$\text{Entropy} = -\log P_{i,j} \sum_{i,j=0}^{N-1} P_{i,j}, \quad (7)$$

where i and j standards position of pixels of GLCM and $P_{i,j}$ is probability of presence of pixel pairs at certain distance and angle.

Contrast measures local variations and texture of shadow depth in GLCM. The larger the contrast, the deeper groove of texture and clearer effect of the image will be shown. Contrast can be calculated by

$$\text{Contrast} = \sum \sum (i-j)^2 P_{i,j}. \quad (8)$$

Homogeneity represents values by the inverse of the contrast weight, with weights decreasing exponentially away from the diagonal, which can be calculated using

$$\text{Homogeneity} = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i-j)^2}. \quad (9)$$

Correlation coefficient concludes that the degree of two variable's activities is associated and can be calculated by

$$\text{CC} = \sum_{i,j=0}^{N-1} (i-\bar{i})(j-\bar{j}) \frac{P_{i,j}}{\delta_x \delta_y}. \quad (10)$$

Angular second moment (ASM) uses $P_{i,j}$ as weight for itself; high values of ASM occurs when the window is very orderly, so it measures the homogeneousness of a given image, and ASM can be calculated using

$$\text{ASM} = \sum_{i,j=0}^{N-1} P_{i,j}^2. \quad (11)$$

In addition to spectral and texture features, combining the geometric characteristics of high-resolution remote sensing images are extremely significant for detailed land cover information extraction. In this article, the article seven spatial geometric features include area, border index, compactness, density, length, length/width, and shape index were selected. Area of an image can be obtained through multiplying number of pixels constituting the image object and the covered area of the object. The boundary index can be calculated as the ratio between the boundary length of the image object and the smallest enclosing rectangle. The tighter the image object, the smaller its border. The density describes the distribution in the pixel space of the image object, that is, how tight the image object is. Density is based on the covariance matrix, which is calculated by dividing the number of pixels constituting the image object by its approximate radius. Length-width ratio can be used as one of the features for road

extraction. It is calculated by the length and the width of the objects. In addition, the length-width ratio can be used to calculate the length of the image object. Shape index was selected to describe the smoothness of the surface of an object. The smoother the surface of the image object, the lower its shape index. The more fragmented the image object, the larger its shape index. It can be calculated by dividing the frame length of an object by the volume of the object.

3.2. Classification Algorithms. In this research, an OBIA-RF method, also known as a combination of OBIA and classifier ensemble method which can take advantage of OBIA and classifier ensemble was constructed, and four state-of-art classification algorithms named KNN, SVM, and DTC were selected for performance comparison. Performance evaluation was carried out by quantitative indicators such as overall accuracy, kappa coefficient, and execution time consumption.

3.2.1. Random Forest. Random forest algorithm is an ensemble learning method proposed by Leo Breiman in 2001 [48], and is one of the most well-known ensemble learning methodology and has advantages of, i.e., performing out-of-sample prediction rapidly, requiring only slight parameter tuning, having capable ranking of the importance of features [28]. Decision trees in RF are generated by randomly selecting sample (bootstrap sampling) subsets in the training sample set and randomly selecting the feature variables to achieve optimal splitting. The obtained decision trees do not need pruning, and the final classification result is obtained by the majority vote method from the classification results of all decision trees in the integration. Gini index which measures the impurity of a given element with respect to the result of the classes is selected as a measure for the best split selection for RF [49]. There are two key parameters in the process of constructing a random forest pattern: the number of spanning trees and the number of randomly selected features. By literature review, the number of selected features is more important than the number of how many trees are trained; especially, generally, each split number of randomly selected features is set as the square root of the number of input characters [50–52]. This parameter can be optimized based on the out-of-bag error estimate. In this research, the number of trees is set as 100 and number of random attribute selection is $\log_2 d$, where d is the total number of features. And then, the number of trees for RF was tuning from 50 to 500 with a step of 10.

3.2.2. Support Vector Machine. SVM is one of the most appealing algorithms for remotely sensed data classification due to their advantages of generalization even with limited training samples which is common in remote sensing data processing [53]. And, as a supervised nonparametric statistical learning method, SVM does not need a training set strictly conforming to the standard independent and identical distribution. The advantages of SVM come from two aspects, transforming original space training set into a very high-dimensional new space and finding a large margin linear boundary in the new space. SVM is a classifier based on theory of structural risk minimization, which tries to lower the

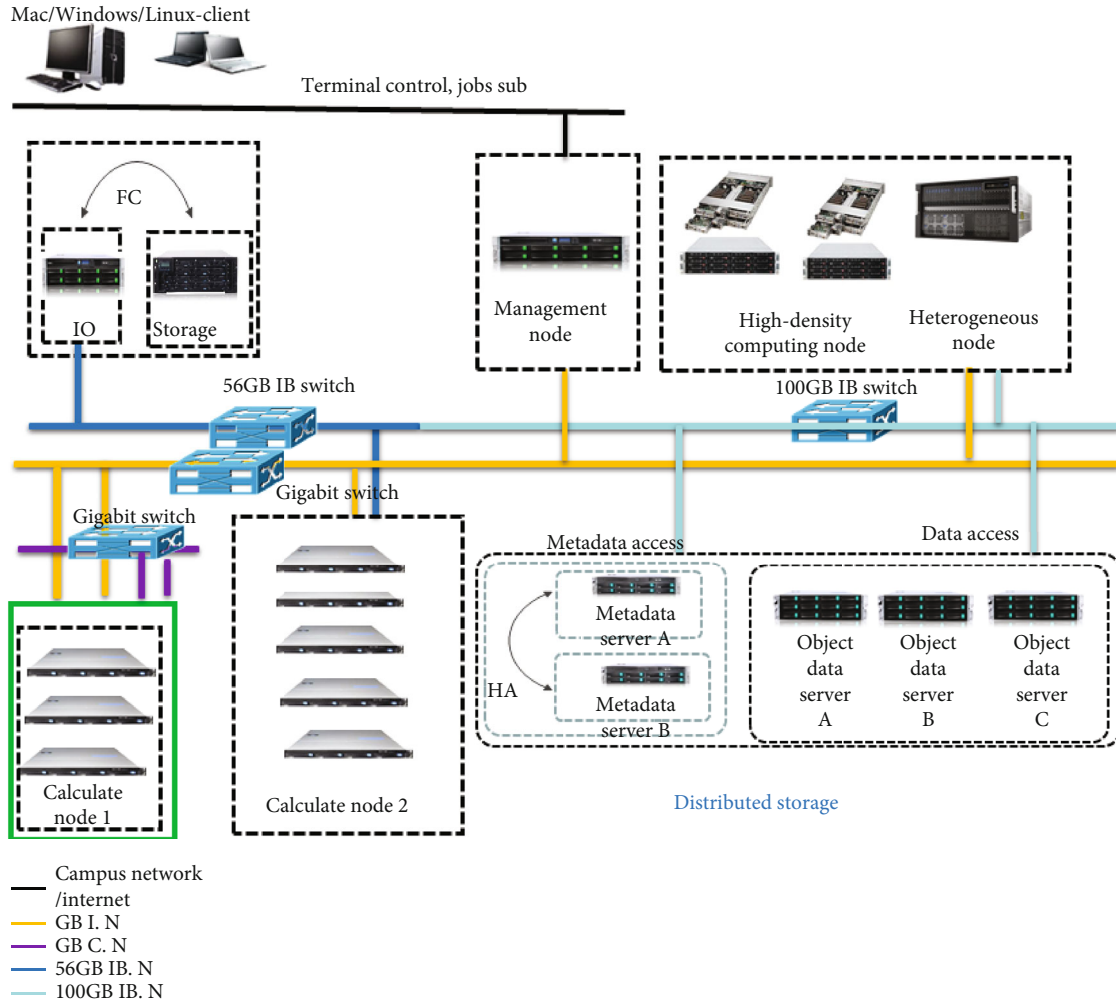


FIGURE 3: Cluster topology of high-performance computing.

generalization error by maximizing the margins on the training data. Thus, SVM looks for an ideal margin by solving optimization problem as

$$\begin{aligned} \min_{\omega, b} \frac{1}{2} \|\omega\|^2, \\ \text{s.t. } y_i(\omega^T x_i + b) \geq 1, i = 1, 2, \dots, m. \end{aligned} \quad (12)$$

Furthermore, for classes that are nonseparable, the optimization can be solved by the so-called ‘kernel stick.’ The optimization procedure seeks to find coefficients α_i and w_0 in Equation (13), where $K(\cdot, \cdot)$ is kernel function. By default, the kernel function is set as the Gaussian kernel, kernel scale is 8, and box constraint is 1 standardized. And then, the scale of kernels for SVM was tuning from 0.1 to 8 with a step of 0.4

$$f(x) = \sum_{i=1}^N \alpha_i K(x, z_i) + \omega_0. \quad (13)$$

3.2.3. k -Nearest Neighbor. The k -nearest neighbor classifier (k -NN) is a kind of nonparametric and memory-based

learning, as well as instance-based learning or lazy learning used for classification and regression [26]. In the classification procedure, a given pixel will be classified by plurality vote of its neighbors in the feature space. The most intuitive k -NN classifier is 1-NN classifier; in this case, a given pixel will be assigned to the class of its closest neighbor in the feature space, which can be described as $C_n^{1NN}(x) = Y(1)$. The useful technique which can help nearer neighbors contribute more than the more distant one is assigning different weights to the neighbors. A common weighting scheme is setting each neighbor a weight of $1/d$, where d is the distance to the neighbor. Classification performance of k -NN can be significantly improved by metric learning, and diversity can be introduced to k -NN classifier by using different subsets of features, different distance metrics, different values of k , etc. In the experiment part, Euclidean distance is selected and number of neighbors is 100, distance weight is setting as equal to evaluate algorithm performance, and the number of neighbors for KNN was tuning from 1 to 300 with step of 10.

3.3. Comparisons and Assessment. Classification accuracy was evaluated by the confusion matrix as well as overall accuracy and the Kappa coefficient [54]. With the help of

TABLE 4: Labeled samples.

	UB (pixels)	UC (pixels)	UR (pixels)	UG (pixels)	UW (pixels)	HS (pixels)
GF-1	628	524	846	1085	427	1246
GF-2	30620	58059	34323	48874	39996	27412

TABLE 5: Overall accuracy and kappa coefficient of selected algorithms.

		DTC	SVM	KNN	RF	
GF-1	P	OA (%)	72.4	87.5	87.4	94.67
		Kappa	0.68	0.85	0.85	0.93
		Time used (s)	0.032	0.57	0.05	27.13
	O	OA (%)	82.86	90.86	89.71	99.43
		Kappa	0.79	0.89	0.87	0.99
		Time used (s)	0.034	0.71	0.1	15.53
GF-2	P	OA (%)	66.67	85.5	89.21	91.26
		Kappa	0.63	0.83	0.87	0.89
		Time used (s)	0.36	589.31	1.89	251.32
	O	OA (%)	64.76	94.97	96.69	98.98
		Kappa	0.6	0.94	0.96	0.98
		Time used (s)	0.58	402.86	86.83	208.02

the confusion matrix, overall accuracy and kappa coefficient can be calculated using

$$OA = \frac{\sum_{i=1}^q n_{ii}}{n} \times 100\%, \quad (14)$$

$$Kappa = \frac{n \sum_{i=1}^q n_{ii} - \sum_{i=1}^q n_{i+} n_{+i}}{n^2 - \sum_{i=1}^q n_{i+} n_{+i}} \times 100\%, \quad (15)$$

where q is the number of classes, n represents the total number of considered pixel, n_{ii} are the diagonal elements of the confusion matrix, n_{i+} represents the marginal sum of the rows in the confusion matrix, and n_{+i} represents the marginal sum of the columns in the confusion matrix [55].

All experiments are performed on high-performance computing system using a portable bash system (PBS), as shown in Figure 3. Each algorithm is programmed as a job which can be submitted to cluster. Finally, time consumption of all algorithms is compared.

4. Results and Discussion

4.1. LULC Mapping. On the basis of segmentation results, spectral features and texture features are integrated and fused to generate multifeature images as inputs to all classifiers. In this research, the best segmentation scale is setting to 105 for GF-1 dataset and 210 for GF-2 dataset through estimation of scale parameter analysis. Land cover types in the study area include 6 categories named UB, UC, UR, UG, UW, and HS, more details can be found in Table 1. Training and testing samples are labeled by an expert of remote sensing with the

assistance of Google Earth. Labeled samples for training and testing are shown in Table 4. In order to test the sample sensitivity of the proposed process chain, limited and sufficient samples were selected from the GF-1 and GF-2 datasets, respectively. And 3-fold out validation method was selected for accuracy assessment.

During the research procedure, the same labeled training and testing samples are used as inputs of the DT, SVM, KNN, and RF classifiers. The default parameters of the constructed DT, RF, SVM, and KNN classifier models are selected for GF-1 and GF-2 image processing. Overall accuracy, kappa coefficient of all experiments, and the best classification results for GF-1 and GF-2 are shown in Table 5 and Figures 4 and 5.

It can be seen that the OBIA-RF algorithm has best classification results with an overall accuracy of 99.43% and 98.98% for GF-1 and GF-2, respectively (Table 5 and Figures 4 and 5). In general, the correct classification accuracy of all categories covered by urban land surface reached 91%. The overall accuracy of the original RF classification is 94.67% for GF-1 and 91.26% for GF-2, and the Kappa coefficient is 0.93 and 0.89, respectively. The overall accuracy of the SVM classification method is 87.5% and 85.5%, and Kappa coefficient is 0.85, 0.83 for the GF-1 and GF-2 datasets, respectively. The overall accuracy of the DTC classification is the lowest, only 72.4% and 66.67% for GF-1 and GF-2.

By analyzing the single-class accuracy of original RF and OBIA-RF classification results, the accuracy of corrected classification of UB, UR, and UC is relative lower than other classes, that is, 91.4%, 90.4%, and 92.2% for the GF-1 dataset, while the lowest single accuracy land use type of the GF-2 dataset is UR and UC for 85.4% and 86.7%, respectively. When the OBIA method is combined with RF accuracy, these difficult identified classes are improved to 91.4%, 90.4%, and 92.2% for GF-1 data and 99.2% and 99.5% for GF-2 data.

When pixel-based approach and object-based approach are compared, research results demonstrate that object-based approach outstands pixel-based approach, which is also demonstrated by other research [37, 38]. While our research further demonstrates that among the selected algorithms, OBIA is more suitable for the classifier ensemble method when compared with stand-of-art single classifier, especially for higher spatial resolution satellite data, GF-2 instead of GF-1.

4.2. General Discussion. Based on Table 5, the best accuracies are achieved by the OBIA-RF model for both GF-1 and GF-2 datasets. Obviously, these experiments showed the superiority of OBIA-RF over the selected state-of-art method in terms of classification accuracy. By statistics, the OBIA-RF method lets 4.62%, 11.89%, 11.79%, and 26.89% better accuracy than RF, KNN, SVM, and DTC for the GF-1 dataset and lets 7.72%, 9.77%, 13.48%, and 32.31% better accuracy than RF, KNN, SVM, and DTC for the GF-2 dataset.

The DTC achieved the worst overall accuracy. SVM model improves classification accuracy by 15.1% and 18.83% at pixel level and improves classification accuracy by 21.95% and 30.21% at object level for GF-1 and GF-2, respectively. The RF model led to further improvement in classification accuracy by 22.27% and 24.59% at pixel level



FIGURE 4: The classification map derived from GF-1 using the OBIA-RF method.



FIGURE 5: The classification map derived from GF-2 using the OBIA-RF method.

for GF-1 and GF-2. RF reduces the correlation between trees through random sampling of observations and features. This can demonstrate the advantage of the classifier ensemble for classification of VHR remotely sensed data. The OBIA method led to final improvement in classification accuracy by 4.62% and 7.72% in this study. This advantage is especially valuable for the relative high benchmark of random forest performance. When the OBIA method was combined with traditional machine learning model, especially classifier ensemble which takes advantage of textual, spatial structure information, and spectral information, classification results will be improved undoubtedly.

Furthermore, we investigate the sensitivity of the proposed model as well as the selected state-of-art machine learning model including DT, SVM, and KNN to parameter choice. Figures 6 and 7 plot the OA as a function of parameter for the corresponding machine learning model selected. The sensitivity of models to parameter choice for the GF-1 and GF-2 datasets (Figures 6 and 7) shows that (a) OA of the DTC model increased with the maximum number of splits, the peak value appears when number of split, GF-1 dataset equals to (290, 430) and GF-2 dataset equals to (480, 250), for pixel and OBIA training, respectively; (b) OA of SVM model increased to peak value when scale of kernels, GF-1 dataset equals to (0.5, 2.5) and GF-2 dataset equals to (0.1, 4.1), for pixel and OBIA and then significantly declines; (c) OA of KNN model appears when neighbors, GF-1 dataset equals to (291, 71) and GF-2 dataset equals to (221, 171), for pixel and OBIA training, respectively; (d) accuracy of RF fluctuates continuously with parameters changes, peak value appears when the number of trees reached, GF-1 dataset equals to (70, 240) and GF-2 dataset

equals to (230, 450), for pixel and OBIA training, respectively; and (e) the OBIA method performed better than pixel-based method for all selected model including the GF-1 dataset and GF-2 dataset.

Sensitivity test of models to parameter choice of GF-1 and GF-2 shows that rank of OA for selected model is RF, DTC, SVM, and KNN. Furthermore, we also wonder which features are most important in the procedure of prediction. With the help of out-of-bag estimation, feature importance in the RF ensemble learning are calculated and shown in Figure 8 for the GF-1 dataset and Figure 9 for the GF-2 dataset. This feature importance rank shows the contribution weight of different features for complex urban surface classification. Figure 8 demonstrates that the most important features for GF-1 remotely sensed data interpretation is texture information (mean value calculated using gray-level cooccurrence matrix(GLCM)) and the second important feature is standard deviation calculated using GLCM, and followed by the important spatial feature calculated by ratio, and then spatial feature of length/width. And, for the GF-2 dataset, the first five rank features for complex urban surface interpretation belongs to spatial feature (length, ratio) and spectral feature (NDVI, NDWI, mean2). By a comprehensive consideration of the processing results of the selected two datasets, a preliminary conclusion can be drawn as that spatial and texture features play an important role for complex urban surface classification with high- and very high-resolution remotely sensed data.

The quality of segmentation directly affects the effect of subsequent classification. In this research, the multiscale segmentation method was selected, which has three user-defined parameters: scale, shape, and compactness. And the

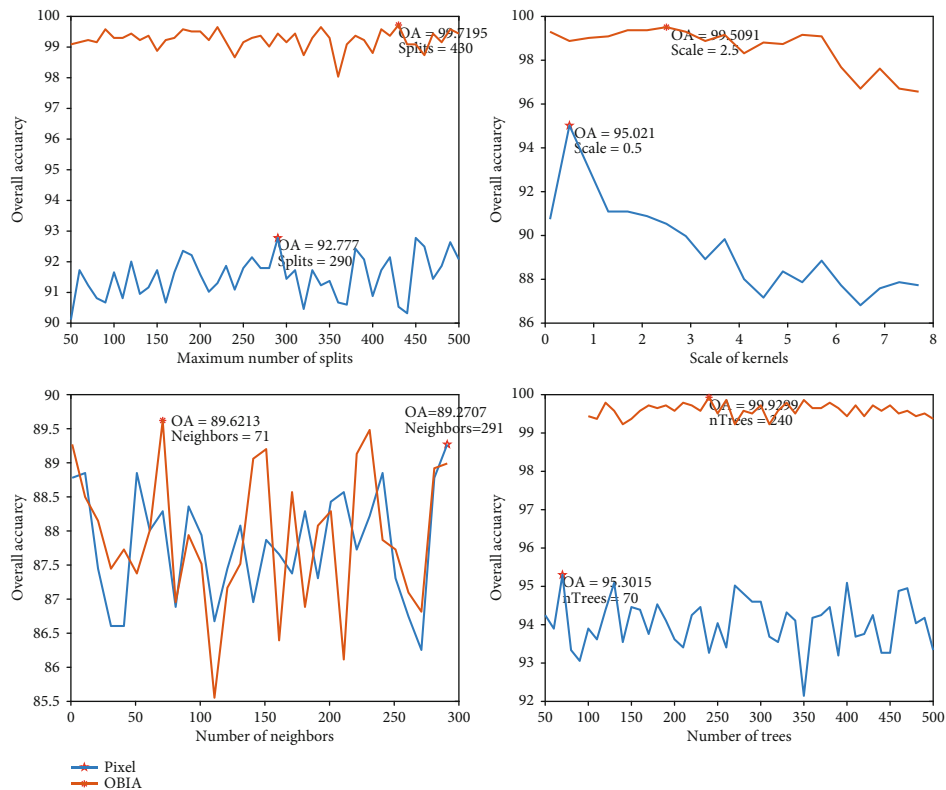


FIGURE 6: Sensitivity of models to parameter choice for the GF-1 dataset.

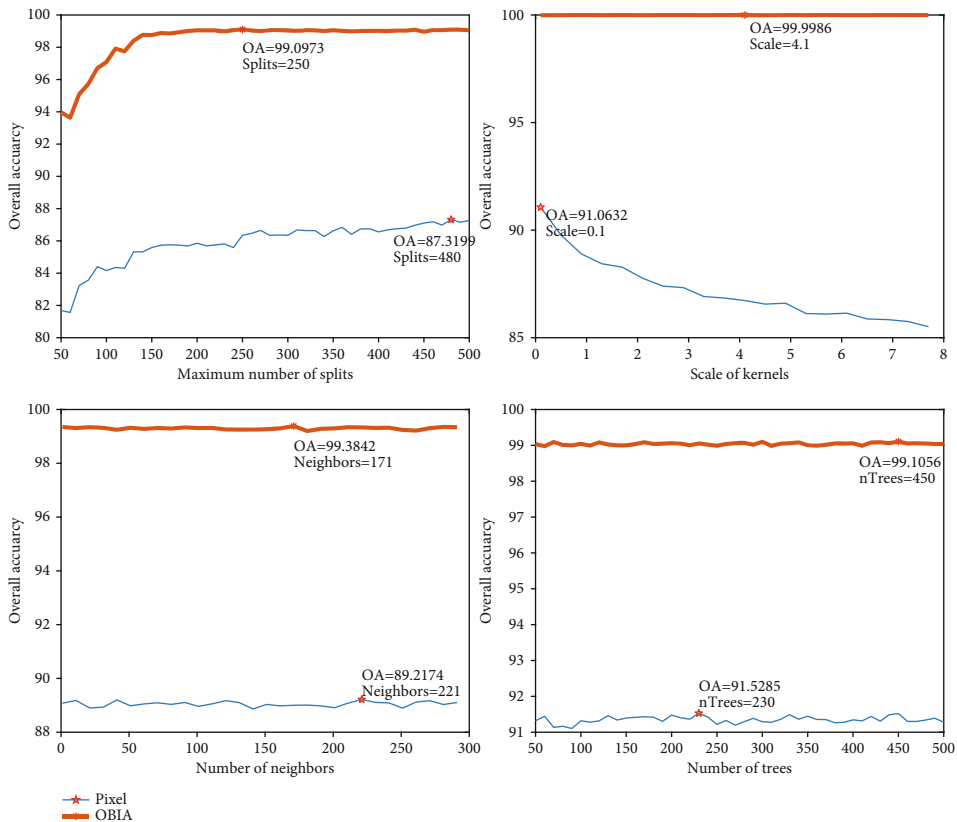


FIGURE 7: Sensitivity of models to parameter choice for the GF-2 dataset.

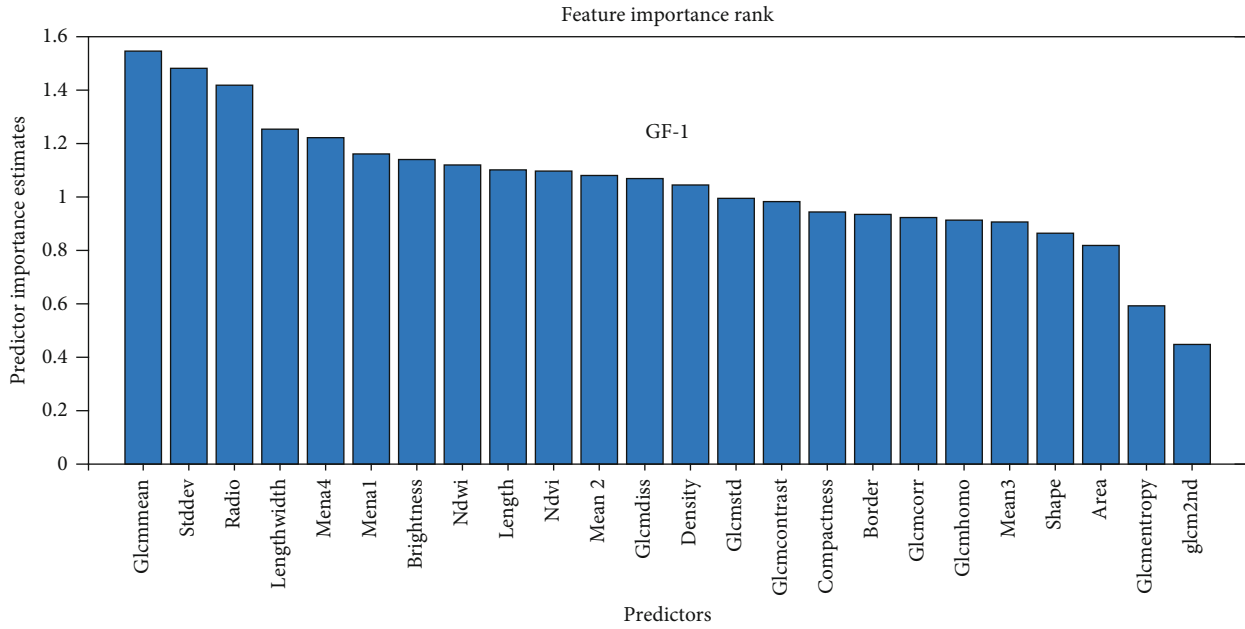


FIGURE 8: Feature importance rank for the GF-1 dataset.

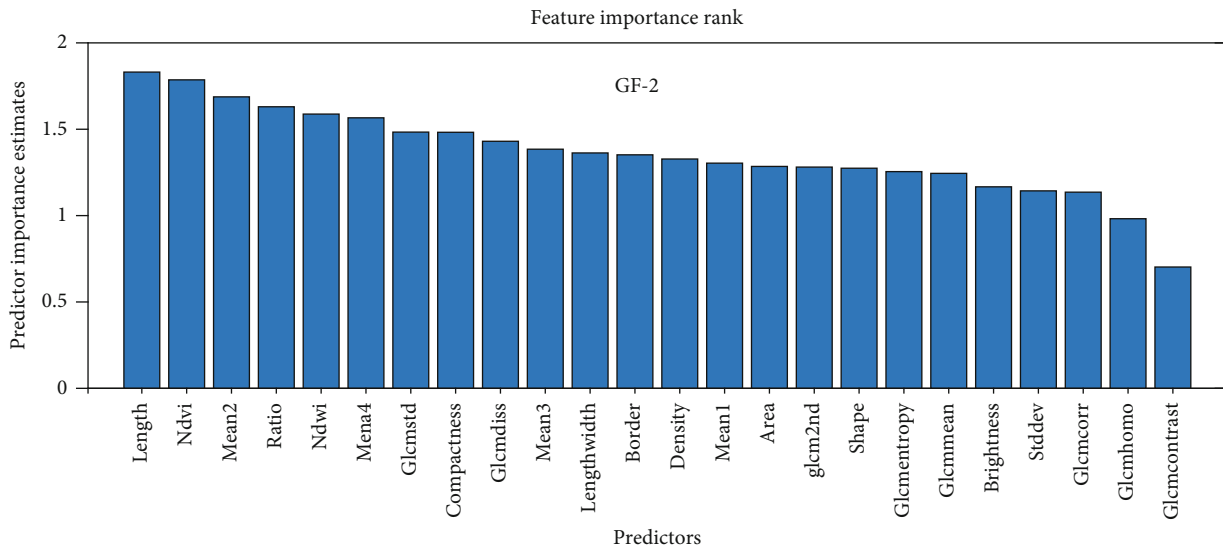


FIGURE 9: Feature importance rank for the GF-2 dataset.

scale parameter that defines the average size of the image object is considered to be the most effective parameter that affects the segmentation quality, while there is no universal rule for this scale determination. By literature review [46], the ESP2 (estimation of scale parameter) scale parameter estimation tool is introduced and combined with visual interpretation to evaluate the optimal scale value and segmentation effect of GF-1 and GF-2 remote sensing images in this research. Based on statistic results (Figure 10), the appearance of first peaks are 105 and 220 for GF-1 and GF-2, respectively, which are the optimal segmentation scales of GF-1 and GF-2 images in this study. The shape and compactness parameters have limited influence on the performance of OBIA, and they were set constant at 0.1 and

0.5, respectively. Therefore, in this research, we chose 0.1 and 0.5 to participate in the segmentation to get the final segmentation map (Figure 10).

Finally, to more explicitly evaluate the practical speed of proposed image classification chain compared to RF, SVM, and KNN, we consider empirical run times. In terms of image performance speed test, to ensure fair comparison, all methods shared the same code. The algorithms were deployed in the Henan Polytech High Performance Computing Center server, shown in Figure 3 (here, node is 1, and thread number of per CPU core per node is 24). Through rough statistic, OBIA-RF spent 223.55 s, SVM spent 403.57 s, KNN spent 86.93 s, and DTC spent 0.64 s for the GF-1 and GF-2 datasets in average. The model with the most

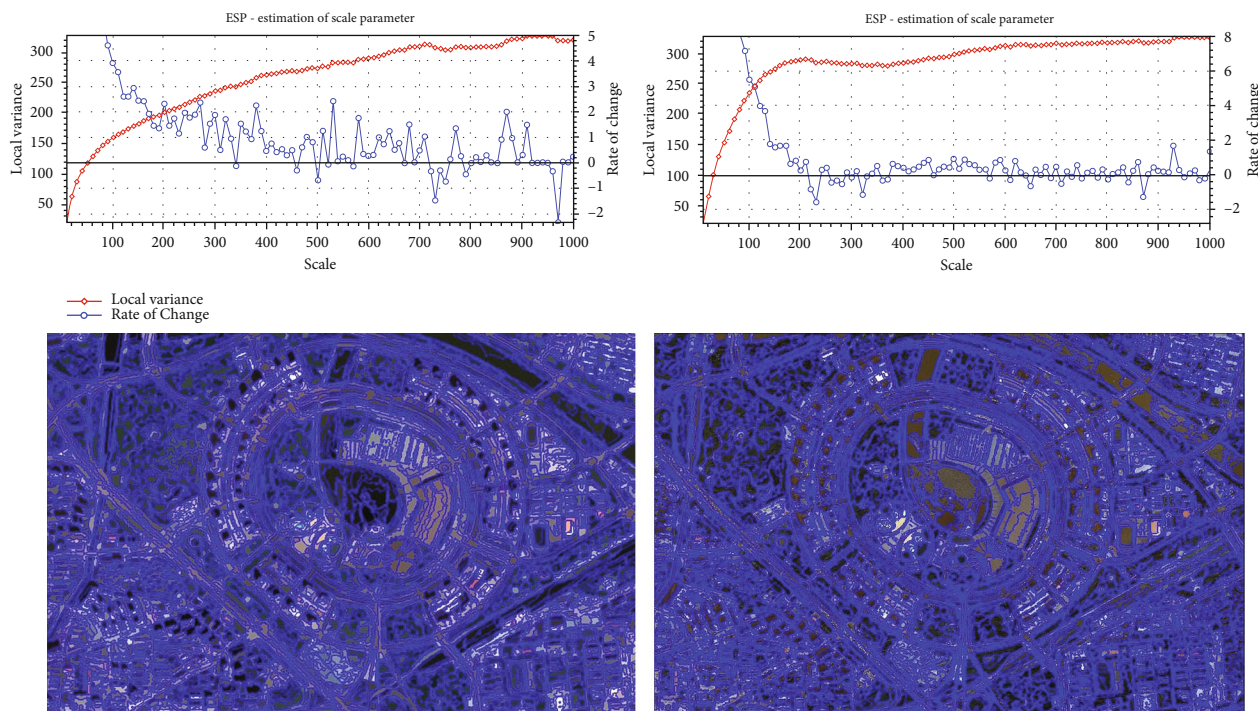


FIGURE 10: Segmentation results of GF-1 and GF-2 based on ESP.

time consumption is SVM, followed by KNN. The most time saving model is DTC, which got less than 70% classification accuracy.

5. Conclusion and Future Work

In this paper, we have proposed a novel urban mapping process chain which can take advantage of both the OBIA and classifier ensemble methods. The novelty in this paper is in the direction of successful evaluation of OBIA-RF on Chinese high-resolution satellite images GF-1 and GF-2 datasets. The performance of the OBIA-RF method has been compared with the state-of-art model such as DTC, SVM, and KNN, and performance of proposed OBIA-RF method has also been examined from urban mapping accuracy, the sensitivity to parametric selection, and time consumption. As the proposed process chain considered only spectral and GLCM texture features for image semantic, other features which might be useful to urban mapping such as local indicator of spatial association, mathematical morphology profiles, and decomposition characteristics of full polarized SAR features will be considered for future research.

Data Availability

The very high-resolution remotely sensed data used in this research is provided by Henan Data and Application Center of the High Resolution Earth Observation System through signing a contract with National Defense Science and Technology Bureau of Henan. Unfortunately, we do not have the priority to share the high-resolution satellite remote sensed data. Anyway, we can share our code used in this research. Researchers can test algorithms using these codes

with their own datasets and repeat experiments to obtain similar research conclusions. Researchers who are interested in this code can download it from hyperlink <https://pan.baidu.com/s/1hUyKWuoSzv2mrDx-NZ192w>, using password ku0w, or contact with the corresponding author.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Authors' Contributions

RM.H did the conceptualization. P.L. did the methodology. P.L. did the formal analysis. RM.H and HW. Z. wrote the manuscript. RM.H proposed the idea and wrote the original manuscript and the following revisions. P.L. provided the funding. RM.H, GY.W., and XL.W. executed all the experiments, XL.W, RM.H, and HW. Z contributed to the revisions and provided valuable comments. Ruimei Han, Pei Liu, Guangyan Wang, Hanwei Zhang, and Xilong Wu contributed equally to this work.

Acknowledgments

We would also like to thank the data provider of Henan Data and Application Center of the High Resolution Earth Observation System. This research was supported by grants from the National Natural Science Foundation of China (41601450), Natural Science Foundation of Hebei Province (grant number D20200409002), and Henan Key Technology R&D Projects (182102310860).



References

- [1] M. Ustuner, F. B. Sanli, and B. Dixon, "Application of support vector machines for landuse classification using high-resolution RapidEye images: a sensitivity analysis," *EUROPEAN JOURNAL OF REMOTE SENSING*, vol. 48, no. 1, pp. 403–422, 2017.
- [2] W. Zhao, S. du, and W. J. Emery, "Object-based convolutional neural network for high-resolution imagery classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 7, pp. 3386–3396, 2017.
- [3] C. Vasilakos, D. Kavroudakis, and A. Georganta, "Machine learning classification ensemble of multitemporal sentinel-2 images: the case of a mixed Mediterranean ecosystem," *REMOTE SENSING*, vol. 12, no. 12, p. 2005, 2020.
- [4] Y. Qian, W. Zhou, C. J. Nytych, L. Han, and Z. Li, "A new index to differentiate tree and grass based on high resolution image and object-based methods," *Urban Forestry and Urban Greening*, vol. 53, p. 126661, 2020.
- [5] P. Lynch, L. Blesius, and E. Hines, "Classification of urban area using multispectral indices for urban planning," *Remote Sensing*, vol. 12, no. 15, p. 2503, 2020.
- [6] Y. Xu, B. Du, L. Zhang et al., "Advanced multi-sensor optical remote sensing for urban land use and land cover classification: outcome of the 2018 IEEE GRSS data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2019, no. 12, pp. 1709–1724, 2019.
- [7] E. O. Yilmaz, B. Varol, R. H. Topaloglu, and E. Sertel, "Object-based classification of Izmir Metropolitan City by using Sentinel-2 images," in *2019 9th International Conference on Recent Advances in Space Technologies (RAST)*, pp. 407–412, Istanbul, Turkey, June 2019.
- [8] D. Sulla-Menashe, J. M. Gray, S. P. Abercrombie, and M. A. Friedl, "Hierarchical mapping of annual global land cover 2001 to present: the MODIS collection 6 land cover product," *Remote Sensing of Environment*, vol. 222, pp. 183–194, 2019.
- [9] L. H. Nguyen, D. R. Joshi, D. E. Clay, and G. M. Henebry, "Characterizing land cover land use from multiple years of Landsat and MODIS time series: a novel approach using land surface phenology modeling and random forest classifier," *Remote Sensing of Environment*, vol. 238, p. 111017, 2020.
- [10] L. B. Zeferino, L. F. T. de Souza, C. H. do Amaral, E. I. F. Filho, and T. S. de Oliveira, "Does environmental data increase the accuracy of land use and land cover classification?," *International Journal of Applied Earth Observation and Geoinformation*, vol. 91, p. 102128, 2020.
- [11] S. Q. Liu and N. Zhang, "Urban land use and land cover classification using multisource remote sensing images and social media data," *Remote Sensing*, vol. 11, no. 22, p. 2719, 2019.
- [12] I. Colkesen and T. Kavzoglu, "Ensemble-based canonical correlation forest (CCF) for land use and land cover classification using sentinel-2 and Landsat OLI imagery," *Remote Sensing Letters*, vol. 8, no. 11, pp. 1082–1091, 2017.
- [13] M. Turker and A. Ozdarici, "Field-based crop classification using SPOT4, SPOT5, IKONOS and QuickBird imagery for agricultural areas: a comparison study," *INTERNATIONAL JOURNAL OF REMOTE SENSING*, vol. 32, no. 24, pp. 9735–9768, 2011.
- [14] J. A. Benediktsson, J. Chanussot, and W. M. Moon, "Advances in very-high-resolution remote sensing [Scanning the Issue]," *PROCEEDINGS OF THE IEEE*, vol. 101, no. 3, pp. 566–569, 2013.
- [15] R. Momeni, P. Aplin, and D. Boyd, "Mapping complex urban land cover from spaceborne imagery: the influence of spatial resolution, spectral band set and classification approach," *REMOTE SENSING*, vol. 8, no. 2, p. 88, 2016.
- [16] Q. Song, Q. Hu, Q. Zhou et al., "In-season crop mapping with GF-1/WFV data by combining object-based image analysis and random forest," *REMOTE SENSING*, vol. 9, no. 11, p. 1184, 2017.
- [17] M. Amani, B. Salehi, S. Mahdavi, B. Brisco, and M. Shehata, "A multiple classifier system to improve mapping complex land covers: a case study of wetland classification using SAR data in Newfoundland, Canada," *INTERNATIONAL JOURNAL OF REMOTE SENSING*, vol. 39, no. 21, pp. 7370–7383, 2018.
- [18] L. Zhou and X. Yang, "Training algorithm performance for image classification by neural networks," *PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING*, vol. 76, no. 8, pp. 945–951, 2010.
- [19] G. Shukla, R. D. Garg, P. Kumar, H. S. Srivastava, and P. K. Garg, "Using multi-source data and decision tree classification in mapping vegetation diversity," *Spatial Information Research*, vol. 26, no. 5, pp. 573–585, 2018.
- [20] F. Wang, Q. Wang, F. Nie, Z. Li, W. Yu, and F. Ren, "A linear multivariate binary decision tree classifier based on K-means splitting," *PATTERN RECOGNITION*, vol. 107, p. 107521, 2020.
- [21] H. Shen, Y. Lin, Q. Tian, K. Xu, and J. Jiao, "A comparison of multiple classifier combinations using different voting-weights for remote sensing image classification," *INTERNATIONAL JOURNAL OF REMOTE SENSING*, vol. 39, no. 11, pp. 3705–3722, 2018.
- [22] G. Lei, A. Li, J. Bian et al., "OIC-MCE: a practical land cover mapping approach for limited samples based on multiple classifier ensemble and iterative classification," *REMOTE SENSING*, vol. 12, no. 6, p. 987, 2020.
- [23] G. J. Briem, J. A. Benediktsson, and J. R. Sveinsson, "Multiple classifiers applied to multisource remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2291–2299, 2002.
- [24] P. Du, J. Xia, W. Zhang, K. Tan, Y. Liu, and S. Liu, "Multiple classifier system for remote sensing image classification: a review," *SENSORS*, vol. 12, no. 4, pp. 4764–4792, 2012.
- [25] M. Xia, N. Tian, Y. Zhang, Y. Xu, and X. Zhang, "Dilated multi-scale cascade forest for satellite image classification," *INTERNATIONAL JOURNAL OF REMOTE SENSING*, vol. 41, no. 20, pp. 7779–7800, 2020.
- [26] P. Thanh Noi and M. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery," *SENSORS*, vol. 18, no. 2, p. 18, 2018.
- [27] C. Pelletier, S. Valero, J. Inglada, N. Champion, and G. Dedieu, "Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas," *REMOTE SENSING OF ENVIRONMENT*, vol. 187, pp. 156–168, 2016.
- [28] M. Belgiu and L. Dragut, "Random forest in remote sensing: a review of applications and future directions," *ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING*, vol. 114, pp. 24–31, 2016.
- [29] A. Zafari, R. Zurita-Milla, and E. Izquierdo-Verdiguier, "Evaluating the performance of a random forest kernel for land

- cover classification," *REMOTE SENSING*, vol. 11, no. 5, p. 575, 2019.
- [30] D. Phiri, M. Simwanda, V. Nyirenda, Y. Murayama, and M. Ranagalage, "Decision tree algorithms for developing rule-sets for object-based land cover classification," *ISPRS INTERNATIONAL JOURNAL OF GEO-INFORMATION*, vol. 9, no. 5, p. 329, 2020.
- [31] A. Zafari, R. Zurita-Milla, and E. Izquierdo-Verdiguier, "A multiscale random forest kernel for land cover classification," *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING*, vol. 13, pp. 2842–2852, 2020.
- [32] M. D. Hossain and D. Chen, "Segmentation for object-based image analysis (OBIA): a review of algorithms and challenges from remote sensing perspective," *ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING*, vol. 150, pp. 115–134, 2019.
- [33] M. Kucharczyk, G. J. Hay, S. Ghaffarian, and C. H. Hugenholtz, "Geographic object-based image analysis: a primer and future directions," *REMOTE SENSING*, vol. 12, no. 12, p. 2012, 2020.
- [34] C. M. D. de Pinho, L. M. G. Fonseca, T. S. Korting, C. M. de Almeida, and H. J. H. Kux, "Land-cover classification of an intra-urban environment using high-resolution images and object-based image analysis," *International Journal of Remote Sensing*, vol. 33, no. 19, pp. 5973–5995, 2012.
- [35] T. Su, T. Liu, S. Zhang, Z. Qu, and R. Li, "Machine learning-assisted region merging for remote sensing image segmentation," *ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING*, vol. 168, pp. 89–123, 2020.
- [36] V. S. Martins, A. L. Kaleita, B. K. Gelder, H. L. F. da Silveira, and C. A. Abe, "Exploring multiscale object-based convolutional neural network (multi-OCNN) for remote sensing image classification at high spatial resolution," *ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING*, vol. 168, pp. 56–73, 2020.
- [37] S. W. Myint, P. Gober, A. Brazel, S. Grossman-Clarke, and Q. Weng, "Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery," *Remote Sensing of Environment*, vol. 115, no. 5, pp. 1145–1161, 2011.
- [38] D. C. Duro, S. E. Franklin, and M. G. Dube, "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery," *REMOTE SENSING OF ENVIRONMENT*, vol. 118, pp. 259–272, 2012.
- [39] B. Fu, Y. Wang, A. Campbell et al., "Comparison of object-based and pixel-based Random Forest algorithm for wetland vegetation mapping using high spatial resolution GF-1 and SAR data," *Ecological Indicators*, vol. 73, pp. 105–117, 2017.
- [40] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING*, vol. 65, no. 1, pp. 2–16, 2010.
- [41] H. Grybas, L. Melendy, and R. G. Congalton, "A comparison of unsupervised segmentation parameter optimization approaches using moderate- and high-resolution imagery," *GISCIENCE & REMOTE SENSING*, vol. 54, no. 4, pp. 515–533, 2017.
- [42] Y. Shen, J. Chen, L. Xiao, and D. Pan, "Optimizing multiscale segmentation with local spectral heterogeneity measure for high resolution remote sensing images," *ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING*, vol. 157, pp. 13–25, 2019.
- [43] S. Georganos, M. Lennert, T. Grippa, S. Vanhuyse, B. Johnson, and E. Wolff, "Normalization in unsupervised segmentation parameter optimization: a solution based on local regression trend analysis," *REMOTE SENSING*, vol. 10, no. 2, p. 222, 2018.
- [44] L. Dragut, D. Tiede, and S. R. Levick, "ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data," *INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE*, vol. 24, no. 6, pp. 859–871, 2010.
- [45] J. Liu, M. Du, and Z. Mao, "Scale computation on high spatial resolution remotely sensed imagery multi-scale segmentation," *INTERNATIONAL JOURNAL OF REMOTE SENSING*, vol. 38, no. 18, pp. 5186–5214, 2017.
- [46] J. Liu, H. Pu, S. Song, and M. Du, "An adaptive scale estimating method of multiscale image segmentation based on vector edge and spectral statistics information," *INTERNATIONAL JOURNAL OF REMOTE SENSING*, vol. 39, no. 20, pp. 6826–6845, 2018.
- [47] L. Dragut, O. Csillik, C. Eisank, and D. Tiede, "Automated parameterisation for multi-scale image segmentation on multiple layers," *ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING*, vol. 88, no. 100, pp. 119–127, 2014.
- [48] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [49] C. Su, S. Ju, Y. Liu, and Z. Yu, "Improving random forest and rotation forest for highly imbalanced datasets," *INTELLIGENT DATA ANALYSIS*, vol. 19, no. 6, pp. 1409–1432, 2015.
- [50] H. Deng and G. Runger, "Gene selection with guided regularized random forest," *PATTERN RECOGNITION*, vol. 46, no. 12, pp. 3483–3489, 2013.
- [51] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *EXPERT SYSTEMS WITH APPLICATIONS*, vol. 134, pp. 93–101, 2019.
- [52] P. Lou, B. Fu, H. He et al., "An optimized object-based random forest algorithm for marsh vegetation mapping using high-spatial-resolution GF-1 and ZY-3 data," *REMOTE SENSING*, vol. 12, no. 8, p. 1270, 2020.
- [53] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: a review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247–259, 2011.
- [54] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sensing of Environment*, vol. 37, no. 1, pp. 35–46, 1991.
- [55] G. M. Foody, "Status of land cover classification accuracy assessment," *Remote Sensing of Environment*, vol. 80, no. 1, pp. 185–201, 2002.

Research Article

On Time-Series InSAR by SA-SVR Algorithm: Prediction and Analysis of Mining Subsidence

Yun Shi ^{1,2}, Qianwen Li,^{1,2} Xin Meng,³ Tongkang Zhang,^{1,2} and Jingjian Shi ^{1,2}

¹School of Geomatics, Xi'an University of Science and Technology, Xi'an, Shaanxi 710054, China

²Key Laboratory of Coal Resources Exploration and Comprehensive Utilization, Ministry of Natural Resources, Xi'an, Shaanxi 710021, China

³Xi'an Research Institute of Surveying and Mapping, Xi'an 710054, China

Correspondence should be addressed to Jingjian Shi; 346870934@qq.com

Received 31 July 2020; Revised 30 September 2020; Accepted 9 October 2020; Published 4 November 2020

Academic Editor: Hyung-Sup Jung

Copyright © 2020 Yun Shi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Given the increasingly serious geological disasters caused by underground mining in the Hancheng mining area in China and the existing problems with mining subsidence prediction models, this article uses the small baseline subset interferometric synthetic aperture radar (SBAS-InSAR) technology to process 109 Sentinel-1A images of this mining area from December 2015 to February 2020. The results show that there are three subsidences: one in Donganshang, one in south of Zhuyuan village, and one in Shandizhaizi village. In the basin, the maximum annual average subsidence rate is 300 mm/a, and the maximum cumulative subsidence is 1000 mm. The SBAS-InSAR results are compared with Global Positioning System (GPS) observation results, and the correlation coefficient is 74%. Finally, a simulated annealing (SA) algorithm is used to estimate the optimal parameters of a support vector regression (SVR) prediction model, which is applied for mining subsidence prediction. The prediction results are compared with the results of SVR and the GM (1, 1). The minimum value of the coefficient of determination for prediction with SA-SVR model is 0.57, which is significantly better than that those of the other two prediction methods. The results indicate that the proposed prediction model offers high subsidence prediction accuracy and fully meets the requirements of engineering applications.

1. Introduction

Ground subsidence in a mining area is a type of vertical deformation of the ground, which is prone to slow regional changes due to the destruction of the structure of the rock mass caused by mining. Long-term underground mining results in movement and deformation of the overlying rock layer and the ground surface and can cause continuous or discontinuous sinking, tilting, curvature, stretching, compression, etc., of buildings and other structures located within the mining area. Deformation, cracks, collapses, collapse pits, landslides (movement), and similar types of damage degrade the land resources and ecological environment in a mining area, significantly impacting the productivity and daily lives of residents and also affect future engineering construction in the area [1]. In a word, the ground subsidence caused by mining is a very complicated process. To effectively

control the ground subsidence in mining areas caused by mining and to reduce losses, it is necessary to accurately understand the causes of mining subsidence and its development process over time.

Traditional levelling and Global Positioning System (GPS) measurements have the disadvantages of sparse measurement points, a small range, and a high cost for monitoring surface subsidence. As an alternative, differential interferometric synthetic aperture radar (D-InSAR) has been proven to be able to probe small surface deformations [2]; it is widely used in the monitoring of landslides [3], urban areas [4], mining area ground subsidence [5], and other geological disasters, but it is susceptible to limitations such as decoherence and atmospheric delay, and it cannot provide continuous time-series deformation information. As a sequential InSAR technology, small baseline subset InSAR (SBAS-InSAR) mitigates the shortcomings of D-InSAR technology

by adopting a short time-space baseline [6–8]. Zhang et al. [9, 10], Hu et al. [11], and Jiang et al. [12] have used time-series InSAR technology to investigate land subsidence in Los Angeles, California, USA; Ningbo, China; and the Lost Hills Oil Field in California. Yin et al. [13], Yang et al. [14], Saygin et al. [15], and Mark et al. [16] have applied time-series InSAR technology for mining area monitoring in Lengshuijiang, China; Datong, China; Zonguldak Province, Turkey; and Springfield, Illinois, USA, and have shown the applicability of SBAS-InSAR technology in mining surface subsidence monitoring applications. He et al. [17] used ALOS-1 data to identify large-scale surface deformations near China's Hancheng coal mine. Through a literature review, it can be found that most applications of InSAR technology in mining subsidence monitoring have focused on data monitoring and analysis, and the research time span has been relatively short. However, for mining subsidence monitoring, it is necessary to combine monitoring and prediction methods to establish a complete mining subsidence prediction model with long time-series capabilities.

The current mining subsidence prediction methods can be divided into four categories: influence function methods, empirical methods, theoretical methods, and other methods [18, 19]. The most commonly used influence function methods are the probability integral method in China and the Budryk-Knothe method elsewhere. The formulas for these two methods are the same. By adjusting the prediction parameters, this method can be applied for prediction under different geological and mining conditions, but the expected values of the parameters must be known and are difficult to obtain. Empirical methods include method based on typical curves, profile functions, and other approaches. A large amount of measured data is used to determine the laws governing the observed deformation. The prediction formulas are simple, but this approach is suitable only for areas with similar geological mining conditions, so its applicability is poor. Theoretical methods rely on complex mechanical parameters and calculations. Previous studies have shown that such methods are very difficult to apply. With the rapid development of computer technology, a large number of complex calculations have become possible, but it is still difficult to obtain mechanical parameters on a large scale. Other methods include neural network backpropagation (BP) [20], grey system theory (GM) [21], and support vector regression (SVR) [22, 23]. Researchers use computer algorithms to establish prediction models to predict mining subsidence in mining areas. Various new ideas are emerging in the field of mining subsidence prediction, but they are still in the exploration stage. Although the BP algorithm considers the nonlinearity of ground subsidence, when few training samples are available, the prediction results obtained are unreliable. The GM algorithm transforms messy original data into regular time-series data in a prescribed way. It is suitable only for medium- and short-term forecasts, and some data fitting results are far from the original data. In this paper, SBAS-InSAR technology is combined with the SA-SVR algorithm to predict mining subsidence for the first time. The SA-SVR algorithm combines SA and SVR. The SVR algorithm can handle any nonlinear situation, has a strong

generalization ability and good theoretical support, and has been widely used in research on the regression fitting of nonlinear functions [24].

This article is organized into five parts. The first part summarizes the current research progress and shortcomings of mining subsidence monitoring and prediction. The second part briefly describes the basic idea of the combined SA-SVR algorithm. The third part presents the extraction of the 4-year average annual deformation rate and the time-series cumulative deformation of a mining area in Hancheng based on SBAS-InSAR technology, analyses the overall deformation behaviour in the mining area before and after mining, provides a decision basis for the subsequent mining management of the mining area, and verifies the SBAS-InSAR monitoring results against GPS data. The fourth part reports the prediction results achieved by using the SBAS-InSAR monitoring results as the training and test sets for the SA-SVR algorithm and finally compares the SA-SVR prediction results with those of SVR and GM. The fifth part summarizes the whole paper and discusses the advantages and disadvantages of the SA-SVR algorithm.

2. Basic Idea of the SA-SVR Algorithm

The SVR algorithm performs parameter regression estimation on the basis of a support vector machine (SVM). This algorithm is suitable for low-information, linear and nonlinear models and for the recognition of high-dimensional spatial patterns from small samples. Research has shown that the SVR algorithm enables stable time-series analysis and statistical forecasting [25]. The basic principle of the method is to find the optimal division plane in a high-dimensional space and to minimize the sum of the energy of the regression coefficients under the constraint that the error of the given sample data with respect to the division boundary is within a certain value. The original SVR algorithm can also be used as a regularization method to balance the error when fitting sample data and the energy of the regression coefficients.

Time-series InSAR technology is used to obtain time-series data on mining area settlement, denoted by $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2) \cdots (\mathbf{x}_n, y_n)\}$, $\mathbf{x}_i \in X = \mathbb{R}^n$, $y_i \in Y = \mathbb{R}$. To study the time-space relationship of settlement in a mining area and use it to make settlement predictions, we need to analyse (\mathbf{x}_i, y_i) ($i = 1, 2, \dots, n$) quantitatively. More specifically, we hope to find a suitable mapping Φ to describe the relationship between \mathbf{x}_i and y_i and use SVR for data learning and prediction. Under the assumption that the mapping Φ between (x, y) follows an approximately linear relationship, we use the following formula:

$$y = \langle \omega, \mathbf{x} \rangle + b \quad (1)$$

where $\langle \cdot, \cdot \rangle$ represents the vector dot product, x is the value of the input variable, and y is the corresponding output value. According to Smola and Scholkopf [26], the SVR algorithm can be applied to solve the following optimization problem:

$$\begin{aligned}
\min : & \quad \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\
& \quad y_i - \langle \boldsymbol{\omega}, \mathbf{x}_i \rangle - b \leq \varepsilon_i + \xi_i \\
s.t. & \quad \langle \boldsymbol{\omega}, \mathbf{x}_i \rangle + b - y_i \leq \varepsilon_i + \xi_i^* \\
& \quad \xi_i, \xi_i^* \geq 0
\end{aligned} \tag{2}$$

where C is a penalty or equilibrium factor and ξ_i, ξ_i^* are the upper and lower bounds of the regression line, respectively, where an excess violation of ε_i is allowed.

The parameters of the SVR model include the accuracy ε_i describing the regression model and the data column, the penalty factor C , and the mapping Φ . The kernel function $K(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle$ can be obtained from this model. These parameters and the kernel function together determine the overall performance of the prediction model. From the constraint conditions of the optimization model given in (2), it can be seen that when the difference between the theoretical value and the actual observed settlement value is less than ε_i , SVR will not penalize such a violation; that is, the corresponding values of ξ_i, ξ_i^* are zero. Therefore, SVR is also called ε -insensitive regression parameter estimation. Generally, if the observed settlement values are the same in terms of accuracy, all ε_i have the same value ε . If the accuracy value ε_i is too small, that is, the accuracy of the data column is too high, then under-fitting or over-fitting can easily occur. Smola and Scholkopf [26] defined the allowable kernel function and presented a detailed discussion of the structure of the allowable kernel function problem. For details, please see Smola and Scholkopf [26].

In the literature on SVR, to solve for the optimal parameters of the optimization problem defined in (2), the original problem is generally converted into the corresponding dual problem using the Lagrange multiplier method, namely,

$$\begin{aligned}
\min : & \quad \frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) + \sum_{i=1}^n \varepsilon_i (\alpha_i + \alpha_i^*) - \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \\
s.t. & \quad \langle (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*), \boldsymbol{\varepsilon} \rangle = 0 \\
& \quad \alpha_i, \alpha_i^* \in [0, C]
\end{aligned} \tag{3}$$

In SVMs, sequential algorithms are usually used to solve for the optimal coefficients $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}^*$ of such a dual problem (see [26]). $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}^*$ essentially correspond to the Lagrangian coefficients of the first and second constraints, respectively, of the optimization problem given in (2). It can be proven (Smola and Scholkopf [26]) that the estimate $\hat{\boldsymbol{\omega}}$ of the unknown parameter in the original regression model defined in (1) is uniquely determined by the Lagrangian multiplication coefficients $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}^*$:

$$\hat{\boldsymbol{\omega}} = \mathbf{X} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) \tag{4}$$

The unknown parameter b is then given by the Karush-Kuhn-Tucker condition [27, 28]. Thus, the optimal prediction model is

$$\hat{y} = \sum_{i=1}^n (\hat{\alpha}_i - \alpha_i^*) \langle \mathbf{x}_i, \mathbf{x} \rangle + \hat{b} \tag{5}$$

In fact, by comparing the original optimization problem (2) and its corresponding dual problem (3), it can be seen that the original optimization problem (2) has more parameters ($\boldsymbol{\omega}$ and b) than the dual problem (3). Nevertheless, if the number of unknown parameters in the original regression model (1) is low, then the number of calculations will not be very different between the original optimization problem (2) and its dual problem (3).

In this paper, an algorithm that combines SA with SVR is adopted to achieve the best prediction effect for mining subsidence. The SA algorithm was proposed by Metropolis et al. in 1953 [27] and started to be widely used after the publication of an article in *Science* by Kirkpatrick et al. in 1983. In the SA algorithm, random factors are introduced into the search process. In the iterative update process, a solution that is worse than the current solution may be accepted with a certain small probability, allowing the algorithm to jump out of locally optimal solutions to reach a better optimal solution. In this paper, an SA optimization algorithm is used in combination with SVR to estimate the parameters of the time-series forecasting model.

The basic principle of the SA algorithm is to imitate the ideal crystal generation process in physics. It is assumed that at a high temperature, solid matter will be in a liquid state, in which the order of the molecules is free and random. Through a careful cooling process, the molecules in this random state will be arranged in an orderly manner to generate crystals in the ideal state. To use an optimized mathematical language to describe the process of forming an ideal crystal state through physical cooling, we need to formulate the ideal crystal state and the physical cooling process in terms of an optimized objective function and an optimization process, respectively. Specifically, the ideal crystal state is the minimum-energy state, that is, the state with the global optimal value of the objective function, and the physical cooling process is equivalent to the iterative process of optimizing the objective function, thus completing the mathematical simulation ideal crystal generation. In particular, Metropolis et al.'s original text from 1953 proposes random sampling from a uniform probability distribution as an iterative process for optimization. If the objective function is fully understood, random sampling from the corresponding probability distribution can instead be performed for optimization (see Peiliang Xu's private message from February 2020) [29].

The basic step of the optimization process in the SA algorithm is to compare the objective function value at a sampled point to the current optimal value. If the former is better than the latter, then the objective function value of the sampled point replaces the current optimal value, and the algorithm enters the next iteration. If the objective function value at the sampled point is worse than the current optimal value,

to prevent the SA algorithm from converging to a locally optimal value too quickly, Metropolis and others have suggested that this worse solution should nevertheless be accepted with a certain probability during the optimization process. For an optimization function $f(x)$, if the current solution is denoted by x_0 and the search point for the next round is denoted by x' , Metropolis and others suggest using the following probability:

$$P(x_0 \Rightarrow x') = \begin{cases} 1, & f(x') < f(x_0) \\ \exp\left(-\frac{f(x') - f(x_0)}{T}\right), & f(x') \geq f(x_0) \end{cases} \quad (6)$$

Accordingly, we decide whether to accept x' as the new solution. In formula (6), T represents the temperature variable, and \Rightarrow represents a state transition.

The SA algorithm obtains the optimal solution through repeated iterations of the above process. However, if T changes too quickly, then the SA algorithm will quickly converge to a locally optimal solution. To ensure that the algorithm will converge in a limited time, it is necessary to adjust the temperature parameter T . The steps of the adjustment process are as follows:

- (i) The initial temperature $T(0)$ is set high enough that all states are acceptable
- (ii) The annealing rate generally decreases exponentially

$$T(n) = \lambda T(n-1), n = 1, 2, 3 \dots \quad (7)$$

Here, λ takes a value in the range [0.8, 0.99] so that each temperature will have a certain probability to be tried.

- (iii) A termination temperature is specified. Once the temperature reaches this preset threshold during the iterative process, annealing is completed

The flow chart of the SA-SVR algorithm is shown in Figure 1.

3. Time-Series InSAR Monitoring and Analysis

3.1. Research Background. The study area is a mine in the southwest of the Hancheng mining area, which is located on the southeastern edge of the Weibei Uplift in the Ordos Basin. The formation is generally a monoclinic structure inclined to the northwest. The geographical location is $110^{\circ}17' - 110^{\circ}36'$ east in longitude and $35^{\circ}21' - 35^{\circ}51'$ north in latitude. The mine is adjacent to the Dongze village structural belt and the Long-ting structural belt in the west and to the Huaxian fault zone in Hancheng and the Yellow River in the east. The geographical location of the mine is shown in Figure 2 [30, 31]. The main coal-bearing strata in the mining area belong to the Shanxi Formation and the Taiyuan Forma-

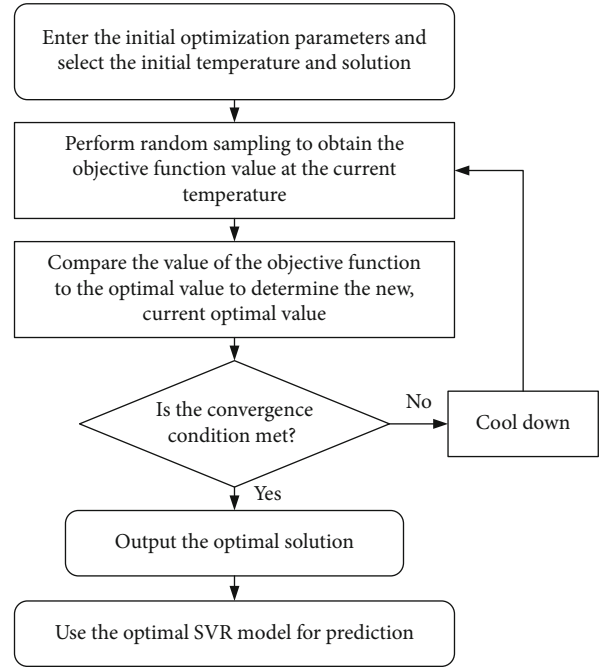


FIGURE 1: SA-SVR algorithm flow chart.

tion; coal seam No. 3 belongs to the Shanxi Formation, and coal seam No. 5 belongs to the Taiyuan Formation. The long-arm comprehensive mechanized coal mining method or the caving mining method has been adopted at all working faces, and the caving method is used to manage the roof. The mining thickness of each working face is 1.5–2.2 m, and the coal seam inclination is $1-3^{\circ}$, being nearly horizontal. This article mainly studies data taken from the 21308 working face of the No. 3 coal seam from December 2015 to February 2020. The strike length is 930 m, and the incline length is 210 m.

An investigation performed in May 2017 showed that only the Donganshang group exhibited multiple cracks in houses and 3 cracks in the road pavement, with a crack length of approximately 3 m, a crack width of approximately 0.05–0.1 m, and an angle between the crack and the road surface of approximately 90° , which caused the ground to form a bulge of 0.1–0.2 m. There were 3 large-scale cracks in the ground, which were approximately 500 m long and 0.2–0.6 m wide, and 3 small-scale cracks, which were approximately 40–80 m long and 0.1–0.3 m wide. In the village, there were 7 landslides, with lengths of 8–15 m, widths of 5–12 m, and drops of 2–5 m.

As research data, 109 scenes of Sentinel-1A radar satellite data in the study area were selected, which cover a time span from December 24, 2015, to February 25, 2020. The Sentinel-1A data are in the C-band, with a wavelength of 5.66 mm; the orbit revisit period is 12 days, the polarization mode is VV, the incidence angle is approximately 33.7° , the resolution of these data is 5×20 m, and the width of a single image is 250 km. The data used in this experiment are all products of the interferometric wide-width mode (IW) and are single-view complex data with slant distance information [32].

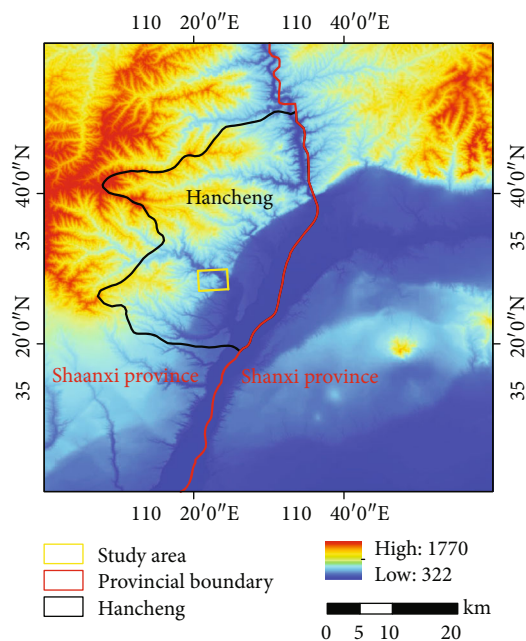


FIGURE 2: Geographical location of the study area.

The SRTM1 digital elevation model (DEM) was chosen as a source of external DEM data; its ground resolution is 30×30 m. For the precise orbital parameters, the precise orbit determination ephemeris data provided by the European Space Agency (ESA) (<https://qc.sentinel1.eo.esa.int/>) were adopted, which can be downloaded 21 days after acquisition of the Sentinel-1A data and have a positioning accuracy of within 5 mm [33]. External DEM and orbital data can be used to remove the levelling effect and orbital errors.

3.2. Data Processing. The SBAS-InSAR data process was realized based on the GAMMA Software. The main steps were as follows: extracting public bursts from the 109 Sentinel-1A scenes to improve the subsequent processing speed and efficiency, performing precise registration with an accuracy of 0.001 pixels, and setting a short time baseline (three consecutive scene images). The research area was cropped out based on the corresponding row and column numbers, and multi-view interference processing was performed at a visual number ratio of 5:1 (range: azimuth). Differential interference, filtering (Goldstein), and unwinding (minimum cost flow method) were performed based on the initial baseline and the precise baseline. The high coherence coefficient method was used to extract points with high coherence, and elevation and atmospheric errors were removed through regression analysis of the high-coherence points. Finally, complete and accurate deformation rate and time-series cumulative shape variables were obtained [34–36].

3.3. SBAS-InSAR Results and Analysis. Following the above data processing method, the annual average subsidence rate (Figure 3) and the time-series cumulative subsidence map (Figure 4) along the line-of-sight (LOS) direction were obtained for the mining area of Hancheng from December 2015 to February 2020. Figure 4 shows the timing accumula-

tion of 11 scenes selected from among all 109 scenes at equal intervals. Figure 3 shows that the entire study area contains three relatively large subsidence areas: the Donganshang subsidence area, Zhuyuan village south subsidence area, and the Shandizhaizi subsidence area. A field investigation showed that there is good consistency between the subsidence area on the Donganshang and the distribution of the working face in the south No. 1 mining area in Hancheng. Both the subsidence area to the south of Zhuyuan village and the Shandizhaizi subsidence area have undergone coal seam mining, and mountain slippage, road cracks, and house cracks have occurred.

The results presented in Figures 3 and 4 show that the maximum annual average settlement rate in the study area is 300 mm/a and the maximum cumulative settlement is 1000 mm from December 2015 to February 2020. The area on the Donganshang presents the most serious subsidence, with a subsidence area of approximately 3.66 km^2 and a cumulative subsidence of approximately 150–1000 mm. All of the working faces in this mining area show different degrees of subsidence. The 21506, 21307, and 21308 working faces stopped being mined in February 2017, and their subsidence rates are approximately 40–295 mm/a. The 21309 working face corresponds to the area with the most severe deformation during the study period. As of February 2020, the settlement rate of the 21309 mining face is the highest. The settlement area to the south of Zhuyuan village is approximately 2.56 km^2 , the settlement rate is approximately 40–257 mm/a, and the cumulative settlement is approximately 100–1000 mm. The Shandizhaizi settlement area is approximately 0.68 km^2 , the settlement rate is approximately 40–129 mm/a, and the cumulative settlement is approximately 100–570 mm.

It can be seen from Figure 4 that each mining face in the mining area shows a different degree of nonlinear settlement. Let the first image (December 24, 2015) be considered the initial image, in which the settlement amount is 0. The three settlement areas had formed by February 16, 2017. Over time and with the mining of the working faces, the subsidence area and the cumulative subsidence continued to increase. As of November 2, 2018, the subsidence area had gradually expanded westward, forming three obvious subsidence basins. As of February 25, 2020, the cumulative settlement had reached its maximum; the impact of the settlement had continued to expand from east to west, and the settlement area and cumulative settlement had gradually increased, indicating that the settlement area had not reached a stable state by that time.

To further analyse the settlement information of the study area, the 21308 working face in the settlement area on the Donganshang is taken as the object of investigation, and profile lines A1A2 and B1B2 are established along the direction and inclination (the position of the section line is shown in Figure 4). The working face trend is advancing along the direction from A2 to A1, and the working face tendency is advancing along the direction from B2 to B1. For image readability, the data from 10 scene images corresponding to some of the time-series cumulative settlement figures were selected to draw the corresponding strike and

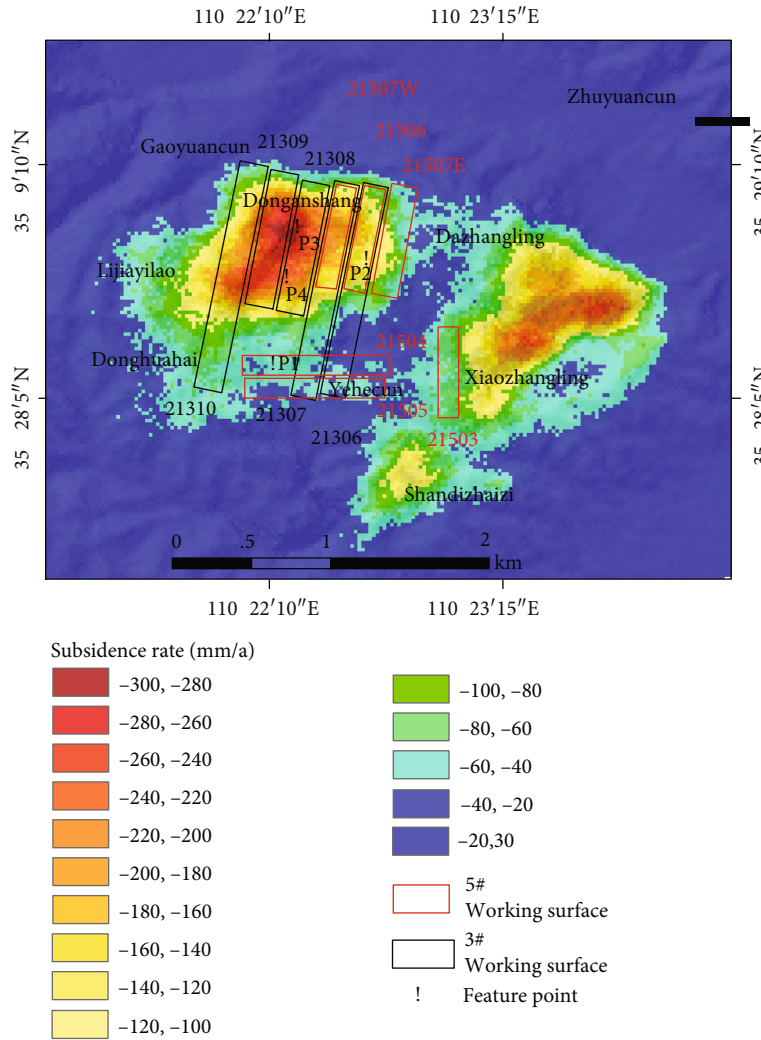


FIGURE 3: Average subsidence rates from December 2015 to February 2020.

longitudinal time-series profile lines using the Origin software (Figure 5). Figure 5 shows that the main settlement areas are between points 11 and 41, while the maximum sinking position is always at point 26. The maximum accumulated settlement amounts to 748 mm and 880 mm. In general, over time, the settlement of each point along the direction and inclination of the working surface is gradually increasing, and an obvious settlement funnel is forming on the spatial scale, which is consistent with the surface mining subsidence in the mining area.

Because this certain mining area in Hancheng has been repeatedly mined at multiple coal seams, to better analyse the pattern of the mining subsidence, four characteristic points (the positions of these characteristic points are shown in Figure 3) on the Donganshang were selected for time-series cumulative settlement analysis, as shown in Figure 6. P1 and P2 are located at the edge of the deformation area on the Donganshang. They are affected by the mining of the 21504 and 21506 working faces, respectively. The cumulative settlement at points P1 and P2 from December 24, 2015, to February 25, 2020, amounted to 255 mm and 392 mm, respectively. P3 and P4 are located in the centre of

the settlement area and are mainly affected by mining at the 21308 working face. The cumulative settlement at these points from December 24, 2015, to February 25, 2020, was relatively large, reaching 888 mm at P3 and 838 mm at P4. In general, the cumulative settlement at these four characteristic points has increased over time. As of February 25, 2020, the settlement had not stabilized.

3.4. Result Verification and Analysis. To quantitatively verify the ground subsidence monitoring results for the mining area as obtained via SBAS-InSAR technology, GPS observation points established in the subsidence area on the Donganshang were used to obtain ground displacement monitoring data, and the monitoring results were compared and analysed. GPS observation data collected from December 24, 2015, to January 12, 2017, at 8 observation points were selected. Because the surface deformation data obtained via SBAS-InSAR were obtained along the radar LOS direction, the three-dimensional deformation data obtained from the GPS observations were obtained in the vertical, east-west, and north-south directions. To compare the results of these two methods, three steps were required. First, it was

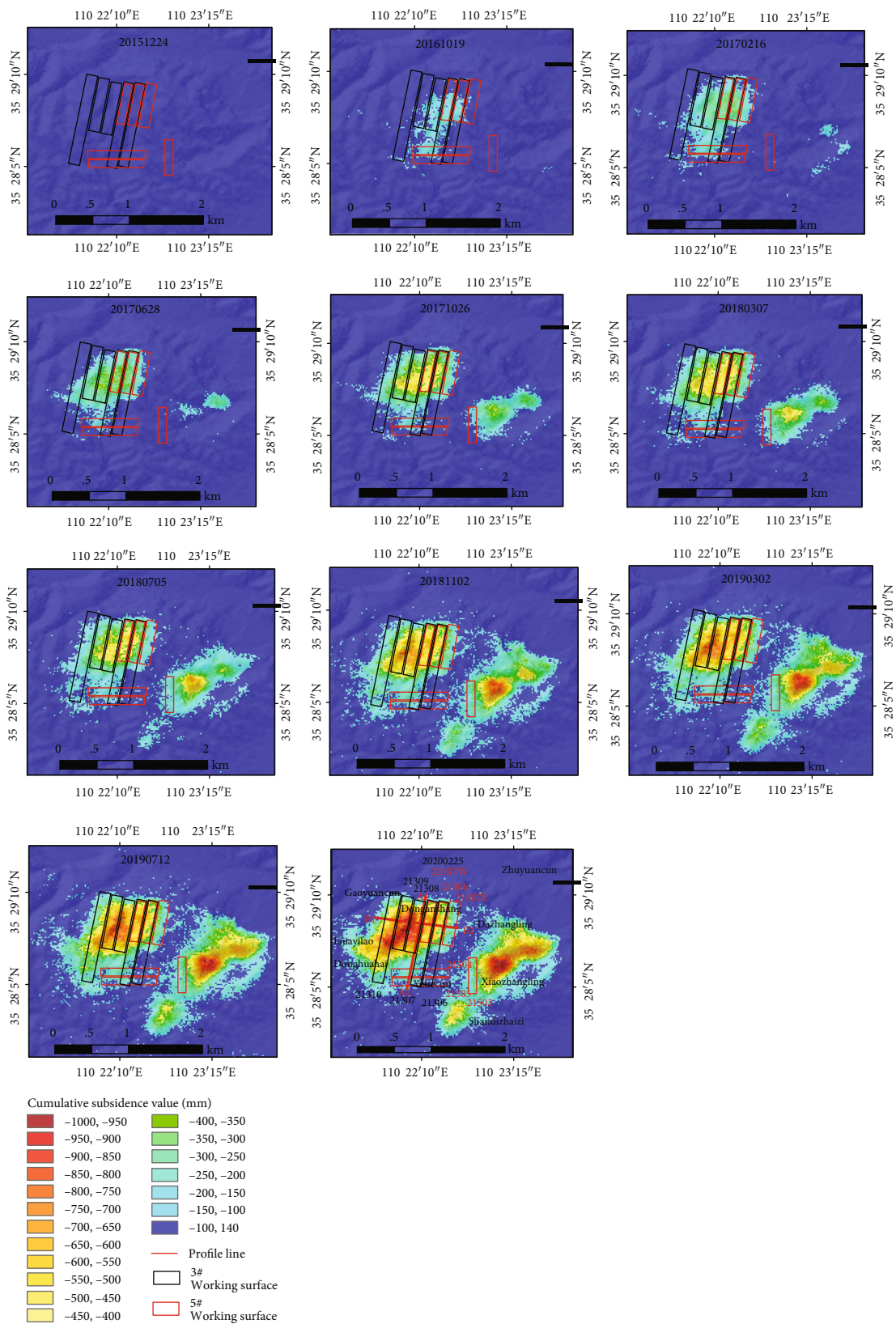


FIGURE 4: Time-series cumulative subsidence from December 2015 to February 2020.

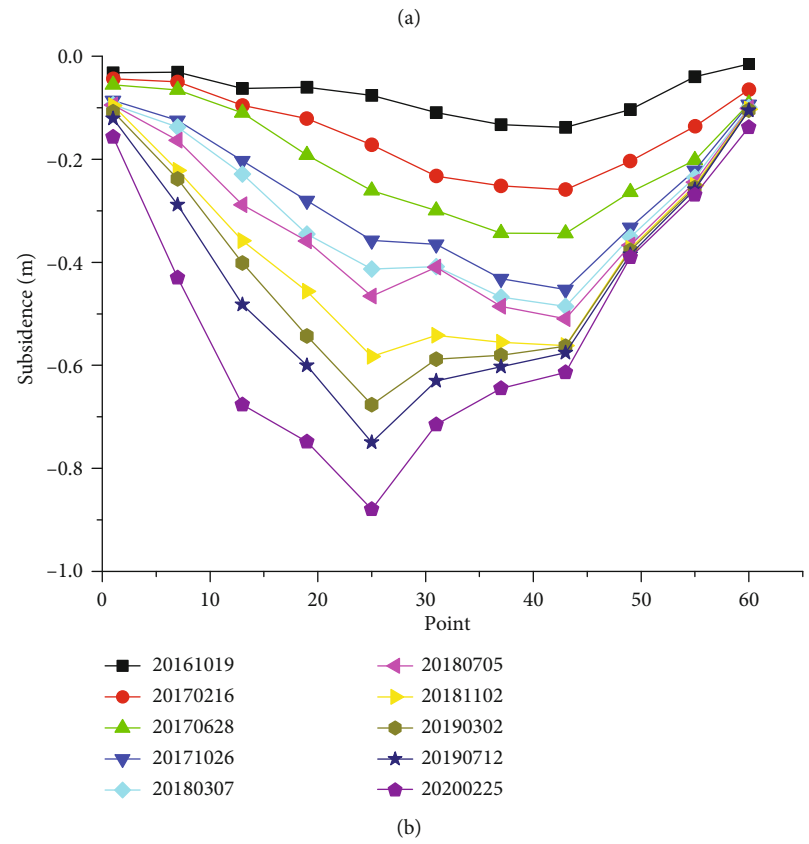
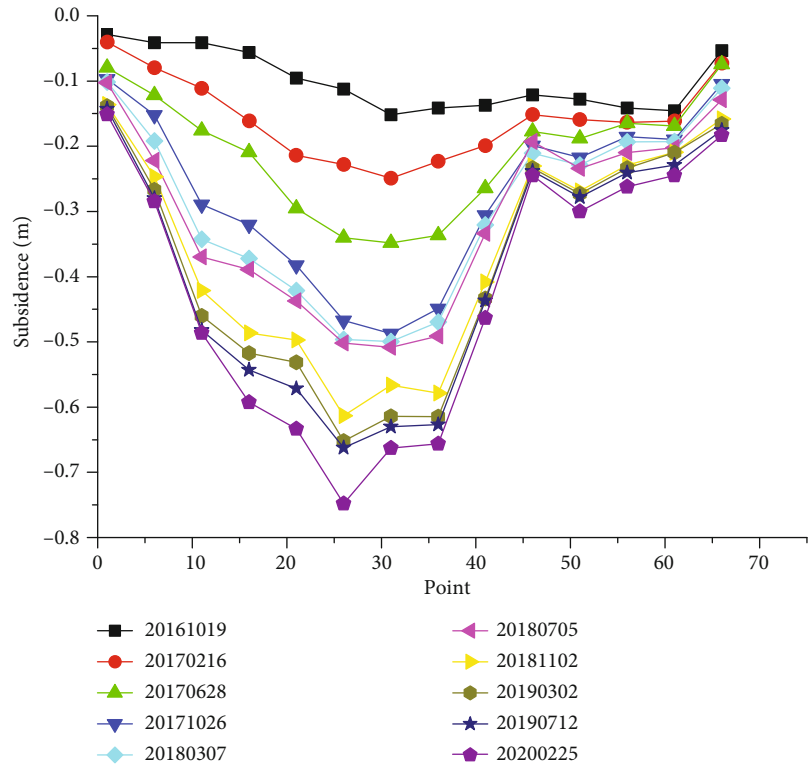


FIGURE 5: Subsidence curves of the working face profiles: (a) A1A2 sinking curve; (b) B1B2 sinking curve.

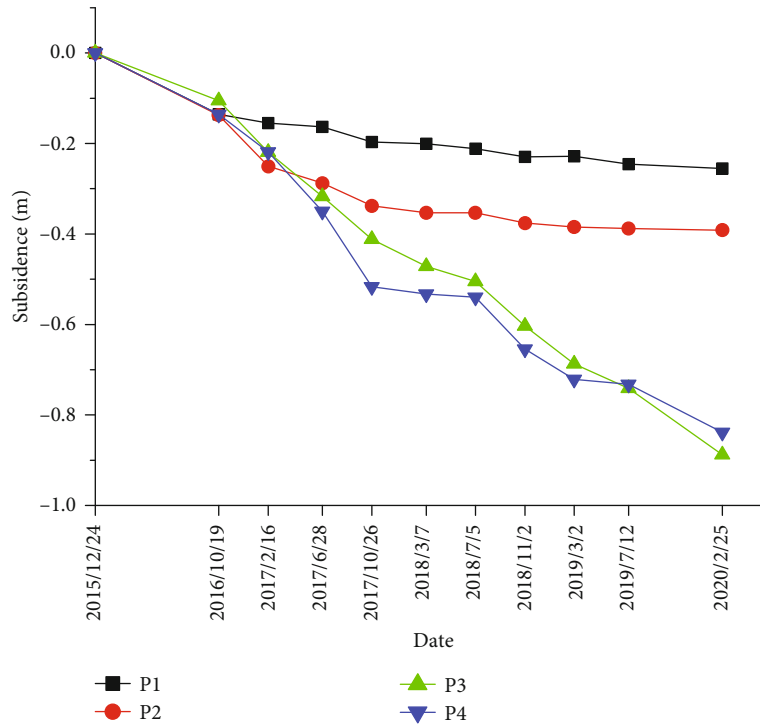


FIGURE 6: Time-series cumulative settlement at characteristic points.

necessary to project the vertical, east-west, and north-south deformations observed from the GPS data to the LOS direction. The projection formula is

$$LOS = [\cos(\theta) \quad -\sin(\theta) \quad \cos(\alpha) \quad \sin(\theta) \quad \sin(\alpha)] \begin{bmatrix} d_v \\ d_E \\ d_N \end{bmatrix} \quad (8)$$

where θ is the local incidence angle of the satellite; α is the azimuth of the satellite's base; and d_v , d_E , and d_N are the vertical, east-west, and north-south deformations, respectively, obtained from the GPS observations.

The above formula was used to convert the GPS-observed deformations into deformations in the LOS direction. Then, the coordinates of the GPS observation points were converted into InSAR data coordinates through Gaussian projection to make the two sets of data consistent in their spatial dimensions. Finally, the extracted values were used via the Point tool of the ArcGIS software to extract the InSAR data deformation points corresponding to the GPS observation points in accordance with the neighbouring pixel values using bilinear interpolation to make the GPS results correspond to the InSAR results. Then, the two sets of results were compared and analysed. As shown in Figure 7, the overall trends of the deformation values obtained from the SBAS-InSAR and GPS observations in the LOS direction are relatively consistent. Only the GPS results and SBAS-InSAR results of 8 observation points at point 4 are abnormal. Because these results were obtained through a combination

of multiple SBAS-InSAR processing experiments, this abnormality may have been caused by human error transmission in the GPS measurements. The smallest deviation of 0.1 cm is observed at point 3. The average absolute error of the 8 points is 1.11 cm, which is within the acceptable range. A correlation analysis of the two sets of data shows that the correlation coefficient is 86%, indicating that they are highly correlated. On the whole, the two sets of results are basically consistent, indicating that SBAS-InSAR technology is reliable for the monitoring of ground subsidence in mining areas and can accurately reflect the corresponding information.

4. Mining Subsidence Prediction

4.1. Model Building. For this experiment, the time-series settlement data based on SBAS-InSAR technology were selected as the training and test samples, and the combined SA-SVR algorithm introduced in the second part of this article was applied to construct a mining area subsidence prediction model and predict the time-series settlement in the mining area. Finally, the SA-SVR prediction results were compared with the prediction results of SVR and GM. To better predict the mining subsidence, a total of 109 sets of data from the Donganshang subsidence area collected from December 24, 2015, to February 25, 2020, were used as the research object in this experiment, and characteristic points were selected along the 21308 working face: points A17 and A25 were selected for the strike section line, and points B30 and B45 were selected for the inclined profile line. Thus, a total of four feature points were used to establish the mining subsidence prediction model.

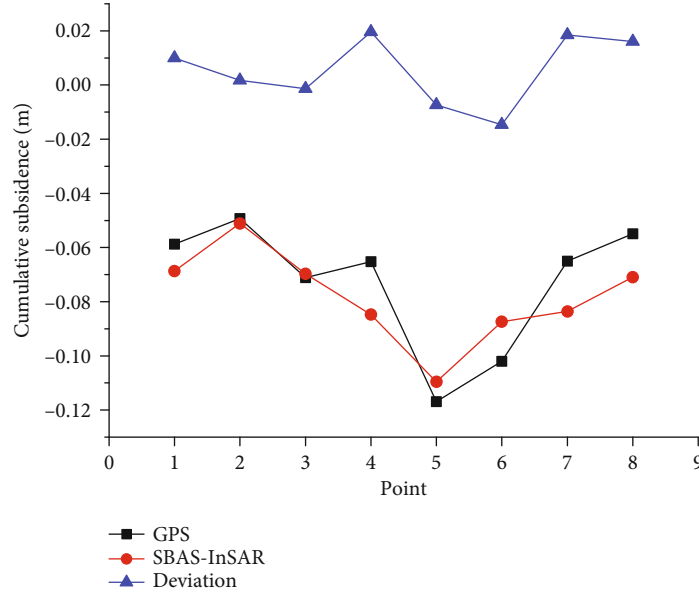


FIGURE 7: Comparison of GPS and SBAS-InSAR results.

First, the parameters of the SVR model were optimally estimated based on the SA algorithm, and then, the optimal estimated parameters were substituted into the SVR model for settlement prediction. The settlement values at each point selected for the experiment constitute a nonlinear time series (\mathbf{x}_i, y_i) ($i = 1, 2, \dots, 104$). The first 104 sets of monitoring data (from December 24, 2015, to December 15, 2019) were used as SVR training samples, and the last five sets of monitoring data (from December 27, 2019, to February 25, 2020) were used as the prediction verification sample. To distinguish between these datasets, they are denoted by

$$(\mathbf{x}_I, \mathbf{y}_I) = \begin{bmatrix} \mathbf{x}_1, y_1 \\ \mathbf{x}_2, y_2 \\ \vdots \\ \mathbf{x}_{104}, y_{104} \end{bmatrix}, (\mathbf{x}_{II}, \mathbf{y}_{II}) = \begin{bmatrix} \mathbf{x}_{105}, y_{105} \\ \mathbf{x}_{106}, y_{106} \\ \mathbf{x}_{107}, y_{107} \\ \mathbf{x}_{108}, y_{108} \\ \mathbf{x}_{109}, y_{109} \end{bmatrix} \quad (9)$$

The first set of data $(\mathbf{x}_I, \mathbf{y}_I)$ was used to determine the unknown parameters in equation (8), namely, $\hat{\omega}$ and $\hat{\mathbf{b}}$. The parameter estimates were then reused in the SVR model for settlement prediction using formula (8):

$$\hat{\mathbf{y}}_{II} = \langle \hat{\omega}, \mathbf{x}_{II} \rangle + \hat{\mathbf{b}} \quad (10)$$

where $\hat{\mathbf{y}}_{II}$ is the predicted value of \mathbf{y}_{II} . The prediction results are listed in Table 1, along with the correlation coefficients. Meanwhile, the displacement map of training, test, and prediction results is shown in Figure 8.

4.2. Model Verification and Analysis. It can be seen from Table 1 and Figure 8 that the prediction model based on the settlement information from the settlement area on the

Donganshang is consistent with the SBAS-InSAR monitoring results. According to Table 1, the maximum absolute error between the predicted and measured values is 5.6 mm, and the maximum root mean square error is 4.1 mm. The maximum average absolute error is 3.8 mm, indicating that the prediction accuracy meets the engineering requirements. For the predictions generated by the mining subsidence prediction model established using the SA-SVR algorithm, the minimum value of the coefficient of determination reaches 0.51. The left side of the dotted line in Figure 8 is the original values and training results of the first 104 sets of training samples (December 24, 2015, to December 15, 2019), and the right side of the dotted line is the original values and prediction results of the last five sets of test samples (2019 from December 7, 2020, to February 25, 2020). It can be seen that the model trend is consistent with the training sample trend as a whole, the prediction error of the test sample is small, and the prediction result is good. It indicates that the subsidence prediction model constructed in this paper has high accuracy, fully meets the requirements of engineering applications, and provides a reliable theoretical basis for mining subsidence monitoring and prediction.

To verify the superiority of the SA-SVR algorithm for the prediction of mining subsidence in the Hancheng mining area in the absence of mining parameter information for this area, characteristic points in time and space were selected to be used in a statistical analysis method for comparing the accuracy of the deformation prediction results of different algorithms. Figure 9 presents a diagram comparing the prediction results for each feature point in the spatial dimension. It can be seen from Figure 9 that the values predicted with the SA-SVR algorithm based on the SBAS-InSAR results show the best fit; the prediction effect at all four feature points is obviously better than that of SVR and GM (1,1), mainly because of the parameter settings of the algorithm.

TABLE 1: Prediction results and correlation coefficients R^2 (unit: mm).

Point	20191227 SBAS predict deviation	20200108 SBAS predict deviation	20200201 SBAS predict deviation	20200213 SBAS predict deviation	20200225 SBAS predict deviation	RMSE	MAE	R^2
A17	-623.7	-632.6	-634.4	-637.3	-640.1	2.8	2.1	0.75
	-623.3	-627.1	-634.9	-638.8	642.7			
	-0.5	-5.5	0.5	1.5	2.5			
A25	-768.1	-772.0	-772.7	-781.8	-783.1	4.1	3.8	0.51
	-763.0	-767.8	-777.5	-782.3	-787.2			
	-5.2	-4.2	4.8	0.6	4.1			
B30	-740.6	-744.7	-748.0	-756.0	-759.2	3.9	3.7	0.68
	-735.0	-740.6	-751.7	-757.3	-762.8			
	-5.6	-4.2	3.7	1.3	3.6			
B45	-521.2	-523.9	-526.2	-528.0	-528.1	0.9	0.7	0.88
	-522.8	-523.8	-525.8	-526.8	-527.8			
	1.6	-0.1	-0.4	-1.2	-0.3			

Note: RMSE (root mean square error) denotes the square root of the ratio of the sum of the squares of the deviations of the observations from their true values to the number of observations; MAE (mean absolute error) denotes the average value of the absolute error and is used to measure the deviation between the observed and true values.

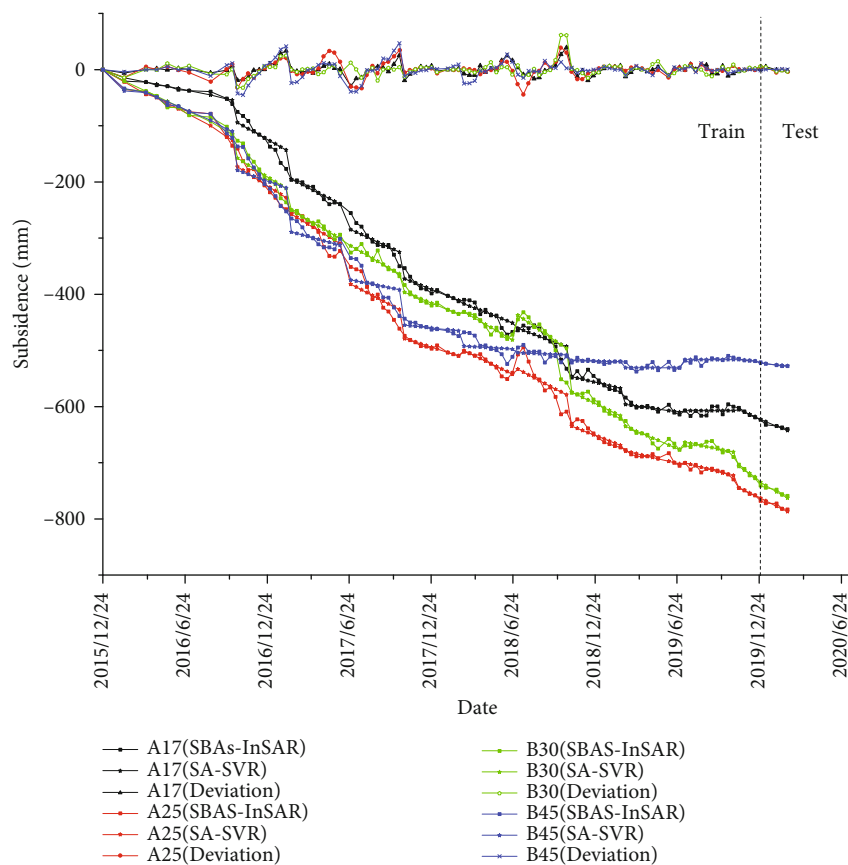
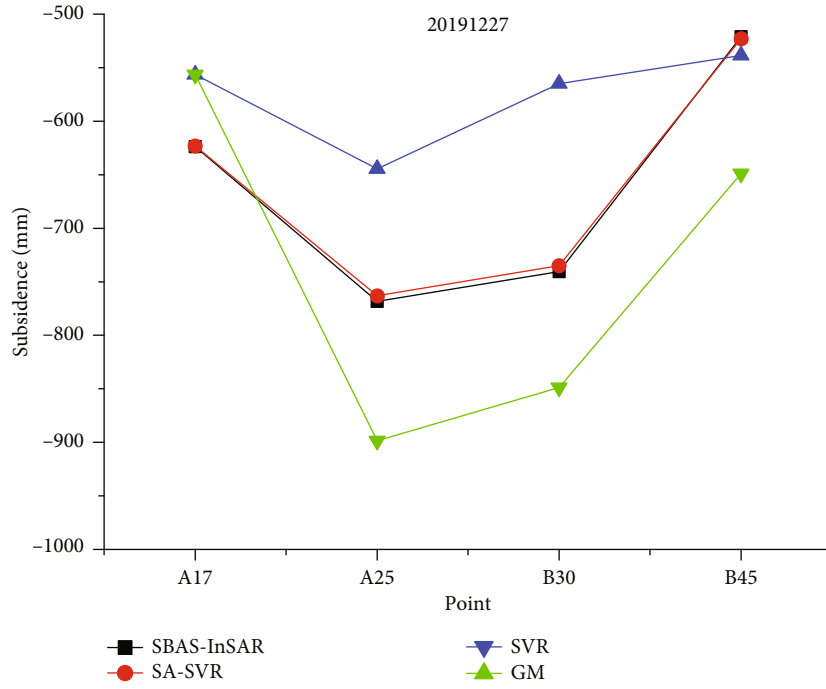


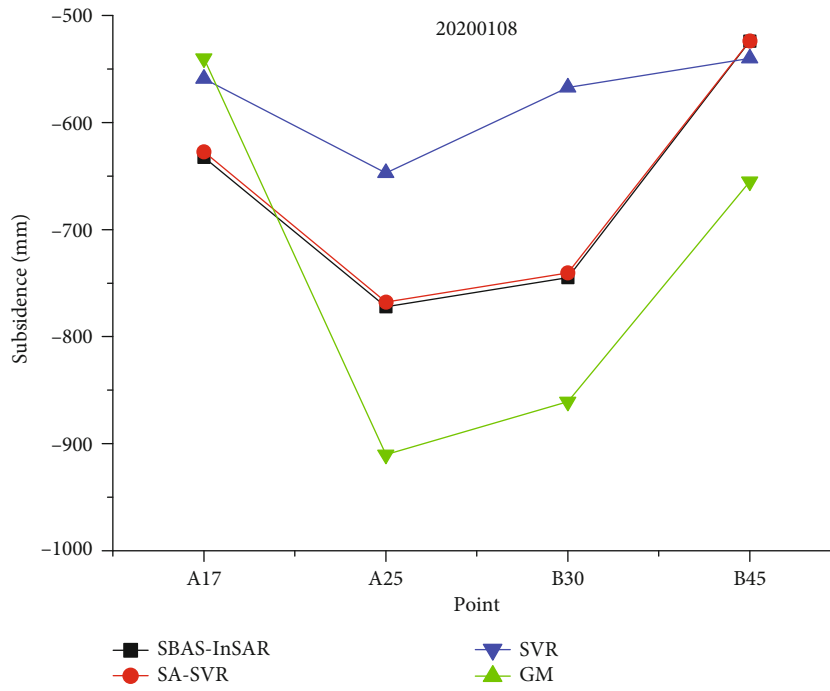
FIGURE 8: The displacement map of training, test, and prediction at 4 points.

In addition, as shown in Figure 10, the prediction accuracy at each selected feature point was analysed in the time dimension. It can be seen that although the SA-SVR predictions for feature points A17 and B45, with smaller settlement values, are very good, the prediction effect is

obviously degraded for feature points A25 and B30, where the values are larger. These findings indicate that the prediction accuracy is affected by the settlement value, with the prediction error being larger when the settlement value is larger.

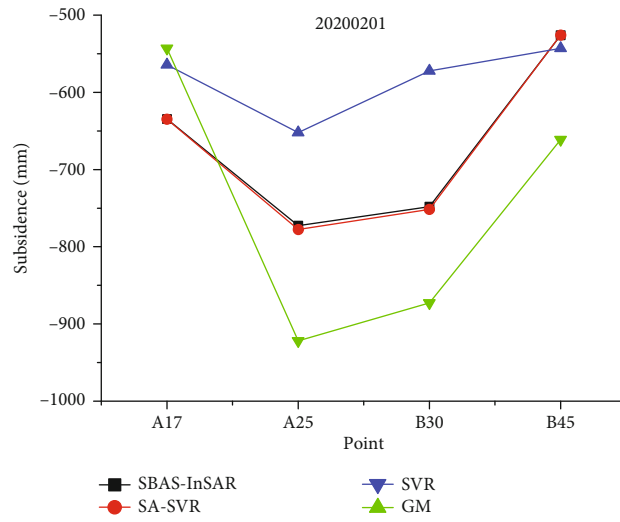


(a)

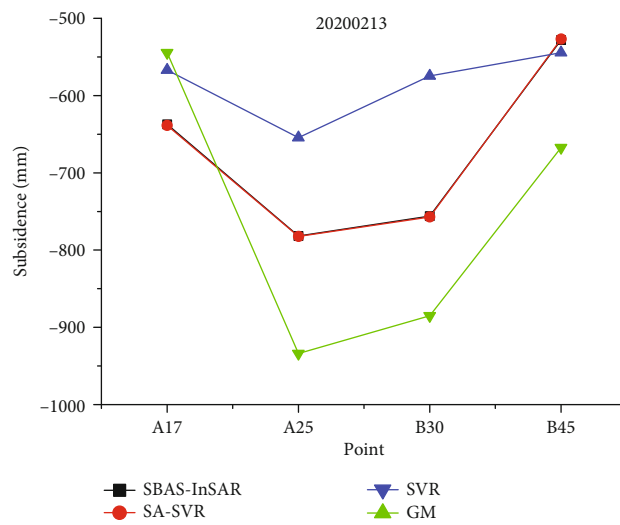


(b)

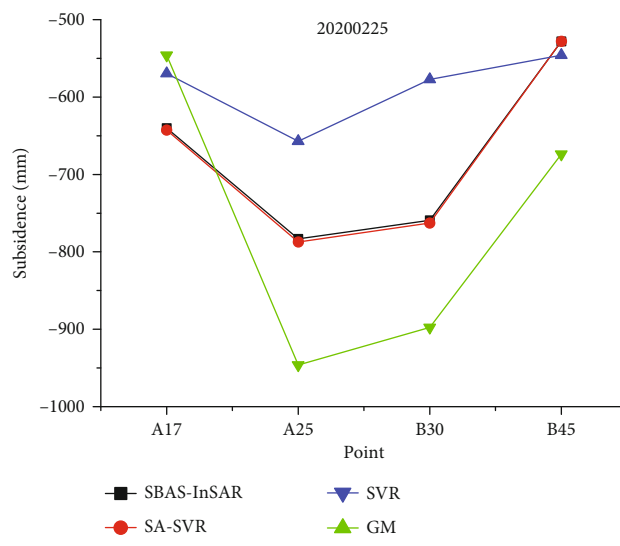
FIGURE 9: Continued.



(c)

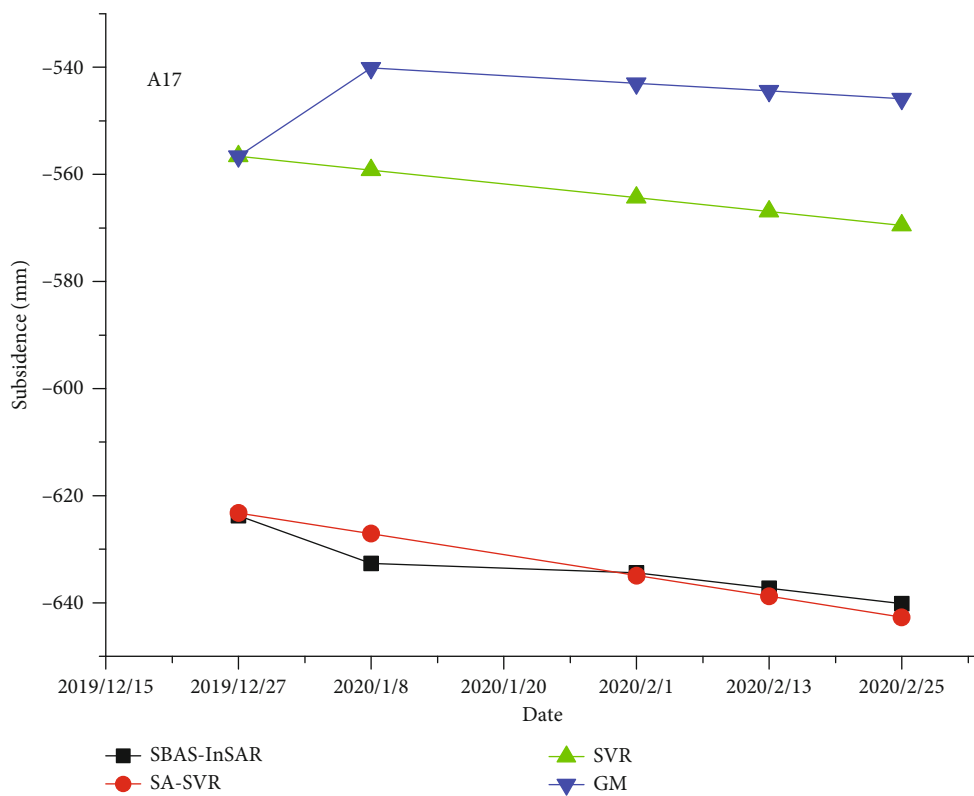


(d)

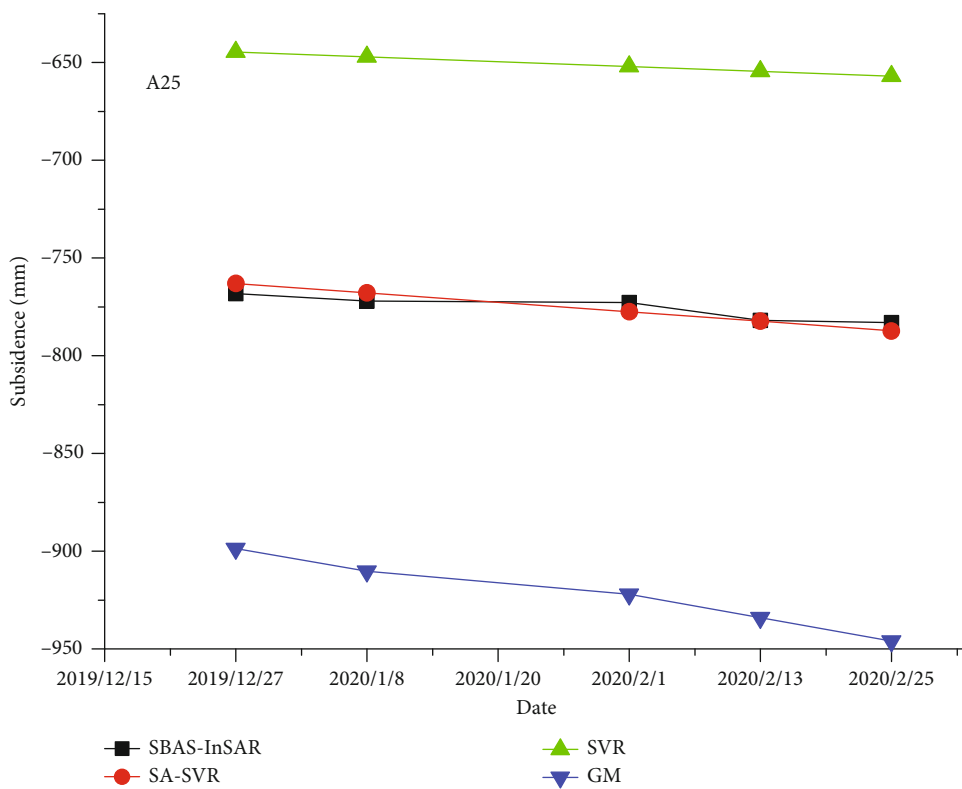


(e)

FIGURE 9: Comparison chart of the results of each point in the spatial dimensions: (a) 20191227. (b) 20200108. (c) 20200201. (d) 20200213. (e) 20200225.



(a)



(b)

FIGURE 10: Continued.

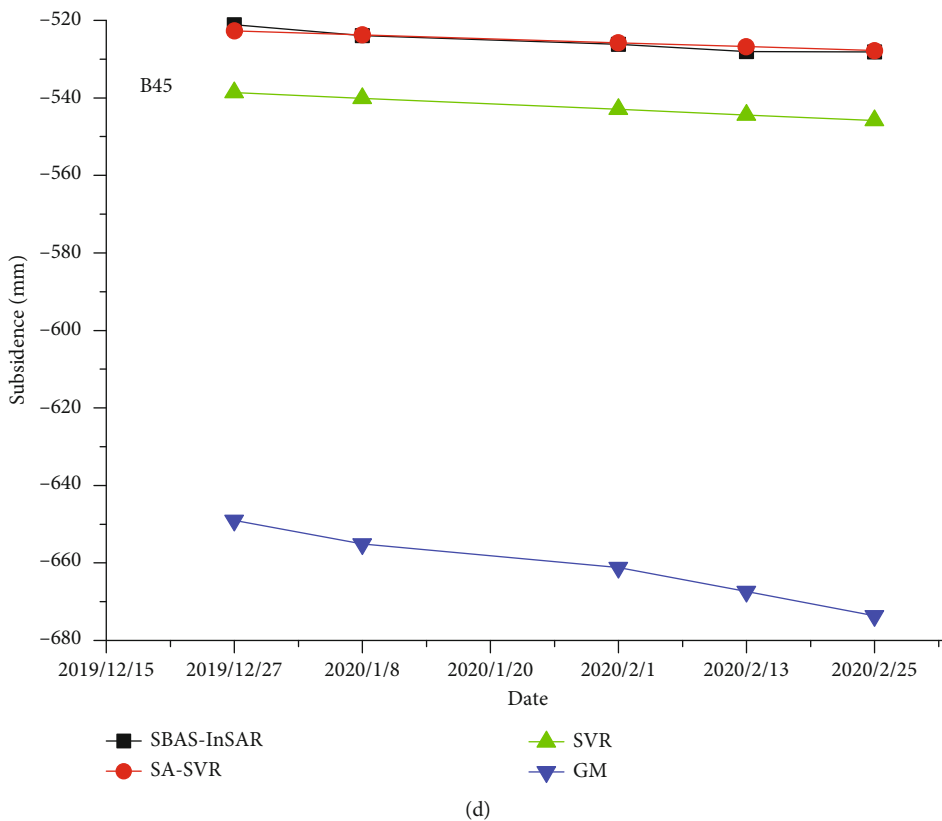
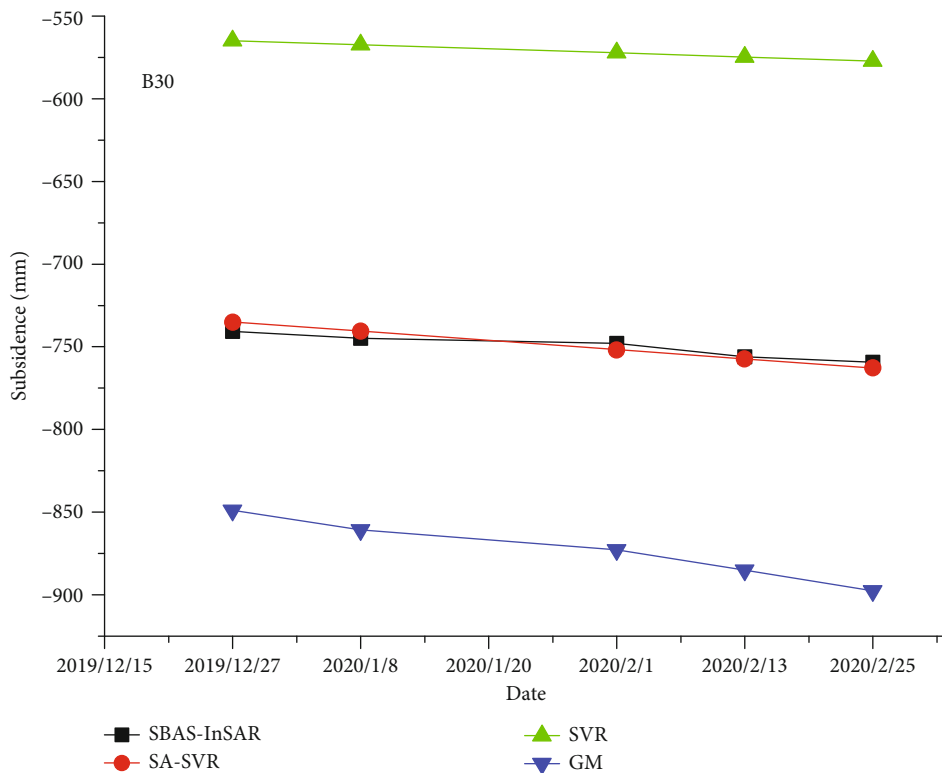


FIGURE 10: Comparison of the trend at each point in the time dimension: (a) A17. (b) A25. (c) B30. (d) B45.

5. Conclusion

This paper has combined SBAS-InSAR technology with the SA-SVR algorithm for the first time for long-term mining subsidence monitoring and prediction for a mining area in Hancheng, China. The surface deformation information of the mining area was extracted based on SBAS-InSAR technology, GPS data were used to verify the monitoring results, and the SBAS-InSAR monitoring results were used as training samples to develop a SA-SVR model for predicting mining subsidence. The SVR and GM algorithms were also used to generate predictions for use as a baseline to compare and analyse the SA-SVR results. The conclusions of this study are as follows:

- (i) The SBAS-InSAR monitoring results are found to be basically consistent with the GPS observations. The minimum deviation is 0.1 cm. The correlation coefficient between the two is 86%, indicating high correlation. This finding illustrates the reliability of the long-term sequence of SBAS-InSAR monitoring data for the surface subsidence in the mining area and reduces cross-validation error
- (ii) The SVR algorithm represents the application of the SVM concept for function approximation and regression estimation. The SVM model is an optimization model proposed by Vaprik [37] in 1995 for regression estimation. To solve the SVM problem, a variety of algorithms have been developed. In 2011, Chang and Lin [38] systematically introduced various mathematical optimization models for solving the SVM problem. The SVM algorithm essentially transforms the original model to be solved into a quadratic mathematical programming optimization model with equality and inequality constraints. Essentially, the original model is expanded; the number of variables changes from the number of unknowns in the regression model to the total number of observations. If there are more observations than unknown parameters of the model, the amount of calculation will increase accordingly. In this paper, the SA algorithm has been directly applied to the original SVM model. By optimizing the parameters of the original model, it becomes more convenient to directly optimize problems with fewer parameters than the number of observations, which may have advantages in terms of convenient programming and ease of calculation. However, in essence, the SA-SVR algorithm obtains an improved solution that is better than the local optimal solution, also called an approximate global optimal solution, at the cost of a large number of calculations
- (iii) The method adopted in this paper improves the accuracy of mining subsidence prediction, indicating that the combination of SBAS-InSAR and SA-SVR can be effectively used for mining subsidence monitoring and prediction

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No: 41674013, No: 41874012).

References

- [1] G. He, L. Yang, G. Ling, F. Jia, and H. Du, *Mining Subsidence*, China University of Mining and Technology Press, Xuzhou, 1991.
- [2] C. Claudie, M. Didier, and K. Christine, "Two examples of the use of SAR interferometry on displacement fields of small spatial extent," *Geophysical Research Letters*, vol. 23, no. 24, pp. 3579–3582, 1996.
- [3] J. Hu, Z. Li, X. Ding, J. Zhu, and Q. Sun, "Spatial–Temporal Surface Deformation of Los Angeles over 2003–2007 from Weighted Least Squares DInSAR," *International Journal of Applied Earth Observation & Geoinformation*, vol. 21, pp. 484–492, 2013.
- [4] F. Calò, F. Ardizzone, R. Castaldo et al., "Enhanced landslide investigations through advanced DInSAR techniques: the Ivancich case study, Assisi, Italy," *Remote Sensing of Environment*, vol. 142, no. 3, pp. 69–82, 2014.
- [5] J. M. Wempen, "Application of DInSAR for short period monitoring of initial subsidence due to longwall mining in the mountain west United States," *Journal of China University of Mining & Technology: English Edition*, vol. 30, no. 1, pp. 33–37, 2020.
- [6] P. Berardino, G. Fornaro, R. Lanari, and E. Sansosti, "A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2375–2383, 2002.
- [7] S. Dong, S. Samsonov, H. Yin, S. Ye, and Y. Cao, "Time-series analysis of subsidence associated with rapid urbanization in Shanghai, China measured with SBAS InSAR method," *Environmental Earth Sciences*, vol. 72, no. 3, pp. 677–691, 2014.
- [8] J. Zhu, Z. Li, and J. Hu, "Research progress and methods of InSAR for deformation monitoring," *Acta Geodaetica et Cartographica Sinica*, vol. 46, no. 10, pp. 1717–1733, 2017.
- [9] L. Zhang, Z. Lu, X. Ding, H.-s. Jung, G. Feng, and C.-W. Lee, "Mapping ground surface deformation using temporarily coherent point SAR interferometry: application to Los Angeles Basin," *Remote Sensing of Environment*, vol. 117, pp. 429–439, 2012.
- [10] J. Hu, Z. Li, J. Zhu et al., "Monitoring 3D surface deformation based on BFGS method fusion InSAR and GPS technology," *Chinese Journal of Geophysics*, vol. 56, no. 1, pp. 117–126, 2013.
- [11] M. Jiang, X. Ding, X. He, and Z. Li, "Timing radar interferometry method based on fast distributed target detection: taking the lost hills reservoir area as an example," *Chinese Journal of Geophysics*, vol. 59, no. 10, pp. 3592–3603, 2016.

- [12] L. Zhang, F. Yang, C. Li, Z. Zhao, and Z. Zhang, "Short baseline set monitoring and analysis of land subsidence in Ningbo," *Science of Surveying and Mapping*, vol. 42, no. 12, pp. 77–82, 2017.
- [13] H. Yin, J. Zhu, Z. Li, X. Ding, and C. Wang, "Ground subsidence monitoring in mining area using D-InSAR SBAS algorithm," *Acta Geodaetica et Cartographica Sinica*, vol. 40, no. 1, pp. 52–58, 2011.
- [14] Z. Yang, Z. Li, J. Zhu, H. Yi, J. Hu, and G. Feng, "Deriving dynamic subsidence of coal mining areas using InSAR and logistic Model," *Remote Sensing*, vol. 9, no. 2, p. 125, 2017.
- [15] A. Saygin, A. Mahmut, S. F. Balik, and C. Ziyadin, "Monitoring of coal mining subsidence in peri-urban area of Zonguldak city (NW Turkey) with persistent scatterer interferometry using ALOS-PALSAR," *Environmental Earth Sciences*, vol. 71, no. 9, pp. 4081–4089, 2014.
- [16] G. Mark and G. Abduwasit, "Evaluation of land subsidence from underground coal mining using TimeSAR (SBAS and PSI) in Springfield, Illinois, USA," *Natural Hazards*, vol. 79, no. 3, pp. 1739–1751, 2015.
- [17] Y. He, B. Qiao, J. Zhang, W. Yan, and W. Wang, "Wide range identification and monitoring Hancheng coal mine by InSAR technology," *Frontiers of Earth Science*, vol. 8, no. 8, pp. 1323–1329, 2018.
- [18] B. Chen, *InSAR Technology and Application Research for Mining Subsidence Monitoring*, China University of Mining and Technology, 2015.
- [19] X. Cui and K. Deng, "Research review of predicting theory and method for coal mining subsidence," *Coal Science and Technology*, vol. 45, no. 1, pp. 160–169, 2017.
- [20] W. Mao, "Prediction of mining subsidence in mining area based on genetic BP neural network model," *Metal Mine*, vol. 2, pp. 164–167, 2016.
- [21] H. Liu and Q. Guo, "Prediction of mining subsidence of Xihaozhuang iron mine based on integrated GPS technology and gray model," *Metal Mine*, vol. 10, pp. 116–119, 2016.
- [22] B. Chen, K. Deng, and H. Fan, "Monitoring and prediction of mining subsidence based on D-InSAR technology and SVR algorithm," *Journal of China University of Mining and Technology*, vol. 43, no. 5, pp. 880–886, 2014.
- [23] J. Zhang, K. Yin, J. Wang, and F. Huang, "Displacement prediction of Bai Shui he landslide based on time series and PSO-SVR model," *Chinese Journal of Rock Mechanics and Engineering*, vol. 34, no. 2, p. 382, 2015.
- [24] H. Qi, "Support vector machine and its application research review," *Computer Engineering*, vol. 10, pp. 6–9, 2004.
- [25] Y. Sun, C. Shao, X. Ji, and L. Zhu, "Traffic accident time series prediction based on the combined model of ARIMA and information granulation SVR," *Journal of Tsinghua University (Science and Technology)*, vol. 54, no. 3, pp. 348–353, 2014.
- [26] S. Alex and S. Bernhard, "A SVR tutorial on support vector regression," *Statistics and Computing*, vol. 14, pp. 199–222, 2004.
- [27] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [28] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to Platt's SMO algorithm for SVM classifier design," *Neural Computation*, vol. 13, pp. 637–649, 2001.
- [29] P. Xu, *Advanced Course of Earth Science and Mathematics: Nonlinear Model, Nonlinear Filtering, Nonlinear Optimization, Robust Estimation and Inversion Problems*, Summer Course of University of Chinese Academy of Sciences, 2019.
- [30] A. Lai, *Sedimentary Characteristics and Accumulation of Coal Measure in Weibei Mining area*, Xi'an University of Science and Technology, 2014.
- [31] Y. Xia, T. Sun, Q. Liang, S. Wang, S. Du, and P. Liu, "Geometry and geodynamic mechanism of buckle folds in Hancheng mining area," *Journal of China Coal Society*, vol. 43, no. 3, pp. 801–809, 2018.
- [32] F. Ling, *Study on Classification Method Using SAR Images for Vegetated Area Identification*, Chinese Academy of Forestry, 2010.
- [33] Y. Zhang, P. Wang, X. Luo, Q. Zhang, and H. Chen, "Monitoring Xi'an Land Subsidence Using Sentinel-1 Images and SBAS-InSAR Technology," *Bulletin of Surveying and Mapping*, vol. 4, pp. 93–97, 2017.
- [34] S. V. Samsonov, P. J. González, K. F. Tiampo, and N. d'Oreye, "Modeling of fast ground subsidence observed in southern Saskatchewan (Canada) during 2008–2011," *Natural Hazards & Earth System Sciences*, vol. 14, no. 2, pp. 5881–5910, 2014.
- [35] L. Bateson, F. Cigna, D. Boon, and A. Sowter, "The application of the intermittent SBAS (ISBAS) InSAR method to the South Wales Coalfield, UK," *International Journal of Applied Earth Observations & Geoinformation*, vol. 34, no. 1, pp. 249–257, 2015.
- [36] J. Zhu, Z. Yang, and Z. Li, "Research progress of 3D surface deformation monitoring and prediction in InSAR mining area," *Acta Geodaetica et Cartographica Sinica*, vol. 48, no. 2, pp. 135–144, 2019.
- [37] V. Vapnik, *The nature of statistic learning theory*, Springer, New York, 1995.
- [38] C. Chang and C. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 1–27, 2011.

Research Article

Hybrid PSO and Evolutionary Game Theory Protocol for Clustering and Routing in Wireless Sensor Network

Yu Song ^{1,2}, Zhigui Liu ¹ and Xiaoli He³

¹School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China

²Department of Network Information Management Center, Sichuan University of Science and Engineering, Zigong 643000, China

³School of Computer Science, Sichuan University of Science and Engineering, Zigong 643000, China

Correspondence should be addressed to Yu Song; songyu@suse.edu.cn and Zhigui Liu; zhigui_liu_swust@hotmail.com

Received 1 July 2020; Revised 18 September 2020; Accepted 9 October 2020; Published 31 October 2020

Academic Editor: Lei Zhang

Copyright © 2020 Yu Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Compared with traditional networks, WSNs have more limited resources such as energy, communication, computing, and storage. The problem of how to achieve energy saving, extend network life cycle, and improve network performance under these limited resources has always been an issue of great interest in WSN research. However, existing protocols do not consider that sensor nodes within the BS threshold may not be clustered. These nodes can directly transmit data to the BS. This simplifies the cluster routing process of the entire WSN and saves more energy. This paper introduces an efficient, and energy-efficient, clustering and equalization routing protocol called the PSOLB-EGT protocol. This protocol introduces a new approach by combining improved particle swarm optimization (PSO) and evolutionary game theory (EGT) algorithms to address the problem of maximizing the network lifetime. The operation of the wireless sensor network is divided into an initialization phase and a data transmission phase. In the initialization phase of the wireless sensor network, the improved PSO algorithm is used to establish clusters and select CHs in areas other than the BS threshold. Entering the data transmission phase, we analyze this problem from the perspective of game theory. We use improved noncooperative evolutionary game theory to build models to solve the problem of the energy waste caused by routing congestion. The proposed PSOLB-EGT protocol is intensively experimented with a number of topologies in various network scenarios, and the results are compared with the well-known cluster-based routing protocols that include the swarm intelligence-based protocols. The obtained results prove that the proposed protocol has increased 9%, 8%, and 5% compared with the ABC-SD protocol in terms of network life, network coverage, and amount of data transmitted, respectively.

1. Introduction

A wireless sensor network (WSN) consists of a large number of microscale, low-power-consumption, and energy-constrained sensor nodes with information sensing, data processing, and wireless communication functions. With the continuous development of network technology and wireless communication technology, WSNs have been widely used in many fields such as the military, environmental monitoring, medical care, and industry.

WSNs are often deployed in hostile environments and need to continuously sense and transmit data unattended. Because a WSN has characteristics such as a large number of nodes, a wide geographical distribution, and a complex

working environment, it is often not realistic to replace the battery to supplement energy after the completion of the layout. Therefore, one of the main challenges in WSNs is the energy consumption problem [1]. Usually, there are two ways to reduce network power consumption: One is to design the hardware equipment of the sensor network to have lower energy consumption, such as a low-power CPU or transmitter. The other is to use more reasonable energy-saving protocols such as a low-energy adaptive clustering hierarchy (LEACH) and the threshold-sensitive energy efficient sensor network protocol (TEEN) [2].

The main task of our study is to optimize the selection and formation of clusters at the initial stage by using an improved protocol. In the data transmission stage, we analyze this

problem from the perspective of game theory and use the improved noncooperative evolutionary game theory (EGT) to solve the routing problem of WSNs [3]. Our proposed method is based on a more reasonable network topology framework with a two-layer cluster structure. Before each round of communication, the improved particle swarm optimization (PSO) algorithm is adopted to reselect cluster-head nodes instead of randomly selecting cluster head (CH) patterns. The PSO algorithm is optimized to avoid local optimization and accelerate global convergence. Compared with previous protocols, this clustering protocol has the following advantages:

- (i) It enables data aggregation at the CH to discard redundant and uncorrelated data. Thereby, it saves energy in the sensor nodes
- (ii) Routing can be more easily managed because only the CH needs to maintain the local route setup of other CHs and thus requires little routing information. This in turn greatly increases the scalability of the network
- (iii) It also saves communication bandwidth because the sensor nodes only communicate with their CH, thus avoiding redundant message exchanges between them

The purpose of the routing algorithm is to select the optimal path to reduce communication delays and energy consumption. Once the optimal set of CHs is elected in the clustering phase, the next step is to find the optimal routing tree from the CHs to the BS while minimizing the total cost. Calculating a desirable route is a challenging problem. Assuming that one path is better than any other path, this path may be used more frequently, which may cause nodes on this path to run out of energy faster [4].

1.1. Related Works. Previous studies have shown that the use of cluster-based layered protocols has broad application prospects for improving the energy efficiency of sensor nodes. In hierarchical routing, when sensor nodes perform multihop communication and data aggregation or fusion in the cluster area, the energy consumption of the sensor nodes is greatly reduced, thereby reducing the amount of information sent to the cluster [5].

LEACH was the first hierarchical routing protocol proposed for WSNs. Many subsequent hierarchical routing protocols are based on LEACH. The LEACH core algorithm consists of three steps [6]. First, CH nodes are randomly selected so that the energy load of the whole network is evenly distributed to each sensor node. Then, data fusion technology is used to reduce the amount of data sent. Finally, the goal of reducing network energy consumption and improving the overall network lifetime is achieved.

$$T(n) = \begin{cases} \frac{p}{1 - p \times [r \bmod (1/p)]}, & \text{if } n \in G, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where p is the probability that the current node becomes the cluster head, r is the current number of rounds, n represents a node, and G is the set of CH nodes that have not been elected in the past $1/p$ rounds. When $r = 0$, each node has the same probability p as the CH. Once a node is elected as the CH, it will not be elected as the CH node in the next $1/p$ round. After $1/p$ rounds, all nodes will be elected as the CH.

Although LEACH can effectively reduce the energy consumption of the entire network, it also has many defects. The most obvious shortcomings are as follows: (i) Each CH communicates with the BS in a single-hop mode. A CH node that is far away from the BS consumes a large amount of energy. (ii) The residual energy of the current node is not considered when selecting the CH. If the energy of the randomly selected CH is too low, it may accelerate node death and reduce network lifetime. (iii) The number of CH nodes in LEACH is usually 5% of the number of nodes. When the distribution density of the sensors varies, a fixed number of CHs cannot optimize the network overhead [7].

It is precisely because of the obvious defects of the LEACH protocol that many scholars are trying to optimize it. Some experts have achieved good results using fuzzy logic methods [8].

WSN routing protocols can be classified into active and reactive types according to different application modes. Active sensor networks continuously monitor surrounding physical phenomena and send monitoring data at a constant rate, while responsive sensor networks only transmit data when observed variables change.

TEEN, which is also based on the LEACH protocol, is the first hierarchical WSN routing protocol for responsive networks. TEEN works in the same way as LEACH, except that it only sends data after the sensor node detects relevant data [9].

After TEEN recluses each cluster area, the CH node needs to broadcast the following three parameters to the members of the cluster: (i) Feature value: this is the physical parameter of the data that the user cares about. (ii) Hard threshold (HT): this is the absolute threshold of the monitored data eigenvalue. When the characteristic value monitored by the node exceeds this threshold, the transmitter is started and reports this value to the CH node. (iii) Soft threshold (ST): this monitors the small range change threshold of the feature value to trigger the node to start the transmitter to report data to the CH.

By setting hard and soft thresholds, TEEN effectively reduces the amount of data sent and is more energy efficient than LEACH. It is suitable for environments that require real-time monitoring of changes. Monitors can also balance the accuracy and requirements of the system of monitoring data by setting different soft thresholds [10].

Although the TEEN protocol is more energy efficient than the LEACH protocol, TEEN still has the following disadvantages: (i) When the threshold has not been reached, the user will not be able to obtain information. (ii) The receiver of the CH node is always active. To receive data from member nodes at any time, the receiver increases the burden on the CH nodes to some extent.

Swarm intelligence optimization originates from a specific phenomenon of group movement in nature. By studying

the group behavior of these real creatures, human beings seek the rules governing them through simulation and imitation, and we build artificial intelligence models to solve complex problems that cannot be solved by conventional methods in daily life and new problems that have not yet been solved. In recent years, some intelligent algorithms, such as ACO, PSO, and GA, have been studied to solve NP-hard optimization problems in WSNs. In 1995, Kennedy and Eberhart proposed and designed a general method for solving practical optimization problems, namely, the particle swarm optimization (PSO) algorithm [11]. By sharing and conveying information about the location information of a food source among groups, the speed of finding food can be accelerated, and the final goal can be achieved through joint effort. The PSO algorithm is easy to realize with high accuracy and is faster and more efficient than other algorithms [11, 12].

People often refer to two important kinds of information in their decision-making process. The first is their own experience, and the second is the experience of others in the group. Similarly, in the process of foraging in birds, each bird's initial state is in a random position, and the direction of flight is also random. However, over time, these initially random birds spontaneously organize into a colony by learning from each other, sharing information, and accumulating foraging experiences. Each bird remembers the best location it finds, which is called a local optimum. In addition, it can also remember that the optimal location searched so far in the flock is the global extreme value, which is called the global optimum. The foraging centers of the whole flock move toward the global optimum.

In the PSO model, each individual can be regarded as a particle, and the flock of birds as a swarm of particles. In a D -dimensional target space, for a group of M particles, the position of the i -th particle is expressed as x_{id} . In other words, the position of each particle is a potential solution. By substituting x_{id} into the objective function, its adaptive value can be calculated and evaluated according to the size of the adaptive value. The best position experienced by individual particles is denoted as $p_{best}(t)$, and the best position experienced by all particles in the whole population is denoted as $g_{best}(t)$. The velocity of the i -th particle is expressed as v_{id} . Standard PSO (SPSO) can be expressed with

$$\begin{aligned} v_{id}(t+1) &= w \times v_{id}(t) + c_1 \times r_1 \times (p_{best}(t) - x_{id}(t)) \\ &\quad + c_2 \times r_2 \times (g_{best}(t) - x_{id}(t)), x_{id}(t+1) \\ &= x_{id}(t) + v_{id}(t+1), \end{aligned} \quad (2)$$

where w is the inertia weight factor and parameters c_1 and c_2 are collectively referred to as learning factors and are, respectively, referred to as cognitive parameters and social parameters. r_1 and r_2 are random numbers in the range $[0, 1]$.

We usually use a linearly decreasing inertia weighting factor w to improve performance. At each update iteration, the value of w decreases linearly from approximately 0.9 to 0.4. Choosing the appropriate inertia weight w can provide a balance between global and local search, and it can result in finding a sufficiently optimal solution with fewer average iterations. Its value is set as follows:

$$w_c = w_{max} - (w_{max} - w_{min}) \left(\frac{t}{T_{max}} \right), \quad (3)$$

where T_{max} is the total number of iterations, t is the current iteration, $w_{max} = 0.9$, and $w_{min} = 0.4$. w_c is a linearly decreasing value. It decreases with the increase in the number of search rounds, which indicates that the effect of the inertial velocity of particles decreases gradually.

Traditional PSO algorithms do not have genetic and crossover operations, relying on the $p_{best}(t)$ and $g_{best}(t)$ of the particles to complete the search. It has the advantages of fewer parameters, simple structure, fast search speed, and easy convergence. However, due to the lack of dynamic adjustment of parameters, it can easily fall into a local optimum, resulting in low convergence precision.

Wireless sensor network routing is a challenging area of research. In general, when we try to optimize routing problems, there are many metrics that should be considered, for example, the distance between nodes, data delay, residual energy of nodes, transmission rate of each link, and distributed characteristics of wireless sensor networks.

In the past, game theory focused on the field of economics, and it was used to study the decision-making process of economic activities so that people could optimize outcomes for their economic interests. Since the 1980s, game theory has been improved and applied more widely. In addition to its applications in economics, it is also widely used in biology, computer science, public policy, and other disciplines and has had an important impact.

Biologists Maynard Smith and Price in 1973 introduced classical game theory into biological evolution analysis and put forward the basic equilibrium concept of EGT, the evolutionary stable strategies (ESSs). In 1978, Taylor and Jonker discovered the relationship between evolutionary stable strategies and replication dynamics, marking the birth of EGT.

In classical game theory, it is assumed that the players in the game are completely rational and the decisions they make are optimal. These players are known to be rational and fully aware of the game. However, in the actual maximum lifetime problem of WSNs, not all nodes participating in the game are completely rational, so the assumptions of classical game theory are not applicable to this problem.

Compared with classical game theory, EGT does not require participants to be completely rational and have sufficient information but only requires participants with limited rationality to learn from each other step by step to make the whole group composed of participants reach equilibrium. It studies the entire group of participants [12]. Therefore, EGT is selected in this paper to study the routing problem of improved protocol in WSNs.

There are two basic concepts in evolutionary game theory, i.e., ESSs and replication dynamic models. The ESS is defined as follows:

If there is a real number α greater than 0, for all the policy sets $m(m \neq n)$ such that

$$U(n, \beta m + (1 - \beta)n) > U(m, \beta m + (1 - \beta)n), \quad (4)$$

for any $\beta \in (0, \alpha)$, then n is called an ESS factor and α is the invasion bound, which is a constant associated with policy m . $\beta m + (1 - \beta)n$ is determined from groups that choose the ESS and groups that adopt a mutation strategy, and it belongs to a mixed group [13].

Replication dynamic theory analyzes the behavior of the whole group based on the principle of “survival of the fittest” in the theory of evolution and changes in the group behavior in the evolutionary game with this method. The expression that determines group behavior is as follows:

$$X(t) = \frac{dx}{dt} = X_i(t)[(U_i(t) - \bar{U}(t))], \quad (5)$$

where $X_i(t)$ represents the proportion of the number of individuals in the group choosing strategy i at time t in the whole group. $U_i(t)$ represents the benefit for individuals in the group who select strategy i at time t . $\bar{U}(t)$ is the average benefit received by each individual in the group at time t [14].

In this section, we review the game theory used to enhance energy conservation and extend the network lifetime. If the individuals who select strategy i in the group obtain more benefits than the average utility of each individual in the group, then the proportion of individuals who select pure strategy i will increase. Conversely, if the utility obtained by the individuals in the group choosing pure strategy i is smaller than the average utility obtained by each individual in the group, the individuals will change their choice to other strategies accordingly [15, 16].

For a better solution, there must be a balance between the WSN energy consumption and the stability of the system. This motivated us to combine PSO and EGT into a new improved protocol. The reasons for the superior performance of the proposed hybrid PSO-EGT protocol over existing protocols are given in what follows.

1.2. Contributions. This paper proposes a CH selection algorithm for WSNs based on improved PSO search and finds the best data transfer path between CHs and the BS through EGT. The main contributions of this paper include the following:

- (i) In clustering, the node within the distance threshold D_0 directly transmits data to the BS, which greatly reduces the energy consumption of the clustering and routing of these nodes
- (ii) A model for calculating the optimal CH number is proposed. Instead of determining the number of CHs as the percentage of the total number of nodes in the past, the model determines the optimal cluster head number according to the actual parameters in the WSN working environment
- (iii) In terms of routing, the working area of the WSN is sorted and planned into different concentric circles according to the threshold value D_0 , and the routing is layered

- (iv) When selecting the transmission path, the shortest path is changed from the previous simple choice to the path jointly determined by the shortest path, the residual power of the relay node, and congestion penalty factors

1.3. Organization. The rest of the paper is organized as follows: In Section 2, the energy and system models are introduced. Section 3 provides two algorithms (i.e., hybrid PSOLB and EGT) for the analysis and design of clustering and routing. The simulation results are presented in Section 4. Finally, Section 5 concludes this paper.

2. Energy and System Model

2.1. Energy Model. Generally, a wireless sensor network node is composed of four modules, i.e., an information sensing module, an information processing module, an information communication module, and an energy supply module. The information sensing module is responsible for collecting and transforming information about the perceived object. The information processing module is responsible for controlling the operation of the entire node and storing and processing the data it collects itself and the data sent by other nodes. The information communication module is responsible for communicating with other nodes, communicating via interactive control messages, and receiving and transmitting service data. The energy supply module is responsible for providing the sensor nodes with the energy required for operation, typically via a large capacity microbattery. It can be seen that each module needs to consume energy. Therefore, it is necessary to establish an energy model, also called an energy consumption model. In contrast to previous studies, we consider the energy consumption of each functional module of the wireless sensor node. The energy consumption model is shown in Figure 1. It can be seen that the energy consumption of the wireless sensor node is mainly determined by three modules. In this section, we will build an information sensing energy consumption model, an information processing energy consumption model, and an information communication energy consumption model. It is necessary to analyze the energy consumption of each module to establish the total energy consumption model of the node [17].

The energy consumption of the information sensing module is related to the power, sensing time, and sensing data quantity of the sensor. Generally, the sensing energy consumption E_s for a node can be expressed as follows:

$$E_s = I_s + V_s + L_s + T_s, \quad (6)$$

where I_s and V_s represent the current and voltage when sensing, respectively; L_s denotes the number of bits in the sensing data; and T_s represents the sensing time.

The information processing energy consumption is mainly determined by the energy consumption per unit byte of information processing and the total amount of data to be processed. The calculation is shown in

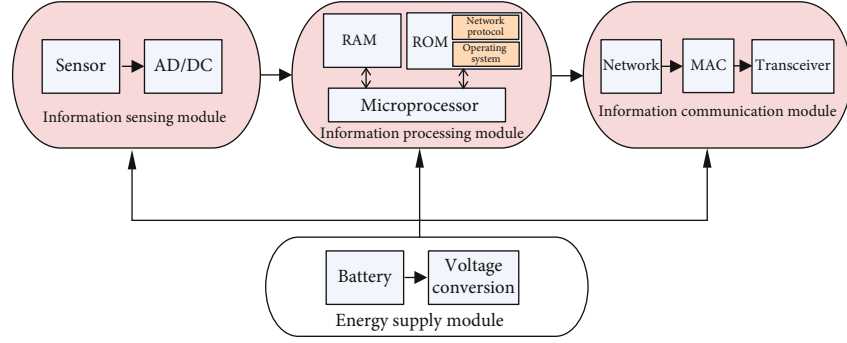


FIGURE 1: Composition of wireless sensor node energy consumption.

$$E_p = l \times E_{da}, \quad (7)$$

$$E_{da} = I_{read} \times V_{read} \times T_{read} + I_{write} \times V_{write} \times T_{write},$$

where l is the total amount of data to be processed and E_{da} represents the energy required for processing one bit of data information. I_{read} , V_{read} , and T_{read} , respectively, represent the current, voltage, and time for reading data when the sensor processes data. In the same way I_{write} , V_{write} , and T_{write} , respectively, represent the current, voltage, and time for writing data when the sensor processes data.

The energy consumption of sensor nodes during the process of data communication is mainly composed of two parts, i.e., the energy consumption of receiving information and the energy consumption of sending information sending [18].

Energy consumed by data transmission is

$$E_{tx}(l, d) = E_{tx-elec}(l) + E_{tx-amp}(l, d) = \begin{cases} l \times E_{elec} + l \times \epsilon_{fs} \times d^2, & d < d_{th}, \\ l \times E_{elec} + l \times \epsilon_{mp} \times d^4, & d \geq d_{th}, \end{cases} \quad (8)$$

and the energy consumed by data reception is

$$E_{rx}(l) = E_{rx-elec}(l) = l \times E_{elec}, \quad (9)$$

$$d_{th} = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}},$$

where l is the total amount of data to be transmitted or received. E_{elec} is the energy consumed to send one bit of data. The amplifier power consumptions ϵ_{fs} and ϵ_{mp} are determined by the transmission distance and the received bit error rate. d is the distance between two sensor nodes. The signal energy consumption model is divided into two categories according to distance (i.e., a free space model and a multipath attenuation model). When the transmission distance is less than the distance threshold d_{th} , the free space energy consumption model is adopted in the communication mode; otherwise, the multipath attenuation model is adopted. d_{th} is a constant, and the value depends on the network environment.

Thus, we can determine the total energy consumed by each sensor [19].

$$E_{total} = E_s + E_p + E_{tx} + E_{rx}. \quad (10)$$

2.2. System Model. In this section, we assume a WSN consists of N nodes (H CH nodes and $N-H$ non-CH nodes) and one BS. There is a set of sensors that are randomly distributed in a designated area. Once the deployment is complete, all sensor nodes become static nodes. In the model we propose, each sensor node can only be assigned to one cluster, and each CH node acts as the CH of exactly one cluster. Each node has local information, including its own unique ID, the CH node ID, the information collection and transmission round, the residual energy level, and the distance to its neighbors, etc. As discussed above, the entire operation of the WSN is divided into T rounds by time. In each round, a normal sensor node sends the monitored data to its cluster head. After receiving the data, the cluster head fuses the data to discard the redundant data and either sends the fused data directly to the BS or forwards it to the BS in multiple hops through other nodes (CH or member nodes) [20].

As shown in Figure 2, according to the general working principle of the WSN, this paper analyzes the initialization stage and data transmission stage. The focus is on optimizing the election of CHs and data transfer [21].

In this paper, the proposed PSO-EGT protocol is based on the classical LEACH protocol. The protocol uses a round as a unit; each round consists of an initialization phase and a data transmission phase for the purpose of reducing unnecessary energy consumption. The specific process of the two phases of each round is shown in Figure 2. The initialization phase is for calculating the optimal number of clusters, electing the CHs, and forming clusters. In the data transmission phase, the WSN performs data aggregation and data transmission and reception [22].

The proposed scheme, called the hybrid PSO-EGT protocol, is based on the common characteristics of both the PSO and EGT algorithms. In this paper, a centralized two-tier PSOLB-EGM protocol is proposed to solve the problem of clustering and routing in WSNs [23]. To implement the hybrid PSO-EGT protocol, the flows to be followed are shown in Figure 3.

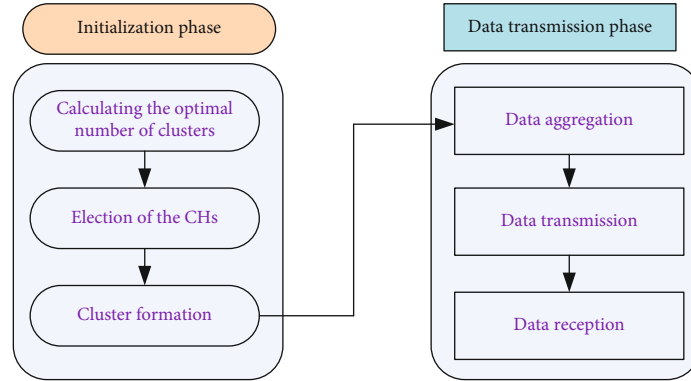


FIGURE 2: Structure of the WSN general working principle.

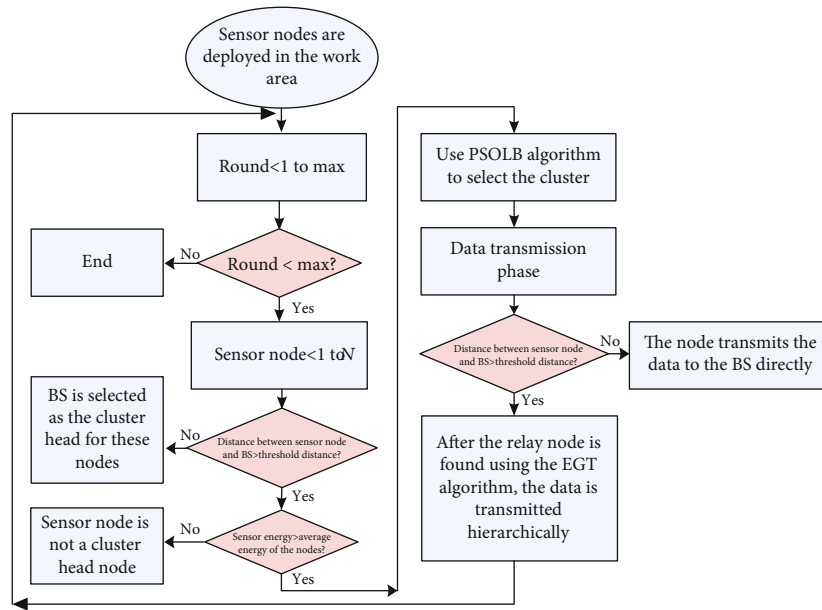


FIGURE 3: Flows of the PSOLB-EGT protocol.

The main task of the protocol in the initialization phase is to select the cluster head.

First, a node must satisfy two conditions at the same time to become a candidate cluster head: the distance from the node to the BS is greater than the threshold. The residual energy of the node is greater than the average residual energy of all nodes.

The cluster head is selected using our improved PSOLB algorithm for candidate cluster head.

In the data transmission phase, nodes within the BS threshold directly transmit data to the BS. Nodes outside the BS threshold use the improved EGT algorithm to transmit data to the BS.

3. Algorithm Design and Implementation

3.1. The PSOLB Algorithm and Parameter Setting. According to the energy consumption model of WSN, the energy consumption of nodes is affected by sensor data volume, sensor

time, node distance, and other factors. Among these influencing factors, the first to be considered is the distance between sensors after clustering, that is, the compactness of clusters [24].

In many existing clustering algorithms, the number of CHs is usually fixed. The algorithm proposed in this paper takes into account that in an actual working environment, sensor node power, cluster number, and other factors are usually related to the size of the monitoring area, BS location, and number of sensor nodes. If there are too many CHs in the network, redundant CH nodes will consume more energy in the network. If the number of CH nodes in the network is too small, remote nodes will not be able to find suitable CHs to join, resulting in the loss of monitoring data and excessive network delay [25].

It is assumed that N sensor nodes are randomly arranged in a monitoring area of size $M \times M$. Since the work of the CH node is mainly divided into receiving information, processing information, and sending information, the energy consumption

of a single CH node is the sum of the energy consumptions of the above three tasks. In the initial condition, we assume that the CH node is far from the BS. According to equations (8) and (12), we can determine the total energy consumption of the i -th CH node as follows:

$$E_{\text{CH}_i} = E_{tx} + E_{rx} + E_{dx} = (a \times l \times E_{\text{elec}}) + (l \times E_{\text{elec}} + l \times \varepsilon_{mp} \times d_{\text{toBS}}^4) + (a \times l \times E_{da}), \quad (11)$$

where $a = N/H$ represents the total number of nodes in the cluster (including one CH node and $a - 1$ non-CH nodes), l represents the total amount of transmitted data, E_{da} is the energy consumption of data aggregation, and d_{toBS} is the distance from the CH node to the BS [26].

Then, according to the Euclidean theorem, the distance between a CH and the BS can be expressed as follows:

$$d_{\text{toBS}} = \sqrt{x_1^2 + y_1^2}. \quad (12)$$

The function of a non-CH node is to send perceived information. We assume that the distance between all non-CH nodes and CH nodes is within the threshold d_{th} . The total energy consumption of the j -th non-CH node in the i -th cluster can be calculated by

$$E_{\text{non-CH}_j} = l \times E_{\text{elec}} + l \times \varepsilon_{fs} \times d_{\text{toCH}_j}^2. \quad (13)$$

Similarly, the distance between the j -th non-CH node and the i -th CH can be expressed as follows:

$$d_{\text{toCH}_j} = \sqrt{x_2^2 + y_2^2}. \quad (14)$$

The expected distance from the CH node to the sink node is

$$\begin{aligned} E(d_{\text{toBS}}^4) &= \int_0^M \int_0^M \left(\sqrt{x_1^2 + y_1^2} \right)^4 \rho(x_1, y_1) dx_1 dy_1 \\ &= \int_0^M \int_0^M (x_1^2 + y_1^2)^2 \rho(x_1, y_1) dx_1 dy_1, \end{aligned} \quad (15)$$

where $\rho(x_1, y_1)$ is the network node density function. Similarly, we can determine the mathematical expectation of $d_{\text{toCH}_j}^2$.

$$\begin{aligned} E(d_{\text{toCH}_j}^2) &= \int_0^M \int_0^M \left(\sqrt{x_2^2 + y_2^2} \right)^2 \rho(x_2, y_2) dx_2 dy_2 \\ &= \int_0^M \int_0^M (x_2^2 + y_2^2) \rho(x_2, y_2) dx_2 dy_2. \end{aligned} \quad (16)$$

The total energy consumption of the entire WSN can be obtained from equation (16).

$$E_{\text{total}} = \sum_{i=1}^H \left(E_{\text{CH}_i} + \sum_{j=1}^{a-1} E_{\text{non-CH}_j} \right). \quad (17)$$

According to equations (15) and (16), the expected value of the total energy consumption is

$$\begin{aligned} E(E_{\text{total}}) &= H \times E(E_{\text{CH}_i}) + (N - H) \times E(E_{\text{non-CH}_j}) \\ &= H \times ((a \times l \times E_{\text{elec}}) + (l \times E_{\text{elec}} + l \times \varepsilon_{mp} \times d_{\text{toBS}}^4) + (a \times l \times E_{da})) \\ &\quad + (N - H) \times (l \times E_{\text{elec}} + l \times \varepsilon_{fs} \times d_{\text{toCH}_j}^2) \\ &= H \times ((a \times l \times E_{\text{elec}}) + (l \times E_{\text{elec}} + l \times \varepsilon_{mp} \times E(d_{\text{toBS}}^4)) + (a \times l \times E_{da})) \\ &\quad + (N - H) \times (l \times E_{\text{elec}} + l \times \varepsilon_{fs} \times E(d_{\text{toCH}_j}^2)). \end{aligned} \quad (18)$$

We take the derivative of equation (18) with respect to H and set its derivative equal to zero, so that we can determine the optimal number of clusters containing the maximum lifetime of all nodes.

$$\frac{dE(E_{\text{total}})}{dH} = 0. \quad (19)$$

In contrast to the traditional PSO algorithm, we first propose that the best result of the last round particle search should be learned; this result is denoted as l_{best} , and we assign the learning weight c_3 to it [27].

$$\begin{aligned} v_{id}(t+1) &= w \times v_{id}(t) + c_1 \times r_1 \times (p_{\text{best}}(t) - x_{id}(t)) \\ &\quad + c_2 \times r_2 \times (g_{\text{best}}(t) - x_{id}(t)) \\ &\quad + c_3 \times r_3 \times \Omega \times (l_{\text{best}}(t-1) - x_{id}(t)), \end{aligned}$$

$$\Omega = \begin{cases} 1, & \text{otherwise,} \\ 0, & l_{\text{best}}(t-1) = p_{\text{best}}(t) \text{ or } l_{\text{best}}(t-1) = g_{\text{best}}(t), \end{cases}$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1), \quad (20)$$

where parameters c_1 and c_2 are collectively referred to as learning factors and are, respectively, referred to as cognitive parameters and social parameters. c_3 is the learning factor of l_{best} . r_1 and r_2 are the intelligent learning factor in this paper. Similarly, r_3 is the intelligent learning factor of l_{best} . Ω is a control parameter. When l_{best} coincides with p_{best} or g_{best} , $\Omega = 0$, and $c_3 \times r_3 \times \Omega \times (l_{\text{best}}(t-1) - x_{id}(t))$ fails to prevent repeated learning of particles.

The PSOLB algorithm allows particles to learn from the best particles in the previous round of searching to avoid falling into a local optimization and to find the global optimization faster. Therefore, PSOLB can meet the requirements of the exploration and development of high-dimensional problems.

In each round, the particle updates its position through three extreme values (i.e., p_{best} , the individual historical optimal solution; g_{best} , the global historical optimal solution; and l_{best} , the global optimal solution in the last round). The

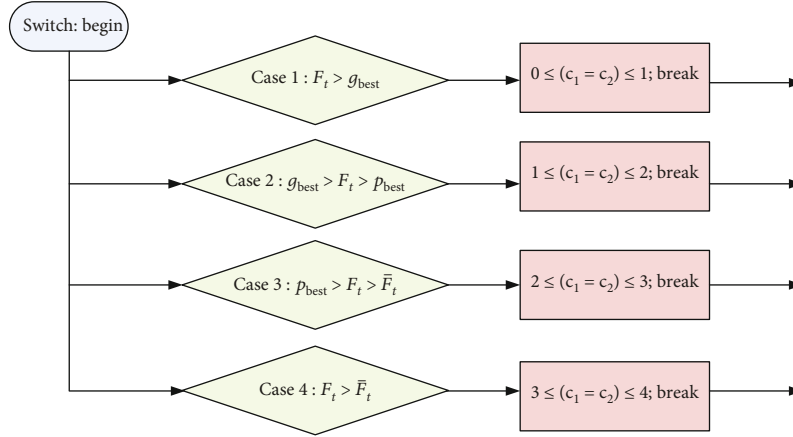


FIGURE 4: Specific improvement methods.

position of the particle is the optimal solution of the distance and is the only criterion for evaluating the particle.

Another improvement to the standard PSO algorithm is to change the learning factors c_1, c_2 from fixed values to dynamic adaptive learning factors. In the basic PSO algorithm, c_1, c_2 have the ability of self-learning and learning from excellent individuals, so that these particles are close to the best points in the group or field. c_1 and c_2 regulate the maximum stride length of particles flying in the optimal direction of individuals or groups, respectively. When the learning factor is small, a particle may wander in a region far away from the target region. When the learning factor is large, the particle can move rapidly to the target region or even beyond the target region. Therefore, the values of c_1 and c_2 will affect the performance of the PSO algorithm [28]. The specific improvement methods are as given in Figure 4.

From Figure 4, we can see that in the proposed method, the learning factors c_1 and c_2 are no longer fixed values of at 2 but are dynamically estimated by the fitness function of the particles. We divide them into four levels. When the particle fitness is greater than g_{best} , this indicates that the particle is close to the optimal solution. At this point, we reduce the value of the particle learning factor $[0, 1]$ and improve the local search ability in order to search for the optimal solution faster. When the fitness of a particle is low, the distance between the particle and the optimal solution is very far. We force particles to take steps out of the local area (c_1 and c_2 take larger values) to search for the best solution faster.

To overcome the shortcomings of the typical linear decrease strategy, the inertia weighting factor in this paper adopts the linear differential decrement method, and its expression is as follows:

$$\begin{aligned} \frac{dw}{dt} &= -\frac{2(w_{\max} - w_{\min})}{T_{\max}} \times t, \\ w_c &= w_{\max} - \frac{(w_{\max} - w_{\min})}{T_{\max}^2} \times t^2, \end{aligned} \quad (21)$$

where t is the current number of search rounds, T_{\max} is the maximum number of search rounds, w_{\max} is the initial

inertia weight value, and w_{\min} is the inertia weight value at the maximum number of search rounds. In this paper, w_{\max} and w_{\min} take values of 0.9 and 0.4, respectively. In the previous linear decrement strategy, the rate of decline of w_c is linear. That is, the presearch and postsearch decline speeds of w_c are the same. When w_c adopts the differential decrement method, the trend of decreasing this value in the PSO algorithm in the early search is slower, and the global search ability is stronger. The downward trend in the search period accelerates, and the local search ability is strengthened. To some extent, the differential decrement method overcomes the limitations of the typical linear decrement strategy and accelerates the convergence of the PSO algorithm. The specific optimization results will be analyzed with the simulation experiment results.

After the optimal number of CHs is determined, the optimization protocol will take the CHs as the core and optimize the minimum energy consumption of the overall network and the maximum life cycle of the nodes.

To achieve the goal of saving energy, the following three values are minimized: the residual energy of non-CH nodes, the distance between non-CH nodes and CH nodes, and the number of hops from a CH to the BS in multihop mode.

The fitness function of the CH node can be expressed as

$$\begin{aligned} F_t &= \mu_1 \times E_{\text{residual-non-CH}} + \mu_2 \times D_{\text{to CH}_i} + \mu_3 \times H_{\text{to BS}}, \\ \mu_1 + \mu_2 + \mu_3 &= 1, \end{aligned} \quad (22)$$

where μ_1, μ_2 , and μ_3 are the control parameters of the three parts of the fitness function. In this protocol, the equivalent values of the three control parameters represent the same effect of the three influencing factors. $E_{\text{residual-non-CH}}$ is the sum residual energy of the non-CH nodes, and its expression is

$$E_{\text{residual-non-CH}}(t) = E_{\text{residual-non-CH}}(t-1) - \sum_{i=1}^H \sum_{j=1}^{a-1} E_{\text{non-CH}_{ji}}(t), \quad (23)$$

where $E_{\text{non-CH}_j}$ is the energy consumption of the j -th non-CH node in the i -th cluster at time t .

$$D_{\text{to CH}_i} = \sum_{j=1}^{a-1} d_{\text{to CH}_j}, \quad (24)$$

where $d_{\text{to CH}_j}$ is the distance from the j -th non-CH node to the i -th CH node.

$$H_{\text{to BS}} = \sum_{i=1}^H H_{\text{to BS}_i}, \quad (25)$$

where $H_{\text{to BS}_i}$ represents the number of hops from the i -th CH node to the BS in multihop mode. If the distance from the CH node to the BS is beyond the threshold H_{th} , the information should be transmitted in multihop mode. Multihop mode is more energy consuming than the single-hop mode, so the number of such CH nodes should be minimized.

Because the CH node needs to undertake more work, the node with more residual energy has a greater chance of being selected as the CH node [29].

Moreover, in order to better complete the following tasks, the optimization protocol stipulates that only nodes with residual energy larger than the average residual energy have the opportunity to be selected as CH nodes. Before selecting CHs, we put the nodes with a residual energy greater than the average residual energy into a set $\Phi = \{E_{\text{residual}} \geq \bar{E}_{\text{residual}}\}$.

Then, the clustering results of each round are calculated by using the proposed PSOLB algorithm [30].

The BS will select the best CH node. Before each round, the WSN will complete the following steps:

- (i) First, neighbor nodes are discovered: each node in the network broadcasts its own ID, residual energy, distance from the neighbor node, distance from the BS, and other information to neighbor nodes
- (ii) Second, neighbor node information is updated according to received data packets
- (iii) Third, the cluster configuration is broadcast: after the BS completes the network configuration, the BS uses flood broadcast again to transmit the configuration to all nodes. It broadcasts the packet containing the configuration. Each node that receives the packet changes its state to the CH node

3.2. The EGT Algorithm. In a WSN, the choice of ideal routing is very challenging. If one path is better than the others, this may cause unbalanced competition between the paths. A well-behaved path may be used more frequently than other paths, resulting in a more crowded path and faster power consumption. Because of the limited energy resources, each node saves energy for its own benefit. Unbalanced competition between paths can lead to path congestion, higher latency, additional packet collisions, and shorter network lifetimes [31].

In this section, we analyze the unbalanced competition problem for paths from the perspective of EGT, and we model the routing in a WSN as an evolutionary anticonoordination routing game. Compared with classical game theory, EGT pays more attention to dynamic strategic changes. The decision-making process can be seen as a strategic evolution over time. The EGT algorithm can be divided into three steps for routing. The specific steps are as follows.

As shown in Figure 5, we simulate the layout of a WSN as a concentric circle. The innermost radius of the concentric circle is R , where $R = d_{th} = \sqrt{\epsilon_{fs}/\epsilon_{mp}}$. Therefore, all nodes within this circular range can directly transmit data to the BS in the single-hop mode. No clustering is carried out in this region. Thus, the overall energy consumption of the WSN is saved. In the concentric circle with a radius of $2R$, the nodes transmit data to the mounting head. The CH transmits data to the innermost concentric circle node through multiple hops and then transmits the data to the BS. Similarly, nodes in the outer concentric circle area first transfer data to the CH in the region, and then the CH in turn transfers the data to nodes in the inner concentric circle, up to the BS [32].

The method of choosing a reasonable route in multihop mode is very important. First, the node establishes a list of all the neighbor nodes, which are arranged from near to far according to the distance from the node. Each node stores five pieces of information: the node ID, the distance from the node to the neighbor node, the distance from the neighbor node to the BS, the residual energy, and the bonus value.

In general, the most efficient way for a CH to transmit data in multihop mode is to find the shortest relay node. To ensure that the global shortest path can be found every time, we use the A^* algorithm to select the relay node. The A^* (A -star) algorithm is the most effective direct search method to solve the shortest path in a static road network, and it is also a common heuristic algorithm for many other problems [33].

$$f(n) = g(n) + h(n), \quad (26)$$

where $f(n)$ is the cost estimation from the initial state to the target state via the state n , $g(n)$ is the actual cost from the initial state to the state n in the state space, and $h(n)$ is the cost estimation of the optimal path from the state n to the target state.

Now that we know the distance from the node to the neighbor and the distance from the neighbor to the BS, we can calculate the shortest path from the current node to the BS.

As shown in Figure 6, in many cases, we will face the problem of routing, that is, two CH nodes will transmit data through the same relay node. The relay node will transfer a large amount of data. When a node is a relay node of multiple CH nodes at the same time, it may easily die because of excessive energy consumption, which is not conducive to a good lifetime for the entire WSN.

EGT is a powerful mathematical tool that models strategic interaction and analysis of competition, conflict, and cooperation with multiple entities.

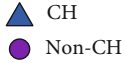
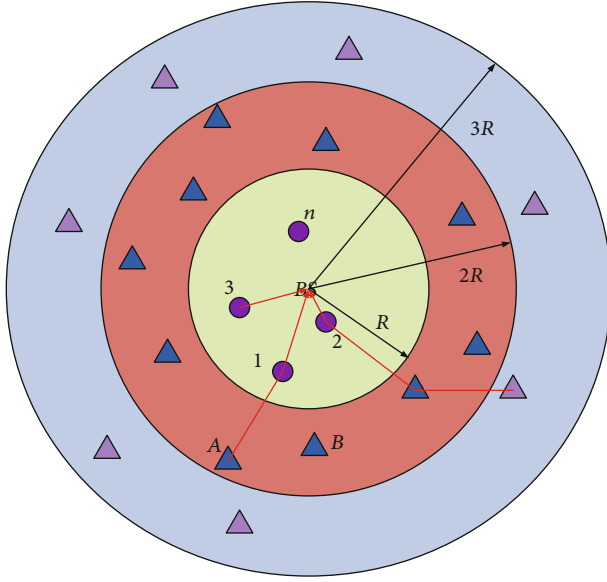


FIGURE 5: Combination of single hop and multihop.

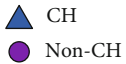
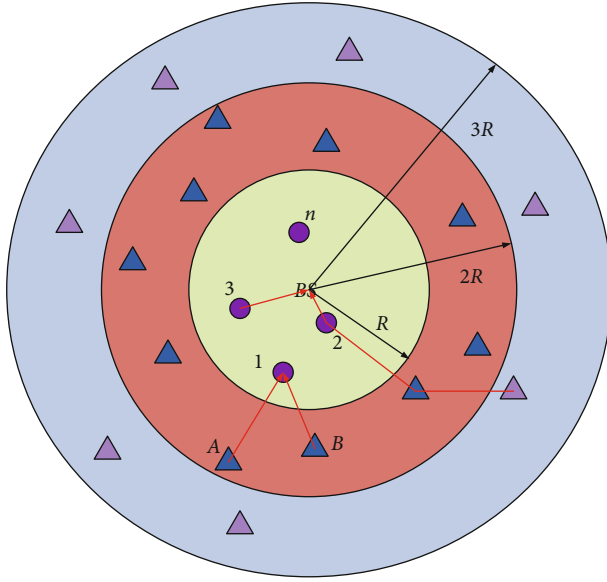


FIGURE 6: Two nodes find the same relay node.

In this section, EGT is selected to study routing congestion in the WSN. An evolutionary game does not require game participants to have complete rationality, but participants with limited rationality can make the whole group they compose reach equilibrium step by step through learning from each other [34].

It can be seen from the previous derivation that the difference form of the EGT replication dynamic process can be

expressed by equation (5). In this paper, $X_1(t)$ represents the proportion of CH nodes in the second concentric circle region that have node 1 selected as the relay node at the same time [35].

$$X_1^A(t) = \frac{N_{\text{node 1 as relay}}}{N_{A \text{ located}}}, \quad (27)$$

where $N_{\text{node 1 as relay}}$ represents the number of nodes that select node 1 as the relay node. N is the total number of nodes in the concentric circle area where node A is located.

$$X_i^N(t) = \frac{N_{\text{node } N \text{ as relay}}}{N_{A \text{ located}}}, \quad (28)$$

where $N_{\text{node } i \text{ as relay}}$ represents the number of nodes that selects node i as their relay node.

The utility function of node i at time t being selected as a relay node is expressed as

$$U_i^N(t) = q_1 \times \frac{1}{E_{tx_i}} + q_2 \times \frac{E_{\text{residual-non-CH}_i}}{E_{\text{initial}}} - q_3 \times \frac{N_{\text{node } i \text{ as relay}} - 1}{N_{\text{node } i \text{ as relay}}}, \quad (29)$$

where E_{tx_i} represents the energy of data transmission through node i : the shorter the transmission distance, the higher the benefit. $E_{\text{residual-non-CH}_i}/E_{\text{initial}}$ is the percentage of the residual energy of relay node i . E_{initial} is the initial energy of the node. The higher the percentage of residual energy, the higher the benefit. We encourage CH nodes to select nodes with high residual energy as the relay nodes. $(N_{\text{node } i \text{ as relay}} - 1)/N_{\text{node } i \text{ as relay}}$ is the penalty parameter. It can be seen from equation (27) that when more nodes select the same relay node at the same time, the penalty parameter increases exponentially. q_1 , q_2 , and q_3 are three weight values that can be adjusted according to actual conditions. In this section, we set the three weight values as follows:

$$q_1 = q_2 = q_3 \approx 0.33, \quad (30)$$

$$X_i^N = \frac{dx}{dt} = \frac{N_{\text{node 1 as relay}}}{N_{i \text{ located}}} \times (U_i^N(t) - \bar{U}(t)) = \frac{N_{\text{node 1 as relay}}}{N_{i \text{ located}}} \times \left(\left(q_1 \times \frac{1}{E_{tx_i}} + q_2 \times \frac{E_{\text{residual-non-CH}_i}}{E_{\text{initial}}} - q_3 \times \frac{N_{\text{node } i \text{ as relay}} - 1}{N_{\text{node } i \text{ as relay}}} \right) - \bar{U}(t) \right), \quad (31)$$

where $\bar{U}(t)$ is the average utility function, and it is expressed by

$$\bar{U}(t) = \sum_{N_{\text{relay}}=1}^{N_{\text{relay}}=i} X_i^N(t) \times U_i^N(t). \quad (32)$$

As seen from Figure 6, when the CH node A selects node 1 as the relay node, the revenue of the CH node is relatively large. When the utility value is greater than the average utility, the CH node A will not change its policy but will continue to

select node 1 as the relay node. When the CH node B also selects node 1 as the relay node, the utility values of CHs A and B are reduced. Therefore, CH A or B will change its policy and select another node as the relay node.

After several evolutions, the game will eventually converge to the evolutionary stable strategy ESS determined by the following equation.

$$X_i^N = 0. \quad (33)$$

3.3. Algorithm Convergence Analysis. After all CHs select a node as the relay node to transmit data, each CH evaluates its own utility according to its own utility function and then compares the obtained utility with the average utility of the group: if its own utility is smaller than the average utility of the group, it randomly chooses to access other relay nodes according to a certain probability; if not, it will keep the previous access network unchanged. See the appendix for a detailed description.

3.4. Analysis of Algorithm Time Complexity. Two important indicators to measure the performance of the optimization algorithm are the accuracy and the speed of the algorithm in solving the problem. The accuracy analysis of the PSOLB-EG protocol has been given in the previous sections. The speed of the algorithm is affected by the time complexity of the algorithm. Therefore, we will focus on the time complexity of each algorithm and give the overall time complexity of each algorithm.

Standard PSO algorithms include processes such as particle swarm initialization, search updates, and particle swarm updates. The computational complexity of the algorithm is $O(N \times I \times D)$, where N is the number of particles, I is the maximum update algebra, and D is the target dimension. However, the protocol we propose is to cluster at nodes other than the BS threshold. The fixed learning factor value is changed to the adaptive value, which increases the comparison time $O(N_{\overline{BS}} \times I \times D + N_{\overline{BS}} \times Ct)$ of the fitness function and the target value; $N_{\overline{BS}}$ represents nodes other than the BS threshold and Ct represents the comparison time. In the same way, all the particles in the process of each round of search have learned the best results of the last round, so the total PSOLB time complexity is $O(N_{\overline{BS}} \times I \times D + N_{\overline{BS}} \times (Ct + Lt))$, where Lt is the learning time of each particle. Because $N_{\overline{BS}}$ is less than N , the proposed protocol is less complex than the standard PSO algorithm.

In terms of data transmission routing, the time complexity of the general exhaustion algorithm is $O(N_{CH} \times N_{CH\text{-neighbor}} \times N_{\text{multihop}})$, where N_{CH} represents the number of cluster heads, $N_{CH\text{-neighbor}}$ represents the number of neighbors of the cluster head, and N_{multihop} represents the number of hops from the cluster head to the BS. The proposed EGT algorithm establishes a cluster head route, and its time complexity becomes $O(N_{\overline{BSCH}} \times N_{\text{upper}} \times N_{\text{multihop}})$, where $N_{\overline{BSCH}}$ represents the number of cluster heads outside the BS threshold, and N_{upper} represents the number of nodes in the upper layer. Because $N_{\overline{BSCH}} < N_{CH}$ and $N_{\text{upper}} < N_{CH\text{-neighbor}}$, the time complexity

of the EGT algorithm is less than that of the general exhaustive algorithm.

Through the above analysis, the PSOLB algorithm takes slightly more time than the standard PSO algorithm, but it is better in terms of search accuracy, so the PSOLB algorithm is more energy efficient than the standard PSO algorithm. Regarding the routing algorithm, the EGT algorithm is superior to other search methods in solving accuracy and time. The overall PSOLB-EGT time complexity is superior to that of the standard PSO algorithm as well as the greedy routing algorithm and some other cluster routing algorithms. The above conclusions will be specifically analyzed in the simulation experiment.

3.5. Analysis of Algorithm Space Complexity. Space complexity is mainly used to measure the storage space occupied by the algorithm. We analyze the space complexity of the hybrid algorithm from two aspects: PSOLB algorithm and EGT algorithm. We assume that each population has N particles, the search space is D -dimensional, and there are a total of M populations. For the standard particle swarm algorithm, the storage space required to store the particle position and velocity is $O(2NDM)$, and the space occupied by the P_{best} of each particle is $O(NDM)$, so the space complexity of the standard particle swarm algorithm is $O(3NDM)$. For the PSOLB algorithm, in addition to the particle position, velocity, and storage space required, each particle needs to store the optimal position of the previous round, so the position space of $O(NDM)$ needs to be increased. In addition, consider that the space for storing the dynamic learning step length is $O(ND)$. Therefore, the total space complexity of the LB-PSO algorithm is $O(4NDM + ND)$.

In the WSN data transmission stage, the space complexity of the EGT algorithm is as follows: assuming that the number of nodes within the BS threshold is V , this part of the nodes directly transmits data to the BS, and its space complexity does not need to be calculated compared to the original algorithm. The current node uses the EGT algorithm to calculate the storage space of the fitness function value of each neighbor node as A , the storage space for the average fitness of the remaining nodes is B , and the storage space for the temporary variables of the improved EGT algorithm is C . Then, the worst space complexity of the improved EGT algorithm is $O((N - V) \times (A + B) + C)$. This has lower space complexity than the standard EGT algorithm.

Thus, it is the LB-PSO algorithm space rather than the standard PSO complexity that increases, but the EGT-improved algorithm reduces the complexity of the algorithm rather than the original space. Therefore, the space complexity of the hybrid algorithm is basically equivalent to the original standard algorithm.

4. System Simulation Analysis

In this section, we use numerical simulation to evaluate the performance of our algorithm PSOLB-EGT. Assume that the network coverage is $200 \text{ m} \times 200 \text{ m}$, where the BS coordinates are located at $(100, 100)$, and 100 to 400 sensor nodes are randomly distributed. The simulation parameters used

in this paper, which accord with those in references [14, 16, 22], are shown in Table 1, and all simulation models and algorithms are coded in MATLAB 2015b.

To evaluate the performance of the algorithm, we compare the clustering and routing of the PSOLB-EGT protocol with those of PSO-C, ABC-SD, TPSO-CR, and JCR. TPSO-CR is a two-tier PSO algorithm which is proposed to solve the problem of clustering and routing in WSN [36]. The ABC-SD algorithm exploits the biologically inspired fast and efficient searching features of the Artificial Bee Colony metaheuristic to build low-power clusters [37]. JCR in terms of energy consumption is the clustered routing protocol proposed in 2016 [30, 31]. Since PSO-C does not include a multihop routing protocol, it adopts a greedy routing algorithm. In the greedy routing algorithm, every node is assumed to know its distance to neighbor nodes. CH selects its nearest neighbor node as a relay node [38].

In addition, we use energy consumption, lifetime, residual energy (number of surviving nodes), packet loss rate, and node capacity (throughput) as indicators in simulating and comparing the four protocols. The throughput capacity of the WSN refers to the total data rate (or the total number of data) sent on the network, that is, the data rate (or the number of data) sent by the CHs to the BS and the data rate (or the number of data) sent from the nodes to the CHs. The size of the throughput is often used in network protocols to indicate the performance of the network. Therefore, throughput is often used as one of the evaluation criteria for routing protocols [39].

In the scenario where the network model is uniformly distributed for N sensor nodes, Michael and Martin in [40] have deduced the upper and lower bounds of the capacity in the application case. When the number of nodes N in the network tends to infinity, the upper and lower boundaries coincide and the conclusion is that the capacity is $\Theta(\log(N))$. The throughput capacity of a single node is $\Theta(\log(N)/N)$. For wireless sensor networks with limited node energy, throughput capacity and available throughput for a single node are important performance metrics [34, 35]. We will prove its importance in later simulations.

On the other hand, the node capacity of a WSN also refers to the ability of the node to forward the amount of data. An important cause of node congestion in WSNs is that the load is greater than the maximum capacity of the node. Usually the WSN node load refers to the amount of information that a node needs to forward at a certain moment. Since the WSN node is limited by hardware resources, the node has a fixed capacity, and if the load exceeds its capacity at a certain time, congestion will occur [41].

$$\omega = \begin{cases} 1, & \text{if } Co > Li = \sum_1^k lki, \\ 0, & \text{otherwise,} \end{cases} \quad (34)$$

$$\eta = \frac{L_i}{C_o}, \quad (35)$$

where Co represents the capacity of node i , L_i represents the load of node i , k represents the number of all neighbor nodes

TABLE 1: Relay node status list.

Node	Relay node list	η	Relay node status
A	1	65%	Live
A	2	70%	Live
A	3	50%	Live

TABLE 2: Relay node status list.

Node	Relay node list	η	Relay node status
A	1	0	Dead
A	2	70%	Live
A	3	50%	Live

TABLE 3: Simulation input parameters.

Parameters	Value
Sensor field region (m ²)	200 * 200
BS location	100, 100
Number of sensors	100, 200, 300, 400
Optimal number of clusters	5, 10, 15, 20
Initial energy of the node (J)	200
Data packet length (bits)	4096
Energy data aggregation (nJ)	5
Number of rounds	2000
Swarm size	25
Particle position	[0, 200]
Node capacity (kbits)	0.1
E_{elec} (nJ/bit)	50
E_{amp} (pJ/bit)	0.0012
E_{fs} (pJ/bit)	10
d_{th} (m)	30

of the node, l_{ki} indicates the amount of data transmitted by the neighbor node k to node i , and η indicates the node capacity occupancy. In the subsequent simulation experiments, we will analyze node capacity in detail [42].

In the PSOLB-EGT protocol, congestion is solved by the following steps: establishing a list of next-hop spare relay nodes, detecting node congestion, establishing an alternate data transmission path, and restoring the original transmission path after congestion is released.

In the PSOLB-EGT protocol, congestion is solved by the following steps: establishing a list of next-hop spare relay nodes, node congestion detection, establishing an alternate data transmission path, and restoring the original transmission path after congestion is released.

As shown in Figure 6, we use equation (29) to calculate the utility function values for each possible relay node of the node and sort them accordingly. For example, the utility function values of the next-hop relay nodes of node A are arranged in the order of node 1 > node 2 > node

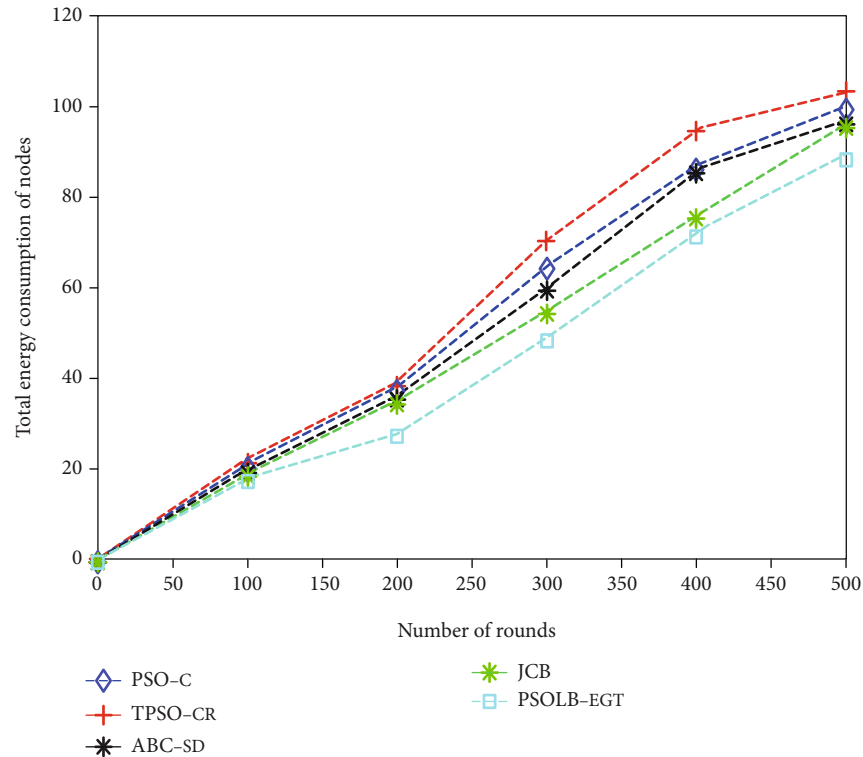


FIGURE 7: Comparison of node energy consumption after running the same cycle for each protocol.

We establish a list of the next-hop relay node statuses for node A. As shown in Table 1, the utility function values of each possible successor node in the next hop of node A are sequentially sorted, and information about their node capacity occupancy, survival, etc. is displayed

According to equations (33) and (35), it is judged whether the relay node in the list is congested. For example, node 1 is not congested and continues to be used as a relay node

If node 1 is congested, node A will automatically elect node 2 as the relay node [43]

In the next round of data transmission, if node 1 congestion is removed, node 1 continues to be used as the relay node. If a node consumes too much energy and dies, the node state changes from live to dead. As shown in Table 2, when relay node 1 dies, node A selects node 2 as the relay node [44]

To test our algorithm, we considered an initial population of 50 particles and let them evolve 2000 times. The parameters required for the experiment are shown in Table 3.

The comparison of the energy consumption of various protocols under the same running cycle is shown in Figure 7. Clearly, the PSOLB-EGT protocol shows the best energy optimization results for the WSN.

We changed the position of the BS in the sensor layout area, as shown in Figure 8. When the BS was in different locations, the total energy consumption of the WSN was compared after running various protocols at the same time for 100 rounds. It can be observed that the PSOLB-EGT protocol is superior to existing protocols in all network scenarios. This is because the PSOLB-EGT protocol uses a more suitable path to transfer aggregate data from the CH to the BS.

We ran five protocols and found that the PSOLB-EGT protocol is superior to the other four protocols when comparing the change in the number of alive nodes in each protocol from 100 to 0. As shown in Figure 9, the PSOLB-EGT protocol is later than other protocols in terms of the first and last node death times. This shows that the protocol can effectively extend the life cycle of WSN.

For the same layout area, with an increase in the number of sensors, each protocol runs for 500 rounds, and we observe that the total residual energy also increases with the increase in the number of sensors. The comparison of residual energy of various protocols is shown in Figure 10.

Through the above simulation experiment, we can observe the following:

- (1) As shown in Figure 7, all other conditions being equal, the PSOLB-EGT protocol consumes less energy than the other four protocols when running the same rounds. It can be inferred that when WSN first started 100 rounds, the energy consumption gap of several protocols was small. As time goes by, the PSOLB-EGT protocol consumes less energy than other protocols
- (2) The results in Figure 8 show that, by changing the location of the BS, the PSOLB-EGT protocol has better scalability than the other protocols
- (3) The PSOLB-EGT protocol has greater advantages than the other four protocols in terms of the first death of nodes. Figure 9 shows that after WSN runs the PSOLB-EGT protocol, its first dead node appears

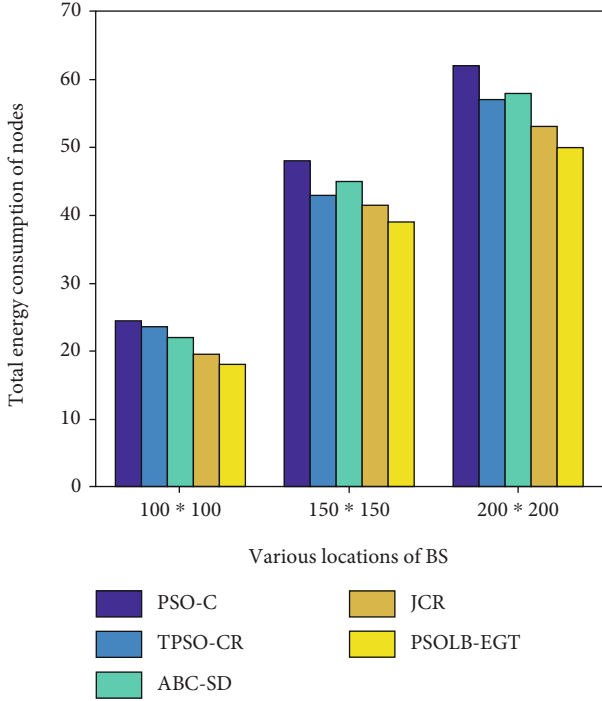


FIGURE 8: Node energy consumption of various protocols when the BS is in different locations.

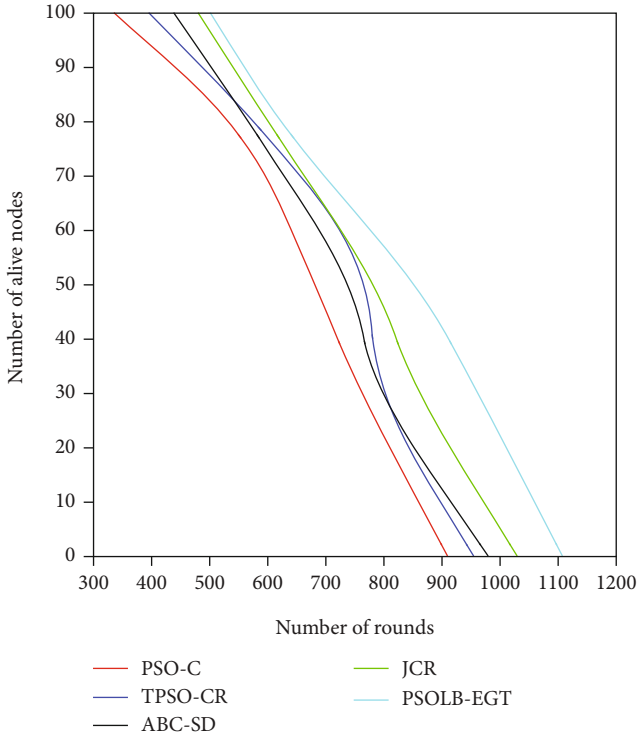


FIGURE 9: Live nodes in the network for various algorithms.

later than other protocols. As the running time of WSN goes by, the node mortality and total running time of the entire WSN are better than those of other protocols

- (4) As the runtime increases, the data transmitted by the WSN continues to increase. Under the same conditions, the WSN uses the traditional algorithm and the improved PSOLB-EGT protocol to perform the node capacity experiment. From the perspective of the average occupancy rate of node capacity, according to the analysis of the experimental results in Figure 11, the other four algorithms increase the average occupancy rate of node capacity as the amount of communication data per unit time increases. The average occupancy rate of the node capacity of the PSOLB-EGT protocol changes slightly
- (5) Figure 12 shows the network throughput comparison between PSOLB-EGT and the other protocols. Throughput is defined here as the number of data packets successfully received at the BS during simulation time. The results represent an average of three different runs for each network size, and it is observed that PSOLB-EGT is superior to the other protocols in network throughput

We implement all protocols to evaluate the rate of packet loss. It is clear from Figure 13 that in all the network scenarios considered, the packet loss rate of the PSOLB-EGT protocol is much lower than that of other protocols. These protocols generate high rates of data loss. This is due to the absence of hybrid-hop communication.

Simulation results in different environments show that the PSOLB-EGT protocol is superior to the existing network protocols in terms of data transmission energy.

We evaluated the number of unclustered sensors per round in each network protocol. The chart in Figure 14 shows the average number of unclustered sensors per round.

As shown in Figure 15, we analyze the node capacity congestion rates of the five protocols for increasing runtimes. In this paper, the total number of nodes successfully transmitting data to the aggregation point is used to describe the node capacity congestion rate of the network as the number of rounds changes. It only makes sense for the WSN to successfully transmit the data generated by the node to the aggregation point. The amount of data successfully delivered reflects the smoothness of the network to some extent. Therefore, the goal pursued by WSN routing is to consume as little energy as possible and transmit as much data as possible to the aggregation point. As seen from Figure 15, the algorithm we proposed has better node throughput capacity.

To verify the energy-saving effectiveness of the PSOLB clustering algorithm separately, we unified all five clustering algorithms (including PSOLB) with the greedy routing algorithm. The operating conditions of the five clustering algorithms are exactly the same. It can be seen from Figure 16 that after the same number of rounds of PSOLB, the remaining energy of the entire WSN is better than it is with any of the other four clustering algorithms. The main

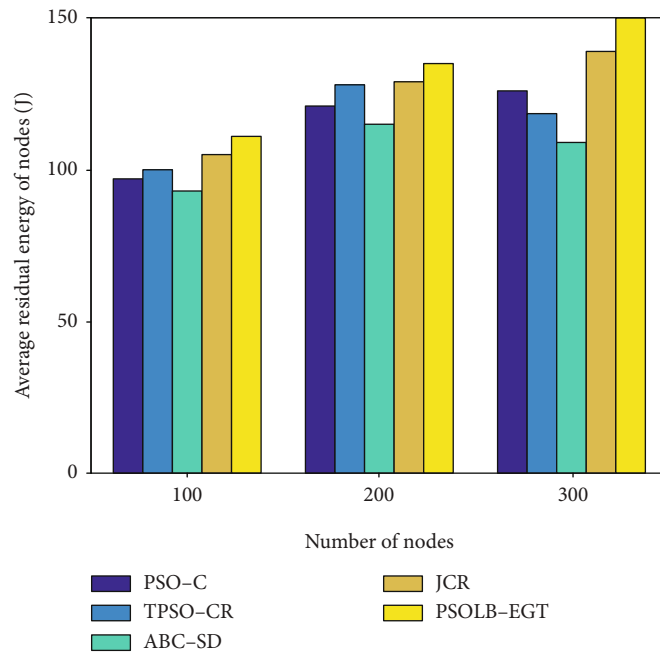


FIGURE 10: Comparison of changes in energy consumption when the number of nodes increases.

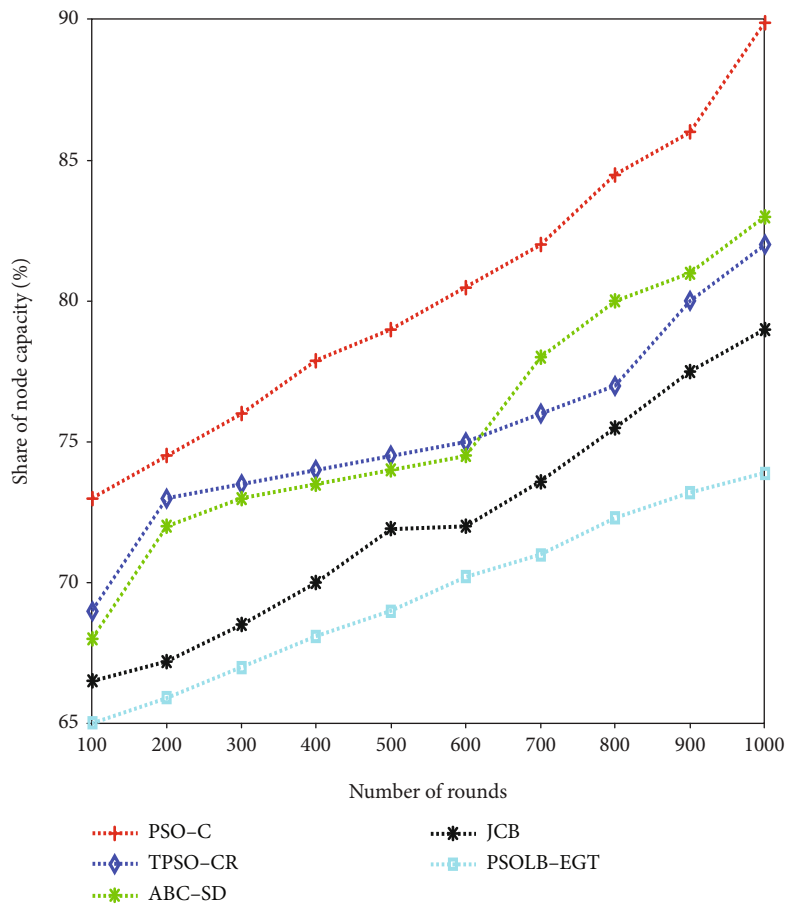


FIGURE 11: Comparison of average occupancy rates of different algorithm nodes.

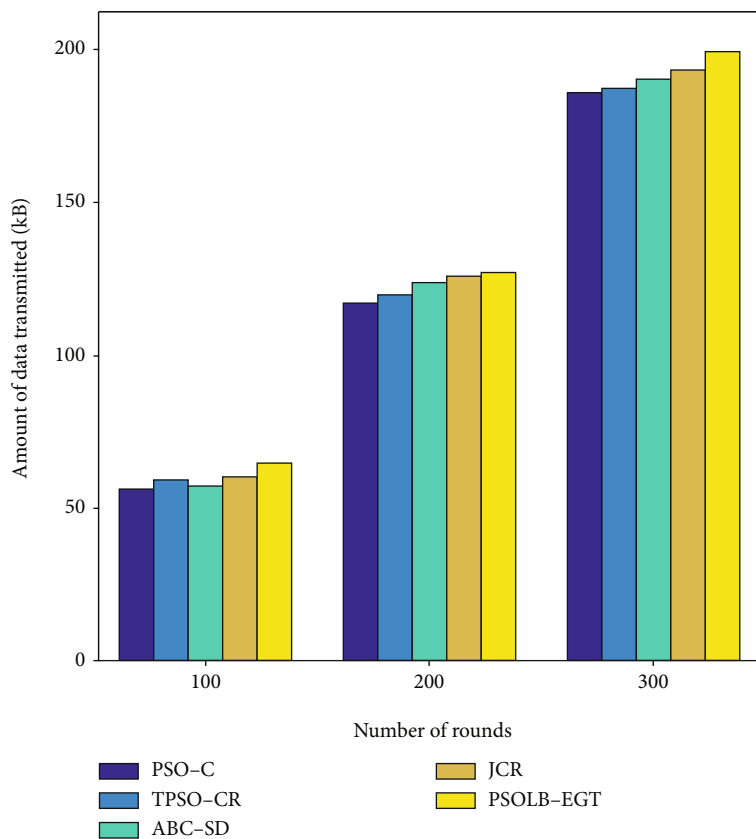


FIGURE 12: Comparison of data transmission amount under the same cycle condition.

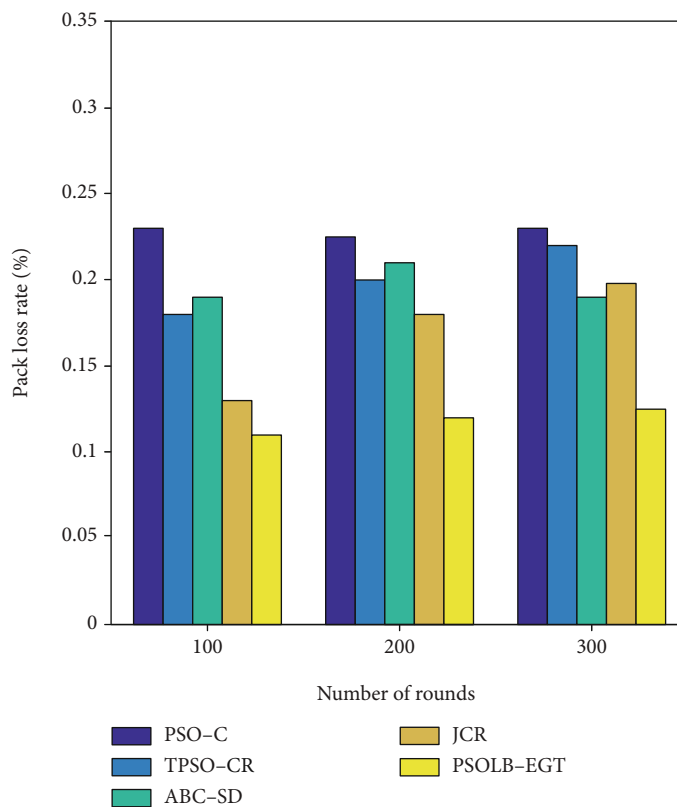


FIGURE 13: Pack loss rate.

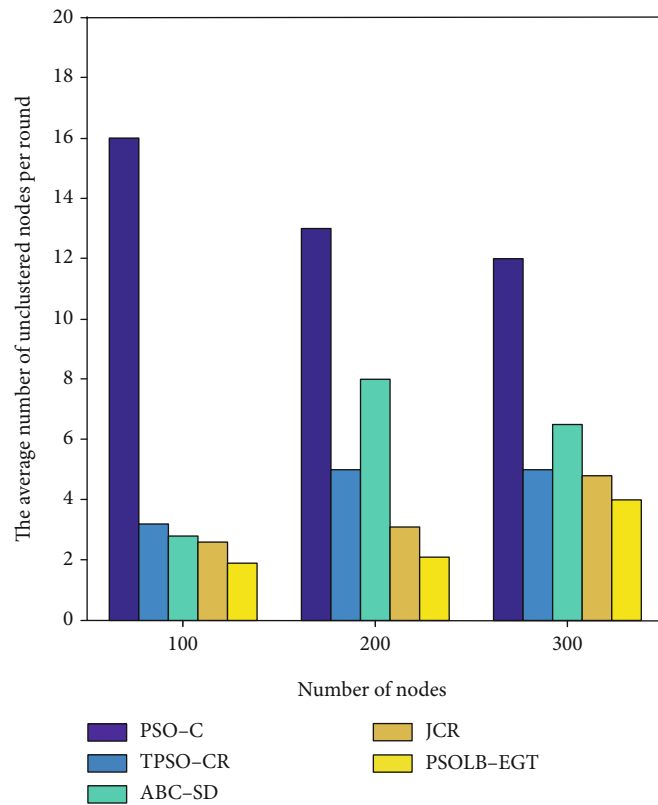


FIGURE 14: Amount of unclustered sensors.

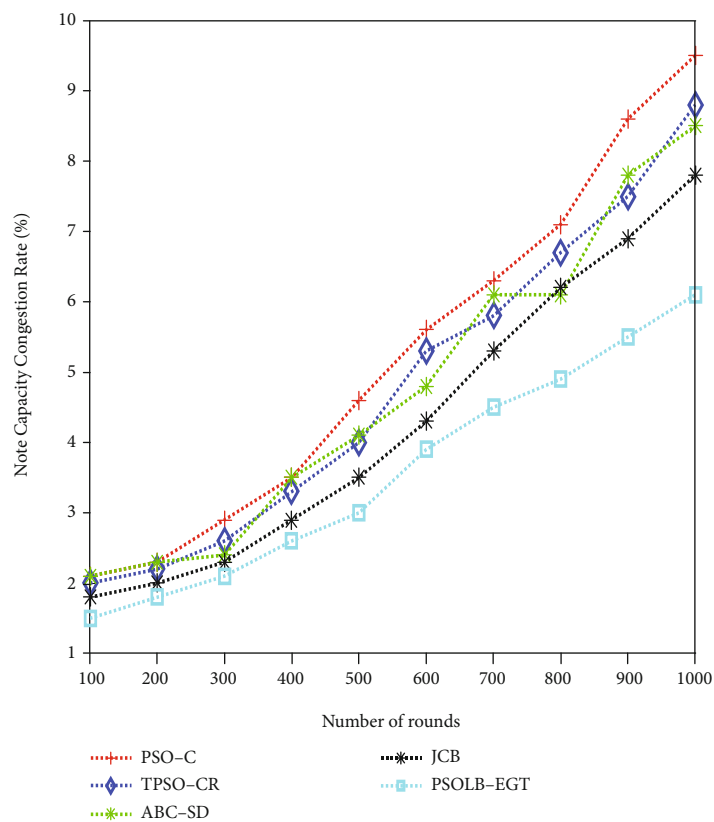


FIGURE 15: Node capacity congestion rate.

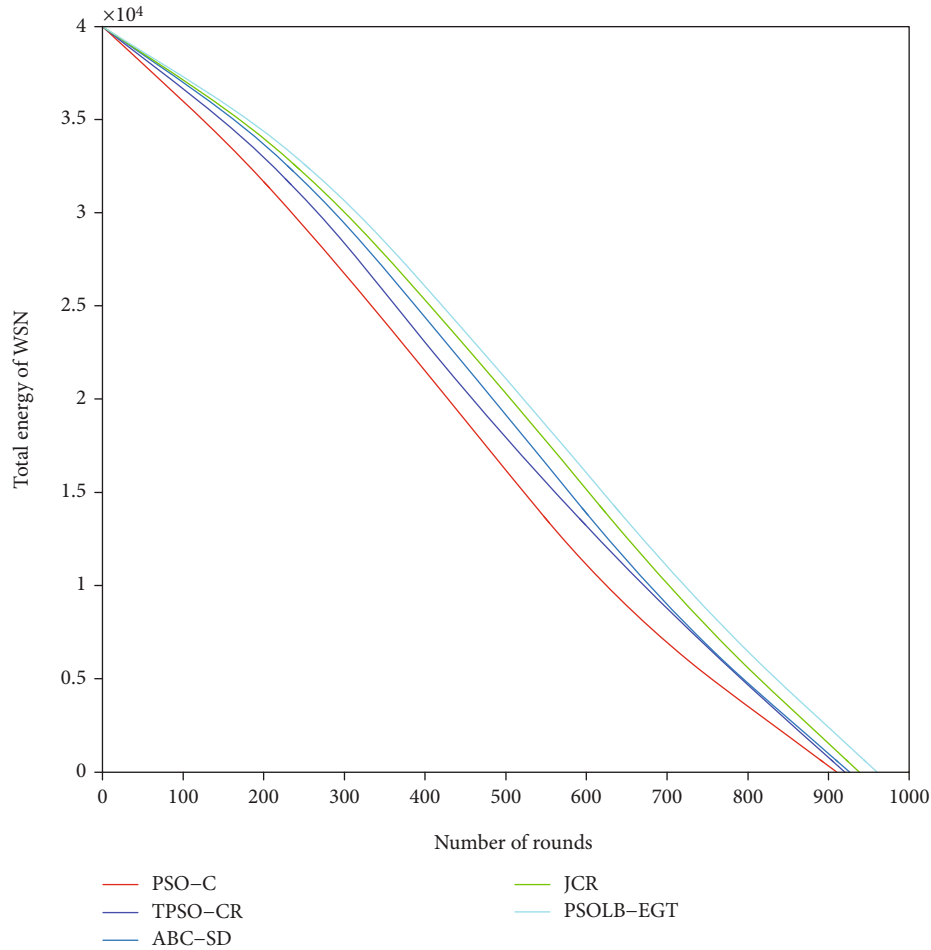


FIGURE 16: Comparison of total WSN energy consumption after using five clustering algorithms plus greedy routing.

effect is that we use adaptive learning factors and optimization measures for the last round of optimal results. However, it can be seen that when the clustering algorithms of JCR and PSOLB use the greedy routing algorithm, the energy consumption of the entire WSN is accelerated.

5. Conclusion

In this paper, we use the improved PSO algorithm and EGT, respectively, to solve two well-known optimization problems of WSN, namely, selection of CHs and routing between the CHs and the receiver. Then, we propose a clustering and routing protocol called PSOLB-EGT. The protocol incorporates an improved CH selection algorithm based on PSO search, which has a better fitness function. Next, an improved routing algorithm based on EGT is proposed to transmit aggregate data from CHs to the receiver for large-scale WSNs. This algorithm uses a novel routing function. The simulation results show that the proposed protocol is superior to existing protocols in the network life cycle, network coverage, and packet transmission capacity [45].

The existing research still has many shortcomings to be further addressed. The prospects for follow-up work include the following:

The routing protocols and positioning techniques studied in this paper are limited to two-dimensional plane space, but three-dimensional space is more in line with the actual application environments. It is important to study the routing and positioning technology of 3D wireless sensors. Therefore, it is necessary to extend the scope of research from two-dimensional space to three-dimensional space and perform further research on clustering routing technology for 3D WSNs.

In addition, a node needs to obtain its own global geographic information when clustering. Therefore, the proposed protocol requires the node to be equipped with a positioning device such as GPS, which increases the hardware requirements of the node to some extent.

The clustering routing algorithm proposed in this paper only assumes that the nodes in the cluster can synchronously send and receive data after receiving the cluster head broadcast and sending data packets to the CH. The corresponding synchronization mechanism is not designed. Therefore, the problem of how to design a reliable and practical mechanism that can synchronize communication between nodes and reduce the communication overhead is a subject that needs to be further studied in sensor network applications.

In the future, we will further optimize the PSOLB-EGT protocol according to the experimental results to improve

the energy saving effect of the protocol. We will study the energy saving optimization of data fusion, another important part of WSNs.

Appendix

Proof of Convergence

When the dynamic process of node routing replication continues, there will be a certain moment when the net utility of each cluster in the same group is the same, and the net utility at this moment is the average net utility of each individual in the group. That is, individual players no longer adjust their strategies to improve their net utility. To calculate the information of each participant and obtain the average utility of the group, it is necessary to set up a central control entity to maintain the information of the participants in the system and inform the participants of the average utility of their group at that time.

According to formulas (29) and (30), $X_i^N = 0$, that is, the equilibrium point of the evolutionary game.

$$\left[q_1 \frac{1}{E_{(Si)}} + q_2 \frac{E_{\text{residual-non-CH}_i}}{E_{\text{initial}}} - q_3 (n_r - 1)^{n_r - 1} \sigma \right] = \left(\sum_1^{n_r} \frac{n_{\text{to node } n_r} \times \delta_{n_r}}{N_{\text{all}}} \right). \quad (\text{A.1})$$

According to this equation (A.1), at this time, the fitness value of a single node is the average fitness value of all nodes in the WSN. That is, a single node no longer adjusts its own strategy to improve the fitness value, that is, the equilibrium point of the evolutionary game.

Data Availability

Please contact with the corresponding author to acquire the underlying data if necessary.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by the Zigong City Key Science and Technology Program (No. 2019YYJC16), by the Artificial Intelligence Key Laboratory of Sichuan Province (Nos. 2017RYY05 and 2018RYJ03), and by the Horizontal Project (Nos. HX2017134 and HX2018264).

References

- [1] S. Yadav and R. S. Yadav, "A review on energy efficient protocols in wireless sensor networks," *Wireless Networks*, vol. 22, no. 1, pp. 335–350, 2016.
- [2] H. Jawad, R. Nordin, S. Gharghan, A. Jawad, and M. Ismail, "Energy-efficient wireless sensor networks for precision agriculture: a review," *Sensors*, vol. 17, no. 8, p. 1781, 2017.
- [3] W. Rehan, S. Fischer, and M. Rehan, "A critical review of surveys emphasizing on routing in wireless sensor networks-an anatomization under general survey design framework," *Sensors*, vol. 17, no. 8, p. 1713, 2017.
- [4] F. Yaghoubi, A. Abbasfar, and B. Maham, "Energy-efficient RSSI-based localization for wireless sensor networks," *IEEE Communications Letters*, vol. 17, no. 6, pp. 973–976, 2014.
- [5] H. Nabizadeh and M. Abbaspour, "IFRP: an intrusion/fault tolerant routing protocol for increasing resiliency and reliability in wireless sensor networks," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 14, no. 1, pp. 52–69, 2013.
- [6] Asaduzzaman and H. Y. Kong, "Energy efficient cooperative LEACH protocol for wireless sensor networks," *Journal of Communications and Networks*, vol. 12, no. 4, pp. 358–365, 2010.
- [7] F. Gao, H. Wen, L. Zhao, and Y. Chen, "Design and optimization of a cross-layer routing protocol for multi-hop wireless sensor networks," in *International Conference on Sensor Network Security Technology & Privacy Communication System IEEE*, Nangang, China, May 2013.
- [8] D. Gao, H. Lin, and X. Liu, "Routing protocol for k-anycast communication in rechargeable wireless sensor networks," *Computer Standards & Interfaces*, vol. 43, pp. 12–20, 2016.
- [9] S. Lee, Y. Noh, and K. Kim, "Key schemes for security enhanced TEEN routing protocol in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 9, no. 6, Article ID 391986, 2013.
- [10] Y. Ge, S. Wang, and J. Ma, "Optimization on TEEN routing protocol in cognitive wireless sensor network," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, 27 pages, 2018.
- [11] Y. Zhou, N. Wang, and W. Xiang, "Clustering hierarchy protocol in wireless sensor networks using an improved PSO algorithm," *IEEE Access*, vol. 5, pp. 2241–2253, 2017.
- [12] D. R. Prasad, P. V. Naganjaneyulu, and K. S. Prasad, "Energy efficient clustering in multi-hop wireless sensor networks using differential evolutionary MOPSO," *Brazilian Archives of Biology & Technology*, vol. 59, no. spe2, pp. 1–16, 2016.
- [13] D. Lin, Q. Wang, D. Lin, and Y. Deng, "An energy-efficient clustering routing protocol based on evolutionary game theory in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 11, no. 11, 2015.
- [14] M. S. Kordafshari, A. Movaghar, and M. R. Meybodi, "A joint duty cycle scheduling and energy aware routing approach based on evolutionary game for wireless sensor networks," *Iranian Journal of Fuzzy Systems*, vol. 14, no. 2, pp. 23–44, 2017.
- [15] X. Liu, "A survey on clustering routing protocols in wireless sensor networks," *Sensors*, vol. 12, no. 8, pp. 11113–11153, 2012.
- [16] N. A. Pantazis, S. A. Nikolidakis, and D. D. Vergados, "Energy-efficient routing protocols in wireless sensor networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 551–591, 2013.
- [17] Y. Jin, L. Wang, Y. Kim, and X. Yang, "EEMC: an energy-efficient multi-level clustering algorithm for large-scale wireless sensor networks," *Computer Networks the International Journal of Computer & Telecommunications Networking*, vol. 52, no. 3, pp. 542–562, 2008.
- [18] Y. Jin, L. Wang, Y. Kim, and X.-Z. Yang, "Energy efficient non-uniform clustering division scheme in wireless sensor networks," *Wireless Personal Communications*, vol. 45, no. 1, pp. 31–43, 2008.

- [19] K. Munusamy, "Distributed energy efficient clustering algorithm for wireless sensor networks," *Journal of Microelectronics, Electronic Components and Materials*, vol. 45, no. 3, pp. 180–187, 2015.
- [20] M. Nesa Sudha, M. L. Valarmathi, and A. S. Babu, "Energy efficient data transmission in automatic irrigation system using wireless sensor networks," *Computers & Electronics in Agriculture*, vol. 78, no. 2, pp. 215–221, 2011.
- [21] J. Hao, G. Duan, B. Zhang, and C. Li, "An energy-efficient on-demand multicast routing protocol for wireless ad hoc and sensor networks," in *GLOBECOM 2013 - 2013 IEEE Global Communications Conference*, pp. 4650–4655, Atlanta, GA, USA, December 2013.
- [22] X. L. He, X. W. Fu, and Y. S. Yang, "Energy-efficient trajectory planning algorithm based on multi-objective PSO for the mobile sink in wireless sensor networks," *IEEE Access*, vol. 7, pp. 176204–176217, 2019.
- [23] N. Gharaei, K. A. Bakar, S. Z. M. Hashim, and A. H. Pourasl, "Energy-efficient intra-cluster routing algorithm to enhance the coverage time of wireless sensor networks," *IEEE Sensors Journal*, vol. 19, no. 12, pp. 4501–4508, 2019.
- [24] F. Fanian and M. Kuchaki Rafsanjani, "Cluster-based routing protocols in wireless sensor networks: a survey based on methodology," *Journal of Network and Computer Applications*, vol. 142, pp. 111–142, 2019.
- [25] S. Chowdhury and C. Giri, "Energy and network balanced distributed clustering in wireless sensor network," *Wireless Personal Communications*, vol. 105, no. 3, pp. 1083–1109, 2019.
- [26] H. Fariborzi and M. Moghavvemi, "EAMTR: energy aware multi-tree routing for wireless sensor networks," *IET Communications*, vol. 3, no. 5, pp. 733–739, 2009.
- [27] J. Hong, J. Kook, S. Lee, D. Kwon, and S. Yi, "T-LEACH: the method of threshold-based cluster head replacement for wireless sensor networks," *Information Systems Frontiers*, vol. 11, no. 5, pp. 513–521, 2009.
- [28] S. Suganthi and S. P. Rajagopalan, "Multi-swarm particle swarm optimization for energy-effective clustering in wireless sensor networks," *Wireless Personal Communications*, vol. 94, no. 4, pp. 2487–2497, 2017.
- [29] J. Albath, M. Thakur, and S. Madria, "Energy constraint clustering algorithms for wireless sensor networks," *Ad Hoc Networks*, vol. 11, no. 8, pp. 2512–2525, 2013.
- [30] Y. X. Peng, Y. H. Li, L. Shu, and W. B. Wang, "An energy-efficient clustered distributed coding for large-scale wireless sensor networks," *Journal of Supercomputing*, vol. 66, no. 2, pp. 649–669, 2013.
- [31] Q. W. Liu and M. D. Liu, "Energy-efficient clustering algorithm based on game theory for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 13, no. 11, 2017.
- [32] A. M. Shamsan Saleh, B. Mohd Ali, M. F. A. Rasid, and A. Ismail, "A self-optimizing scheme for energy balanced routing in wireless sensor networks using SensorAnt," *Sensors*, vol. 12, no. 8, pp. 11307–11333, 2012.
- [33] I. C. Paschalidis and R. M. Wu, "Robust maximum lifetime routing and energy allocation in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 8, no. 8, Article ID 523787, 2012.
- [34] A. A. M. Rahat, R. M. Everson, and J. E. Fieldsend, "Evolutionary multi-path routing for network lifetime and robustness in wireless sensor networks," *Ad Hoc Networks*, vol. 52, pp. 130–145, 2016.
- [35] K. Saleem, A. Derhab, M. A. Orgun et al., "Cost-effective encryption-based autonomous routing protocol for efficient and secure wireless sensor networks," *Sensors*, vol. 16, no. 4, p. 460, 2016.
- [36] R. S. Y. Elhabyan and M. C. E. Yagoub, "Two-tier particle swarm optimization protocol for clustering and routing in wireless sensor network," *Journal of Network and Computer Applications*, vol. 52, pp. 116–128, 2015.
- [37] A. A. A. Ari, B. O. Yenke, N. Labraoui, I. Damakoa, and A. Gueroui, "A power efficient cluster-based routing algorithm for wireless sensor networks: honeybees swarm intelligence based approach," *Journal of Network and Computer Applications*, vol. 69, pp. 77–97, 2016.
- [38] S. Tyagi and N. Kumar, "A systematic review on clustering and routing techniques based upon LEACH protocol for wireless sensor networks," *Journal of Network and Computer Applications*, vol. 36, no. 2, pp. 623–645, 2013.
- [39] M. Sabet and H. Naji, "An energy efficient multi-level route-aware clustering algorithm for wireless sensor networks: a self-organized approach," *Computers & Electrical Engineering*, vol. 56, pp. 399–417, 2016.
- [40] M. Tekkalmaz and I. Korpeoglu, "Distributed power-source-aware routing in wireless sensor networks," *Wireless Network*, vol. 22, no. 4, pp. 1381–1399, 2016.
- [41] D. K. Sharma, S. K. Dhurandher, D. Agarwal, and K. Arora, "kROP: k-means clustering based routing protocol for opportunistic networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 4, pp. 1289–1306, 2019.
- [42] K. Haseeb, N. Islam, A. Almogren, and I. U. Din, "Intrusion prevention framework for secure routing in WSN-based mobile Internet of Things," *IEEE Access*, vol. 7, pp. 185496–185505, 2019.
- [43] K. Haseeb, N. Islam, T. Saba, A. Rehman, and Z. Mehmood, "LSDAR: a light-weight structure based data aggregation routing protocol with secure internet of things integrated next-generation sensor networks," *Sustainable Cities and Society*, vol. 54, article 101995, 2019.
- [44] K. A. Darabkh, M. Z. El-Yabroudi, and A. H. El-Mousa, "BPA-CRP: a balanced power-aware clustering and routing protocol for wireless sensor networks," *Ad Hoc Networks*, vol. 82, pp. 155–171, 2019.
- [45] A. Wang, W. B. Heinzelman, A. Sinha, and A. P. Chandrakasan, "Energy-scalable protocols for battery-operated micro-sensor networks," *Journal of Vlsi Signal Processing Systems for Signal Image & Video Technology*, vol. 29, no. 3, pp. 223–237, 2001.

Research Article

Spatiotemporal Characterization of Land Subsidence in Guandu (China) Revealed by Multisensor InSAR Observations

Wu Zhu ¹, Xue-qi Zhang ¹, Zhan-ke Liu ², and Qian Zhu ¹

¹College of Geology Engineering and Geomatics, Chang'an University, Xi'an, China

²The First Geodetic Surveying Brigade of Ministry of Natural Resources, Xi'an, China

Correspondence should be addressed to Zhan-ke Liu; lzk_111@163.com

Received 29 May 2020; Revised 6 August 2020; Accepted 18 September 2020; Published 22 October 2020

Academic Editor: Sang-Hoon Hong

Copyright © 2020 Wu Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Excessive groundwater exploitation has brought about severe ground subsidence in Guandu (China), threatening the stability of urban infrastructure. Mapping of the spatiotemporal variations of ground deformation is urgently needed for disaster prevention and mitigation. In this study, multisensor interferometric synthetic aperture radar (InSAR) observations were applied to Guandu to derive the time series deformation from 2007 to 2019. The annual deformation velocity revealed three severe subsiding regions in Guandu. Based on the ascending and descending Sentinel-1 images with overlapping temporal and spatial coverage, two-dimensional vertical and horizontal east–west deformation was calculated and indicated that the deformation in Guandu was dominated by vertical direction. After connecting the multisensor results, long-term ground deformation spanning from January 9, 2007, to September 1, 2019, was produced and showed that the north subsiding region experienced fast followed by slow subsidence, whereas the south subsiding region experienced slow followed by fast subsidence. This difference was due to the changes of groundwater pumping centers and rates. The cumulative maximum subsidence reached 400 mm during the period of 2007–2019. The similar variations in temporal domain between the change of groundwater level and ground deformation suggested that groundwater exploitation accounted for the severe subsidence in Guandu. Our results may provide scientific evidence regarding the sound management of groundwater exploitation to mitigate potential damage to infrastructure and the environment.

1. Introduction

Guandu is located at a longitude of $102^{\circ}38' - 103^{\circ}03'$ east and a latitude of $24^{\circ}54' - 25^{\circ}17'$ north [1]. Situated in the mideastern Yunnan–Guizhou plateau in Southwest China, Guandu is a typical lacustrine sediment basin with altitude from 1886 to 2731 m [2]. Since the launch of China's reform program in 1978, the population and economics of Guandu have experienced rapid development. Accordingly, the demands of water for domestic and industrial consumption increased rapidly over the past forty years. In this context, phenomenon of groundwater exploitation was serious in Guandu. The related document showed that groundwater level of Guandu had declined about 40.3 m between 1984 and 1998

[3]. As many pioneer studies indicated, groundwater exploitation certainly resulted in ground subsidence [4, 5]. Therefore, severe subsidence was observed in Guandu and caused substantial damage to homes, roads, canals, pipelines, and other infrastructure [6]. To prevent and mitigate the disasters, it is urgent to characterize and monitor the spatiotemporal variations of ground deformation in Guandu [7].

Early studies initiated a few years ago using leveling observation recorded the extent of subsidence in Guandu and showed that the subsidence reached 240 mm between 1987 and 1998 [7, 8]. Although these observations provided useful information, they had low spatial resolution, and therefore, more detailed and comprehensive ground deformation was difficult to see [9]. Interferometric synthetic aperture radar

(InSAR) has demonstrated its potential for high-density spatial mapping of ground deformation associated with earthquakes [10], volcanoes [11], and other geologic processes [12]. Recent advanced InSAR techniques have improved our understanding of the process of ground deformation, such as the deep learning approach [13], artificial intelligence technique [14], optimal phase unwrapping algorithm [15], signal retrieval for decorrelating targets [16], four-dimensional filtering approach [17], improved synthetic aperture radar (SAR) image coregistration algorithm [18], atmospheric delay correction method [19], and coherent point selection algorithm [20]. Additionally, an increasing amount of SAR satellites have provided a large set of multisensor SAR images, which have been employed to reconstruct the spatiotemporal evolution of ground deformation [21]. These advanced techniques have facilitated land subsidence monitoring in many urban areas, such as in Beijing [22], Xian [23], Shenzhen [24], Taiyuan [25], Mexico [26], Italy [27], and Turkey [28]. Regarding Guandu, existing research focused on the whole city and there was no information available for spatiotemporal characterization analysis of ground deformation so far [7–9].

To better understand the spatiotemporal evolution of ground deformation in Guandu, the multisensor InSAR observation was utilized to obtain the long-term and two-dimensional time series deformation. For this, 20 ALOS-1 images, 40 COSMO-SkyMed (CSK) images, 24 ALOS-2 images, and 91 Sentinel-1A images were collected in this study. The annual deformation velocity was estimated to characterize the spatial pattern of subsidence in Guandu. Subsequently, two-dimensional and long-term time series deformation was obtained to analyze the temporal evolution of subsidence. Finally, the correlation between subsidence and groundwater changes was discussed.

2. Study Area and Datasets

2.1. History and Geologic Setting of Study Area. Guandu, situated in the eastern part of Kunming (China) (Figure 1(a)), has one of the mildest climates in China, characterized by short, cool, dry winters with mild days and crisp nights and long, warm, humid summers [1]. The annual average precipitation is about 1,002 mm, of which 86%-90% concentrates in the summer and autumn (May to October). The annual average evaporation is about 1,900 mm, which is greater than the annual average precipitation [3]. Groundwater is the main water supply in Guandu, where the annual average groundwater exploitation is more than 20,000 m³, among which 90% is used as domestic water consumption and 10% is used as industrial water consumption [8]. Due to the pollution of shallow pore water, exploitation of groundwater is mainly from the karst water in the study area. Figure 1(b) shows the stratigraphic profile along the line PP' in Figure 1(a), where clay, sand, and dolomite are widespread in the study area. It is observed from Figure 1(b) that the contact layer between soil and bedrock is not the water storage layer, indicating that the soil aquifer is directly connected to the bedrock aquifer. In this context, the groundwater level of soil aquifer rapidly declines with the decline of groundwater level of bedrock aquifer. This decline causes the water-

release compression and soil consolidation, resulting in the ground subsidence. Thus, groundwater exploitation of karst water leads to the several subsidence over the study area. Early studies showed that groundwater level of Guandu had declined about 40.3 m between 1984 and 1998, which caused the cumulative subsidence of 250 mm [8], as shown in Figure 1(c).

2.2. SAR Datasets. 175 SAR images in total, including 20 ALOS-1 images, 40 COSMO-SkyMed (CSK) images, 24 ALOS-2 images, and 91 Sentinel-1A images, were collected to derive the deformation time series over the study area. Table 1 and Figure 1(a) show the parameters and coverage of these SAR images, respectively. The collected multisensor SAR datasets with different imaging parameters, e.g., azimuth and incidence angle, spatial and temporal resolution, orbit direction, and wavelength, allow us to comprehensively analyze the observed surface deformation [29]. Meanwhile, two-dimensional ground deformation was calculated by integrating of the multisensor SAR datasets with overlapping temporal and spatial coverage [30]. After setting the spatial and temporal baseline thresholds, a total of 624 interferograms was generated from the collected multisensor SAR images. Based on the generated interferograms, the ground deformation between January 2007 and September 2019 was estimated over the study area. Light detection and ranging (LIDAR) digital elevation model (DEM), which had a spatial resolution of 3 m and centimeter-level height precision, was acquired as an external DEM to remove the topographic phase from the differential interferograms.

3. Methodology

Using the collected 175 multisensor SAR images, ground deformation was estimated by multitemporal SAR interferometry technique. The interferometric pairs were generated by setting small temporal and spatial baselines, which limited the noise effects and preserved the temporal and spatial coherence characteristics of the interferograms. As a result, totally 624 interferometric pairs were produced, among which 59 pairs were from ALOS-1 images, 88 pairs were from COSMO-SkyMed images, 71 and 15 pairs were, respectively, from ALOS-2 ascending and descending images, and 91 and 300 pairs were, respectively, from Sentinel-1 ascending and descending images. The topographic and orbital phases were simulated and subsequently removed from the differential interferometric phase by using the collected LIDAR DEM and SAR orbital data. Then, the process of adaptive spectral filtering with a window size of 32 was executed to suppress the interferometric noise [31]. After that, phase unwrapping using the Minimum Cost Flow (MCF) method was carried out to retrieve the absolute phase [32]. Finally, the atmospheric delay errors were reduced through the cascade of a low-pass filtering implemented in the two-dimensional spatial domain, followed by a temporal high-pass filtering [33]. After these processes, the unwrapped interferograms from multisensor datasets were resampled to a common study area (red rectangle in Figure 1(a)) and used to estimate

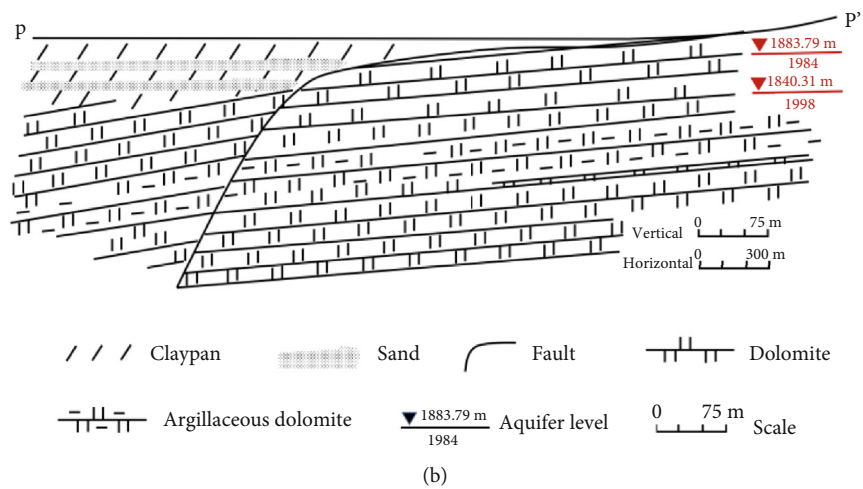
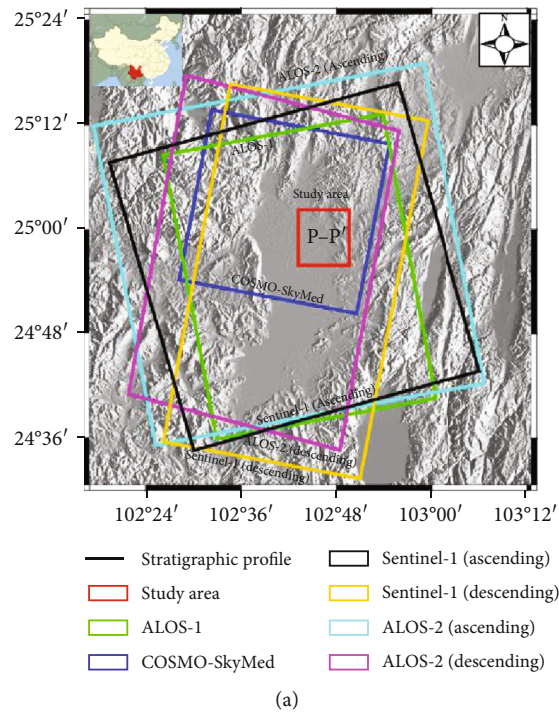


FIGURE 1: Continued.

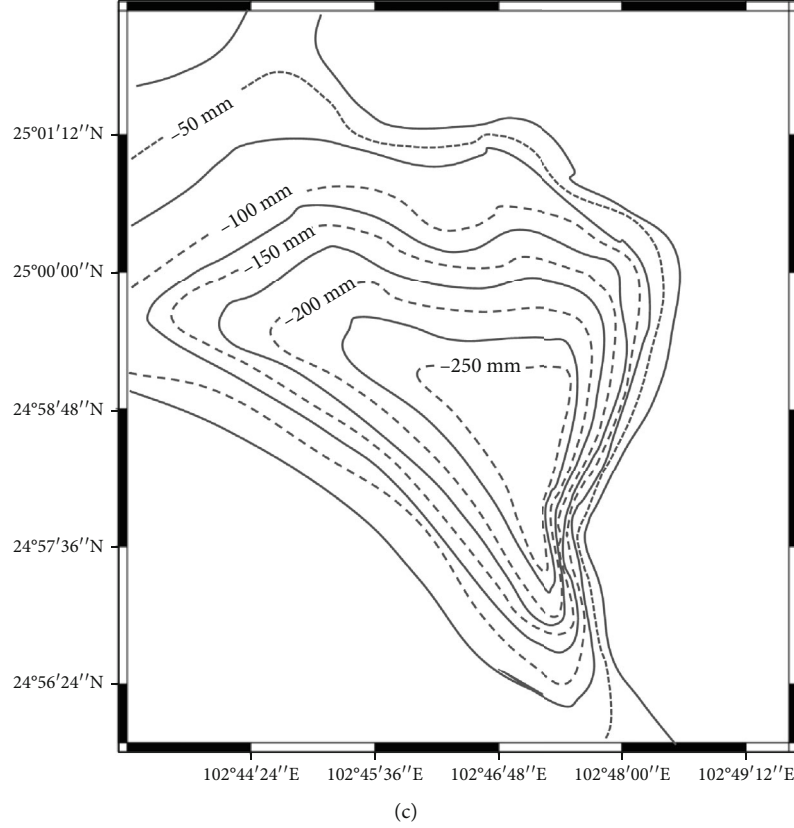


FIGURE 1: (a) Study area and synthetic aperture radar (SAR) data coverage used in this study; corresponding data are superimposed on the digital elevation model (DEM); (b) stratigraphic profile along the line PP' in (a); (c) ground subsidence due to groundwater exploitation between 1984 and 1998.

TABLE 1: Parameters of SAR data in this study.

Satellite	ALOS-1 (ascending)	COSMO-SkyMed (descending)	ALOS-2 (ascending)	ALOS-2 (descending)	Sentinel-1 (ascending)	Sentinel-1 (descending)
Azimuth angle (°)	-10	-170	-10	-169	-12	-169
Incidence angle (°)	38	29	31	40	39	39
Number of SAR images	20	40	18	6	31	60
Number of interferograms	59	88	71	15	91	300
Data period	09/01/2007- 07/03/2011	13/06/2011- 06/01/2016	26/09/2014- 11/01/2019	06/06/2015- 26/08/2017	23/01/2015- 17/02/2017	25/05/2015- 01/09/2019

the vertical deformation velocities, two-dimensional ground deformation, and long-term vertical deformation time series.

3.1. Vertical Deformation Velocities. The vertical deformation velocities of multisensor SAR datasets with different imaging parameters were calculated to characterize the spatial pattern of ground deformation over the study area. Using the produced unwrapped interferometric pairs, the vertical deformation velocities were estimated by [34]

$$\text{Ph}_{\text{rate}} = \frac{\sum_{j=1}^N \Delta t_j \varphi_j}{\sum_{j=1}^N \Delta t_j^2}, \quad (1)$$

where Ph_{rate} was the deformation velocities, N was the number of interferograms involved in the estimation, Δt_j was the time interval between master and slave pair, and φ_j was the unwrapped interferograms. Meanwhile, the standard deviation of Ph_{rate} was estimated by [35]

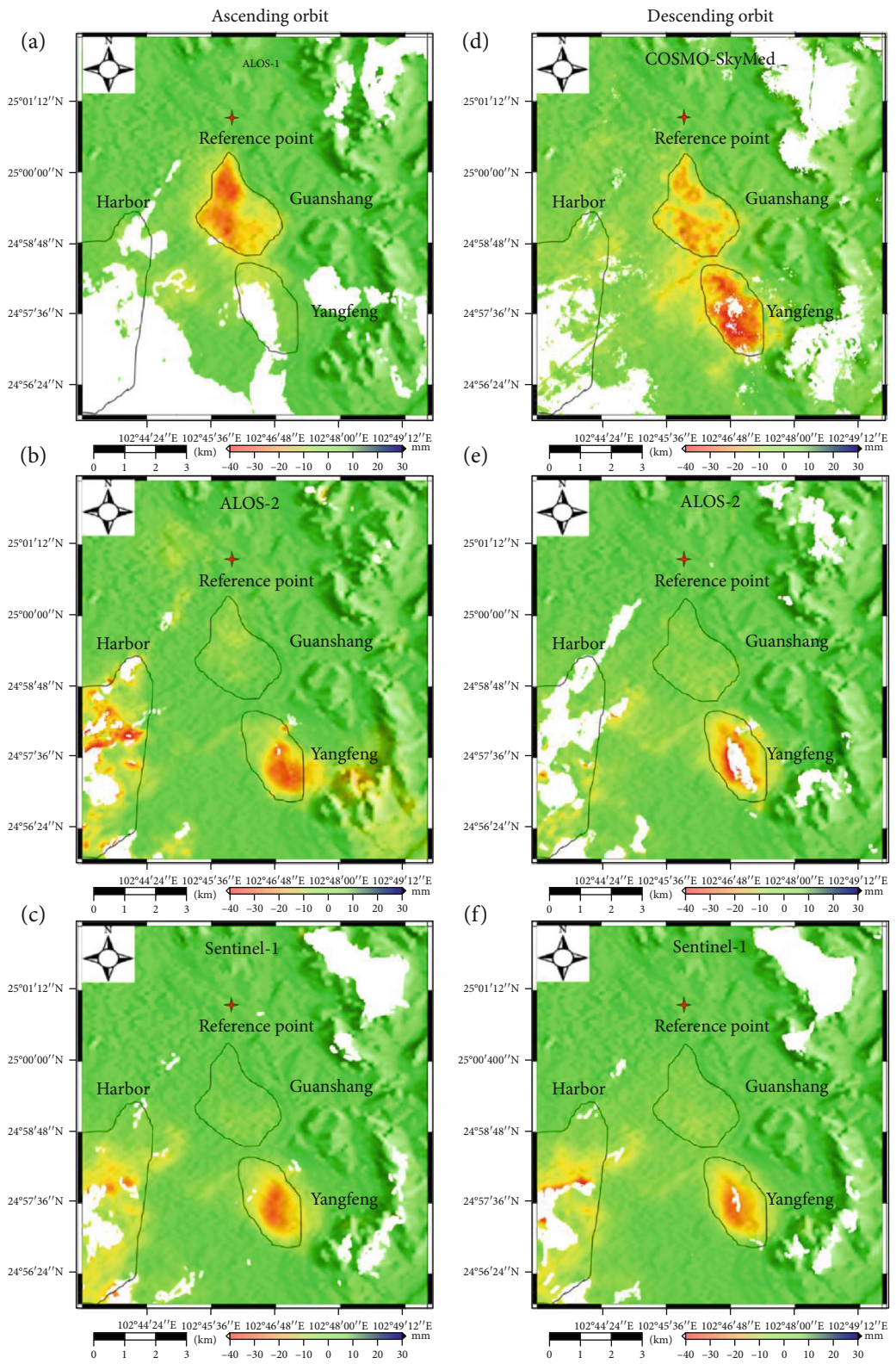


FIGURE 2: Deformation velocities from (a) ALOS-1, (b) ALOS-2, and (c) Sentinel-1 datasets with ascending orbit and (d) COSMO-SkyMed, (e) ALOS-2, and (f) Sentinel-1 datasets with descending orbit.

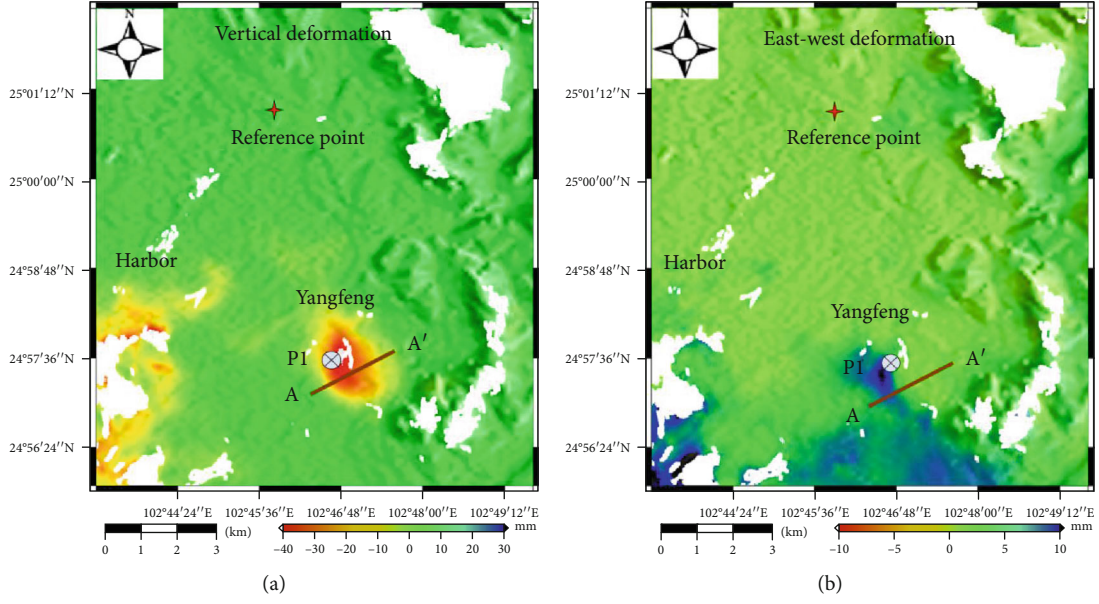


FIGURE 3: (a) Vertical and (b) horizontal deformation rate maps. The red cross shows the reference point; the cross circle shows the location of time series deformation point P1; the brown solid line shows the location of profile line of AA'.

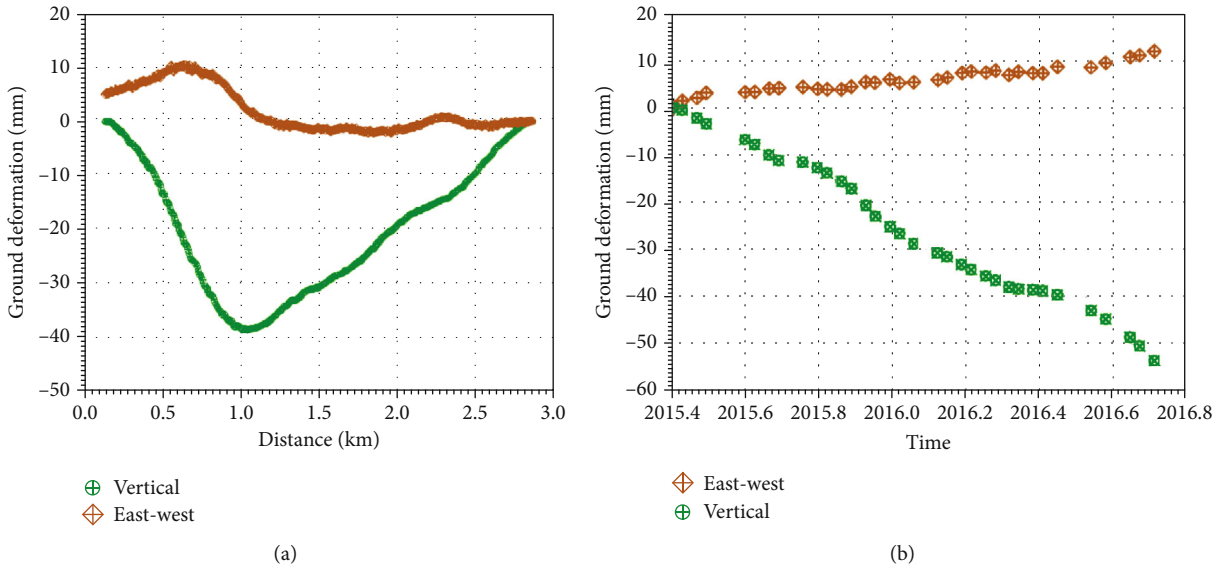


FIGURE 4: (a) Deformation rate along the line AA' and (b) time series deformation at point P1 in vertical and horizontal east–west directions.

$$\text{Var}(\text{Ph}_{\text{rate}}) = \frac{\sum_{j=1}^N \left(\varphi_j - (4\pi/l) \cdot \text{Ph}_{\text{rate}} \cdot \Delta t_j \right)^2}{\Delta t_j^2}, \quad (2)$$

where l was the radar wavelength and $\text{Var}(\text{Ph}_{\text{rate}})$ was the standard deviation of Ph_{rate} . In this case, six vertical deformation velocities from ALOS-1, COSMO-SkyMed, ascending and descending ALOS-2, and ascending and descending Sentinel-1 datasets were produced over the study area, which allowed us to compare the observed deformation.

3.2. Two-Dimensional Ground Deformation. In areas with overlapping spatiotemporal coverage, it is possible to map the two-dimensional vertical and horizontal east–west ground deformation through combing of multisensor SAR datasets. In this case, the Sentinel-1 datasets in ascending and descending orbits were used to derive the two-dimensional ground deformation. It is worth noting that the ALOS-2 datasets were not involved in the two-dimensional deformation estimation due to the limited acquisitions in descending orbit. The multidimensional small baseline subset (MSBAS) method was adopted to calculate

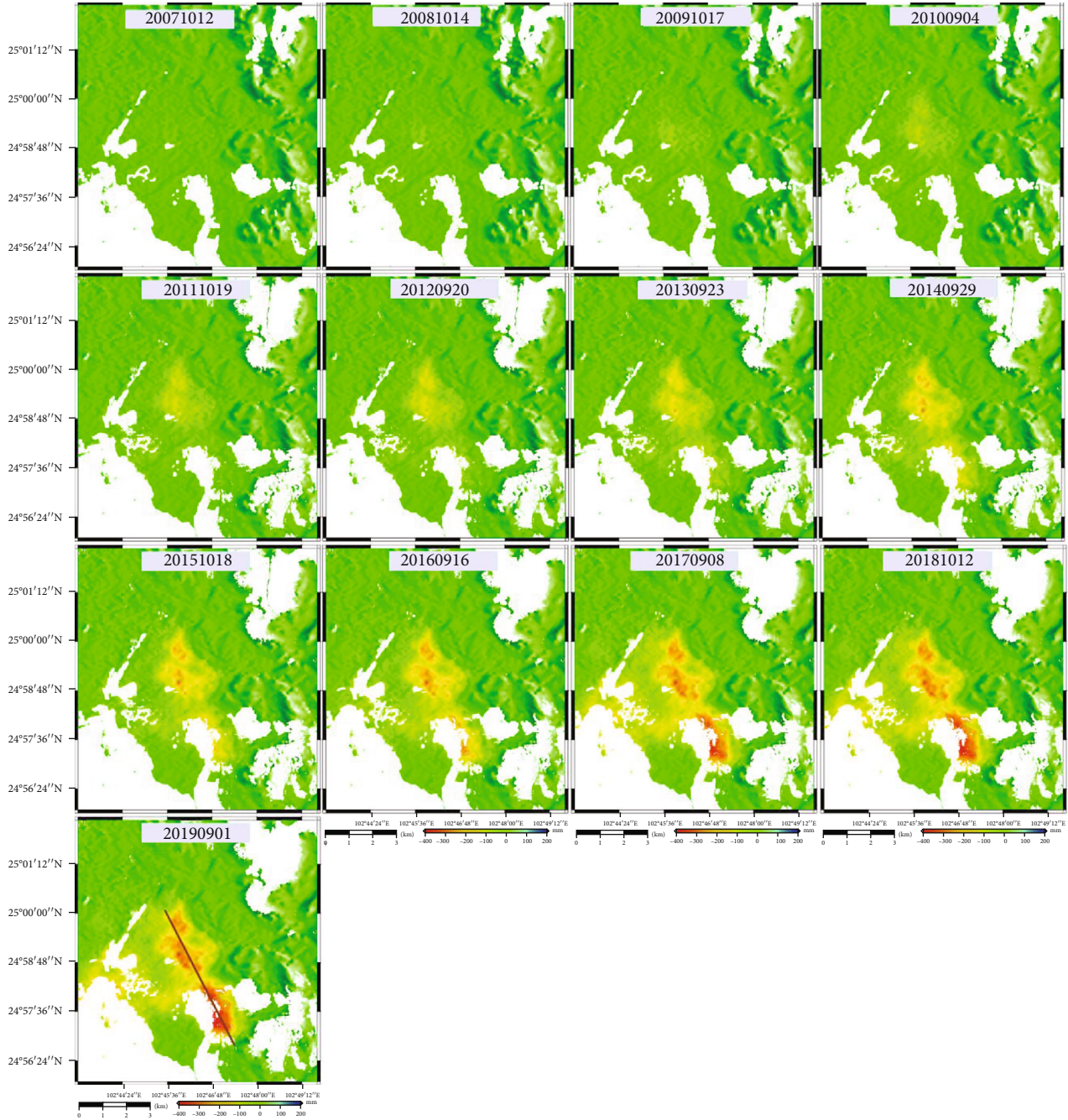


FIGURE 5: Time series deformation in September or October from 2007 to 2019.

the two-dimensional ground deformation [36]:

$$\begin{pmatrix} \hat{A} \\ \lambda L \end{pmatrix} \begin{pmatrix} V_E \\ V_U \end{pmatrix} = \begin{pmatrix} \hat{\Phi} \\ 0 \end{pmatrix}, \quad (3)$$

$$\hat{A} = \{-A \cos \theta \sin \phi, A \cos \phi\},$$

where matrix A consisted of time intervals between consecutive SAR acquisition, θ was the azimuth angle, ϕ was the inci-

dence angles; V_E and V_U represented unknown horizontal east-west and vertical velocities that were to be determined, $\hat{\Phi}$ represented the unwrapped, geocoded, and resampled interferometric phase, λ was a regularization parameters, and L was a zero-, first-, or second-order difference operator. The unknown parameters V_E and V_U for each pixel were solved by applying the singular value decomposition (SVD), and deformation time series were reconstructed from the computed deformation rates by numerical integration. In this case, the first order regularization with λ equal to 0.1

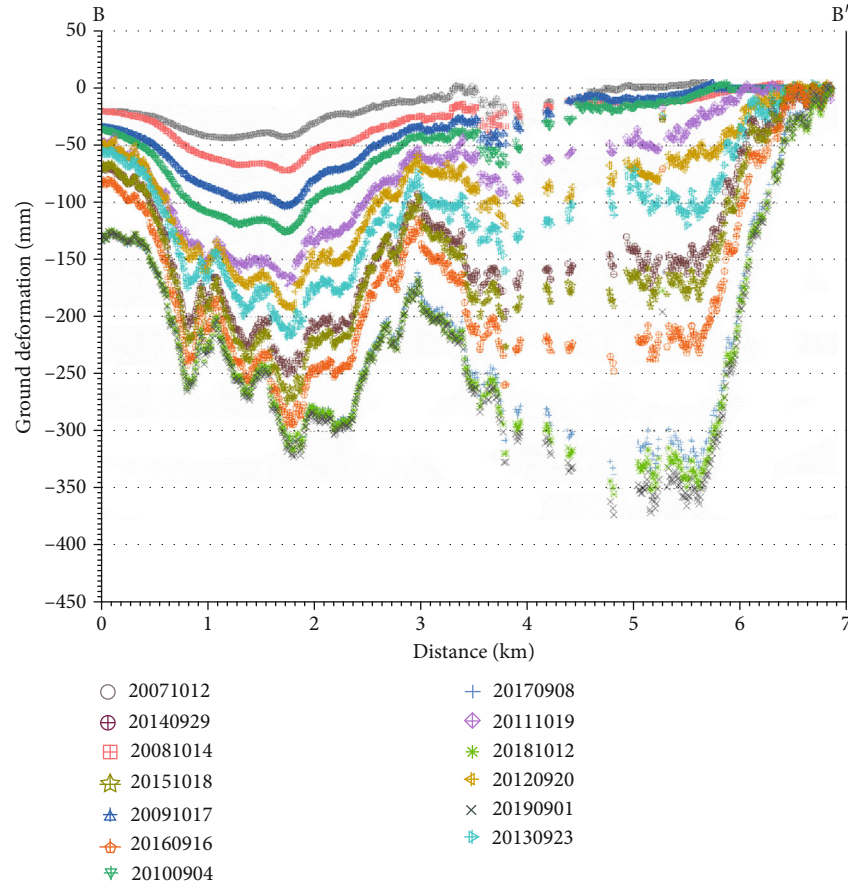


FIGURE 6: Time series deformation along the line BB' marked in Figure 5.

was chosen to be involved in the calculation based on the repeated experiments. Finally, two-dimensional vertical and horizontal east–west ground deformation from May 23, 2015, to September 16, 2016, was produced over the study area, which allowed us to analyze the ground deformation both in vertical and east–west direction.

4. Results and Analysis

4.1. Vertical Deformation Velocities with Different SAR Datasets. Figure 2 shows the vertical deformation velocities, which were derived from ALOS-1, ALOS-2, and Sentinel-1 datasets with ascending orbit and COSMO-SkyMed, ALOS-2, and Sentinel-1 datasets with descending orbit. It is worth noting that all measurement values were relative to a common reference point (red cross in Figure 2) that was considered to be stable over the entire time period. Positive and negative values represented uplift and subsidence. Three prominent deformation regions were observed from Figure 2. The first deformation region named as Guanshang was located at the center of deformation maps and presented the shape of approximate pear. The second deformation region with the shape of approximate oval was named as Yangfeng and located at the lower center of deformation maps, which was considered as the extension of Guanshang deformation region. The third deformation region was located at the lower left of deformation maps and named as

Harbor. For these three deformation regions, it was found that they showed the different deformation magnitudes with respect to the different datasets. The Guanshang deformation region displayed the severe subsidence in ALOS-1 and COSMO-SkyMed datasets, while it was not prominent in ALOS-2 and Sentinel-1 datasets. As for the Yangfeng deformation region, the severe subsidence was observed in COSMO-SkyMed, ALOS-2, and Sentinel-1 datasets, while it was not prominent in ALOS-1 datasets. Compared to the Guanshang deformation region, the Harbor deformation region presented the opposite deformation magnitude: severe subsidence appeared in ALOS-2 and Sentinel-1 datasets, while insignificant deformation was observed in ALOS-1 and COSMO-SkyMed datasets. This phenomenon was due to the different imaging time regarding to the different SAR datasets. As shown in the introduction, the subsidence in Guandu was highly correlated with groundwater extraction, meaning that the degree of groundwater extraction with respect to the different times may have led to the different deformation magnitudes. The relevant information was investigated, and the degree of groundwater extraction was found to be extremely intensive in Guanshang in early acquired ALOS-1 and COSMO-SkyMed datasets. Therefore, the Guanshang deformation region showed the significant signal in ALOS-1 and COSMO-SkyMed datasets, while it was not significant in ALOS-2 and Sentinel-1 datasets. The Yangfeng and Harbor deformation regions experienced the

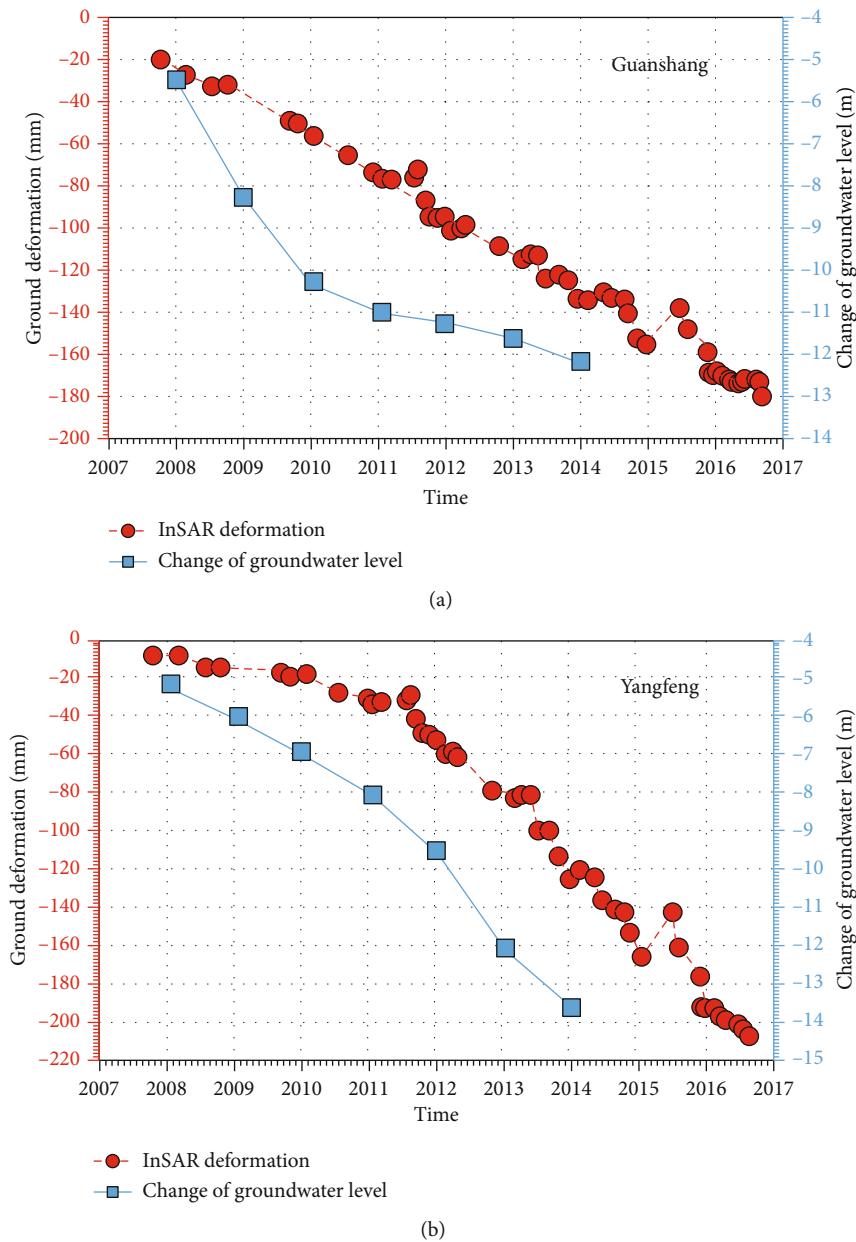


FIGURE 7: The relationship between groundwater level changes and ground deformation in Guanshang (a) and Yangfeng (b) over the study area.

slow followed by fast subsidence in temporal domain and therefore showed the significant signal in later acquired ALOS-2 and Sentinel-1 datasets. Further inspection indicated that there were subtle differences in deformation shapes between descending and ascending orbits, particularly for the Yangfeng deformation region, where the deformation shape presented in descending orbit was elongated in ascending orbit. Through analyzing the vertical deformation velocities from different SAR datasets, it was summarized that three prominent subsiding regions were observed over the study area.

4.2. Two-Dimensional Vertical and Horizontal Deformation. The difference in deformation shape between descending

and ascending orbits indicated the existence of horizontal deformation. In order to investigate this deformation, the deformation velocities and time series deformation maps along the vertical and horizontal east–west directions were derived from the ascending and descending Sentinel-1 images, which were acquired from May 23, 2015, to September 16, 2016. The positive and negative values in vertical direction, respectively, represent uplift and subsidence, while positive and negative values in east–west direction, respectively, represent eastward and westward movement. The vertical and east–west deformation velocities are presented in Figure 3. Compared to Figure 2, the Yangfeng and Harbor deformation regions were retained in Figure 3, while the deformation in Guanshang region was not obvious. As analyzed in Section



FIGURE 8: Long-term vertical deformation time series spanning from 9 January 2007 to 1 September 2019 by combing of the collected multi-sensor SAR datasets.

4.1, this phenomenon was due to the temporal variations of deformation. The statistics showed that the maximum subsidence and eastward movement rates were 38 mm/year and 11 mm/year for Yangfeng region, respectively. For Harbor

deformation region, the maximum subsidence rate was up to 25 mm/year, and the maximum eastward movement rate was 12 mm/year. Compared with the Yangfeng deformation region, the phenomenon of decorrelation was relatively severe

at Harbor deformation region. This decorrelation was due to the ongoing constructions according to the filed investigation. The cause of observed deformation in Figure 3 was attributed to the groundwater exploitation considering the history and geologic setting of study area. The pioneering researches suggested that this aquifer deformation occurred in both the vertical and horizontal directions in response to fluid withdrawal from confined aquifer system [37–39]. In the vertical direction, the decline of groundwater level led to the water-release compression and soil consolidation with the result of ground subsidence. In the horizontal direction, the eastward movement was related with a fault located at the east of the study area (as shown in Figure 1(b)), which impeded the horizontal propagation of fluid-pressure changes. The magnitude of aquifer deformation was a function of the distance from the center of pumping, the pumping rate, the time since the initiation of pumping, the hydraulic diffusivity of the aquifer, and the presence of heterogeneous barriers or boundaries that may impede flow and/or displacement [39]. Unfortunately, it was difficult to evaluate the observed vertical and horizontal deformation because of the lack of relevant supporting data. However, the standard deviations of two-dimensional deformation velocities were below 3 mm for most regions, suggesting the reliability of our observed deformation.

To characterize the spatiotemporal variation of deformation over the study area, the deformation rates along the profile line AA' and time series deformation at point P1, both of which are marked in Figure 3, are displayed in Figure 4. It was observed in Figure 4(a) that the vertical deformation presented the shape of a funnel, while east–west deformation presents the shape of approximate parabola. As for the temporal variations in Figure 4(b), both vertical and horizontal east–west deformation showed an approximately linear variation with time, where the cumulative deformation between May 23, 2015, and September 16, 2016, was -55 mm and 12 mm in vertical and east–west directions, respectively.

4.3. Long-Term Vertical Deformation Time Series. To investigate the temporal variation of ground deformation over the study area, the collected multisensor SAR datasets were combined to produce the long-term vertical deformation time series. During this procedure, the line-of-sight (LOS) deformation was directly converted to the vertical direction since it was dominated by the vertical deformation in this case, which has been proved in the last section. The final time series deformation was from January 9, 2007, to September 1, 2019, as shown in Figure 5. For saving the page space, we only showed the results in September or October from 2007 to 2019, as shown in Figure 6. The prominent subsiding region was detected, where the cumulative maximum subsidence reached 400 mm. Further observations suggested that the subsidence experienced two different stages of development. In the first stage spanning from 2007 to 2014, the rapid subsidence was observed at the north subsiding region, while the subsidence was relative slow for the south subsiding region. However, this phenomenon was changed in the second stage, which spans from 2014 to 2019. The accelerated subsidence appeared at the south subsiding region, while the decelerated subsidence was observed at the north subsid-

ing region. We think this phenomenon may be related with the changes of groundwater pumping centers and rates.

Figure 7 displays the time series deformation along the line BB' marked in Figure 6. It was found that the subsiding region was composed of two different subsiding funnels, where the cumulative maximum subsidence was 330 mm and 370 mm, respectively. Meanwhile, two subsiding funnels presented the different developing processes: the north region showed the fast followed by slow subsidence, while the south region showed the slow followed by fast subsidence.

5. Correlation between Ground Deformation and Groundwater Exploration

Groundwater is the main water supply in Guandu, where the annual average groundwater exploitation is more than 20,000 m³. To investigate the changes of groundwater, the groundwater observations at Guanshang and Yangfeng stations were collected in this study, as shown by the blue rectangles in Figure 8. It is observed in Figure 8(a) that there is a rapid decline between 2008 and 2010 followed by a slow decline between 2010 and 2014 for Guanshang groundwater level. It is different with Guanshang that Yangfeng station (Figure 8(b)) presents a gradual decline between 2008 and 2012 followed by a rapid decline between 2012 and 2014. The cumulative declines of groundwater at Guanshang and Yangfeng stations between 2008 and 2014 are, respectively, 12.2 and 13.8 m. Considering the presented stratigraphic structure in Figure 1(b), the subsidence easily occurred in the presence of groundwater exploitation, which caused the water-release compression and soil consolidation. The red circles in Figure 8 present the similar variations in temporal domain, particularly for Yangfeng station. However, there is a subtle inconsistent variation between the change of groundwater level and ground deformation from 2008 to 2011 for Guanshang station; the groundwater level illustrates the accelerated decline while ground deformation shows the steady decline. This inconsistency may have been caused by the inaccurate groundwater observation in 2010 since Guanshang groundwater gauging station was temporarily damaged during the building construction between 2010 and 2011 [7]. Based on the analyses of ground deformation and groundwater changes, groundwater exploitation accounts for the severe subsidence in Guandu.

6. Conclusion

With the rapid development of population and economics in Guandu, ground subsidence has caused substantial damage to homes, roads, canals, pipelines, and other types of infrastructure. In this context, time series deformation of Guandu was derived from multisensor InSAR observations to characterize and monitor its spatiotemporal variations. This observation may provide scientific evidence regarding the sound management of groundwater exploitation. Based on this study, the following conclusions were made.

- (1) Severe subsiding regions were detected through multisensor InSAR observations. Using 20 ALOS-1

images, 40 COSMO-SkyMed (CSK) images, 24 ALOS-2 images, and 91 Sentinel-1A images, the annual deformation velocity and time series deformation maps of Guandu were retrieved through multisensor InSAR processing. The results showed that three severe subsiding regions were observed, where the cumulative subsidence reached 400 mm during the period of 2007–2019. The deformation of Guandu was dominated by the vertical direction through analyzing the two-dimensional time series deformation

- (2) Groundwater exploitation accounted for the subsidence in Guandu. Comparison between the change of groundwater level and ground deformation indicated that they presented the similar variations in temporal domain, suggesting that groundwater exploitation caused the severe subsidence in Guandu

Data Availability

Sentinel-1A/B images were from the European Space Agency (ESA) project, and these data were downloaded from the Sentinel-1 Scientific Data Hub. ALOS-2 images were provided by the Japan Aerospace Exploration Agency (JAXA). Several figures were prepared using Generic Mapping Tools software.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was funded by Chang'an University (Xi'an, China) through the National Key R&D Program of China (2019YFC1509802 and 2018YFC1505101), Natural Science Foundation of China projects (NSFC) (Nos. 41790445, 41941019, 41731066, and 41504005), Natural Science Basic Research Plan in Shaanxi Province of China (2018JQ4031 and 2018JQ4027), and the Fundamental Research Funds for the Central Universities, CHD (Nos. 300102269207, 300102269303, and 300102268204).

References

- [1] Y. Li, N. Zhang, and K. Zhang, "Analysis of groundwater monitoring, exploitation and utilization in Yunnan Province," *Pearl River*, vol. 37, no. 8, pp. 44–47, 2016.
- [2] N. Lu, A. J. Hernandez, and R. D. Ramsey, "Land cover dynamics monitoring with Landsat data in Kunming, China: a cost-effective sampling and modelling scheme using Google Earth imagery and Random Forests," *Geocarto International*, vol. 30, pp. 1–34, 2014.
- [3] Y. Li and N. Zhang, "Assessment of groundwater overdraft zones in Kunming basin," *Yangtze River Scientific Research*, vol. 34, no. 6, pp. 35–38, 2017.
- [4] T. L. Holzer and A. I. Johnson, "Land subsidence caused by ground water withdrawal in urban areas," *GeoJournal*, vol. 11, no. 3, pp. 245–255, 1985.
- [5] T. Woldai, G. Oppliger, and J. Taranik, "Monitoring dewatering induced subsidence and fault reactivation using interferometric synthetic aperture radar," *International Journal of Remote Sensing*, vol. 30, no. 6, pp. 1503–1519, 2009.
- [6] R. L. Hu, Z. Q. Yue, L. C. Wang, and S. J. Wang, "Review on current status and challenging issues of land subsidence in China," *Engineering Geology*, vol. 76, no. 1–2, pp. 65–77, 2004.
- [7] Y. Yang, X. M. Huang, Q. Zhang, and W. Zhu, *Investigation report of surface cover and ground deformation of Kunming, China*, Yunnan Institute of Surveying and Mapping Engineering, Kunming, China, 2017.
- [8] G. T. Meng and F. Li, "A case study of land subsidence in southern Kunming downtown area," *Hydrogeology and Engineering Geology*, vol. 1, no. 6, pp. 29–33, 2007.
- [9] W. Zhu, W. L. Li, Q. Zhang et al., "A decade of ground deformation in Kunming (China) revealed by multi-temporal synthetic aperture radar interferometry (InSAR) technique," *Sensors*, vol. 19, no. 20, p. 4425, 2019.
- [10] C. Wang, X. Shan, C. Wang, and X. L. Ding, "Using finite element and Okada models to invert coseismic slip of the 2008 Mw 7.2 Yutian earthquake, China, from InSAR data," *Journal of Seismology*, vol. 17, no. 2, pp. 347–360, 2013.
- [11] M. J. Jo, H. S. Jung, and J. S. Won, "Detecting the source location of recent summit inflation via three-dimensional InSAR observation of Kilauea volcano," *Remote Sensing*, vol. 7, no. 11, pp. 14386–14402, 2015.
- [12] M. Zheng, K. Deng, H. Fan, S. Du, and H. Zou, "Monitoring and analysis of mining 3D time-series deformation based on multi-track SAR data," *International Journal of Remote Sensing*, vol. 40, no. 4, pp. 1409–1425, 2019.
- [13] N. Anantrasirichai, J. Biggs, F. Albino, and D. Bull, "A deep learning approach to detecting volcano deformation from satellite imagery using synthetic datasets," *Remote Sensing of Environment*, vol. 230, article 111179, 2019.
- [14] S. Valade, A. Ley, F. Massimetti et al., "Towards Global Volcano Monitoring Using Multisensor Sentinel Missions and Artificial Intelligence: the MOUNTS Monitoring System," *Remote Sensing*, vol. 11, article 1528, 2019.
- [15] H. Yu, H. Lee, N. Cao, and Y. Lan, "Optimal baseline design for multibaseline InSAR phase unwrapping," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5738–5750, 2019.
- [16] R. Prébet, Y. Yan, M. Jauvin, and E. Trouvé, "A data-adaptive EOF-based method for displacement signal retrieval from InSAR displacement measurement time series for decorrelating targets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5829–5852, 2019.
- [17] K. Wnuk, G. Walton, and W. Zhou, "Four-dimensional filtering of InSAR persistent scatterers elucidates subsidence induced by tunnel excavation in the Sri Lankan highlands," *Journal of Applied Remote Sensing*, vol. 13, no. 3, article 34508, 2019.
- [18] W. Zou and L. Chen, "Determination of optimum tie point interval for SAR image coregistration by decomposing autocorrelation coefficient," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 5067–5084, 2019.
- [19] Z. Li, Y. Cao, J. Wei et al., "Time-series InSAR ground deformation monitoring: atmospheric delay modeling and estimating," *Earth-Science Review*, vol. 192, pp. 258–284, 2019.
- [20] S. Wu, L. Zhang, X. Ding, and D. Perissin, "Pixel-wise MTInSAR estimator for integration of coherent point selection and

- unwrapped phase vector recovery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 5, pp. 2659–2668, 2019.
- [21] Y. Aimaiti, F. Yamazaki, and W. Liu, “Multi-sensor InSAR analysis of progressive land subsidence over the Coastal City of Urayasu, Japan,” *Remote Sensing*, vol. 10, no. 8, article 1304, 2018.
- [22] X. Liu, P. Wang, Z. Lu et al., “Damage detection and analysis of urban bridges using terrestrial laser scanning (TLS), ground-based microwave interferometry, and permanent scatterer interferometry synthetic aperture radar (PS-InSAR),” *Remote Sensing*, vol. 11, no. 5, p. 580, 2019.
- [23] M. Peng, C. Zhao, Q. Zhang, Z. Lu, and Z. Li, “Research on spatiotemporal land deformation (2012–2018) over Xi’an, China, with multi-sensor SAR datasets,” *Remote Sensing*, vol. 11, no. 6, p. 664, 2019.
- [24] P. Ma, W. Wang, B. Zhang et al., “Remotely sensing large- and small-scale ground subsidence: a case study of the Guangdong–Hong Kong–Macao Greater Bay Area of China,” *Remote Sensing of Environment*, vol. 232, article 111282, 2019.
- [25] W. Zhu, Q. Zhang, X. Ding, C. Zhao, C. Yang, and W. Qu, “Recent ground deformation of Taiyuan basin (China) investigated with C-, L-, and X-bands SAR images,” *Journal of Geodynamics*, vol. 70, pp. 28–35, 2013.
- [26] E. Chaussard, S. Wdowinski, E. Cabral-Cano, and F. Amelung, “Land subsidence in Central Mexico detected by ALOS InSAR time-series,” *Remote Sensing of Environment*, vol. 140, pp. 94–106, 2014.
- [27] J. Delgado Blasco, M. Fomelis, C. Stewart, and A. Hooper, “Measuring urban subsidence in the Rome metropolitan area (Italy) with Sentinel-1 SNAP-StaMPS persistent scatterer interferometry,” *Remote Sensing*, vol. 11, no. 2, p. 129, 2019.
- [28] M. Imamoglu, F. Kahraman, Z. Cakir, and F. B. Sanli, “Ground deformation analysis of Bolvadin (W. Turkey) by means of multi-temporal InSAR techniques and Sentinel-1 data,” *Remote Sensing*, vol. 11, no. 9, p. 1069, 2019.
- [29] Q. Zhang, W. Zhu, X. Ding, C. Zhao, C. Yang, and W. Qu, “Two-dimensional deformation monitoring over Qingxu (China) by integrating C-, L- and X-bands SAR images,” *Remote Sensing Letter*, vol. 5, no. 1, pp. 27–36, 2014.
- [30] T. Fuhrmann and M. C. Garthwaite, “Resolving three-dimensional surface motion with InSAR: constraints from multi-geometry data fusion,” *Remote Sensing*, vol. 11, no. 3, p. 241, 2019.
- [31] I. Baran, M. P. Stewart, B. M. Kampes, Z. Perski, and P. Lilly, “A modification to the Goldstein radar interferogram filter,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 2114–2118, 2003.
- [32] M. Costantini, “A novel phase unwrapping method based on network programming,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 813–821, 1998.
- [33] A. Ferretti, C. Prati, and F. Rocca, “Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 5, pp. 2202–2212, 2000.
- [34] D. T. Sandwell and E. J. Price, “Phase gradient approach to stacking interferograms,” *Journal of Geophysics Research*, vol. 103, no. B12, pp. 30183–30204, 1998.
- [35] D. T. Sandwell and L. Sichoix, “Topographic phase recovery from stacked ERS interferometry and a low-resolution digital elevation model,” *Journal of Geophysics Research*, vol. 105, no. B12, pp. 28211–28222, 2000.
- [36] S. V. Samsonov, W. Feng, A. Peltier, H. Geirsson, N. Oreye, and K. F. Tiampo, “Multidimensional small baseline subset (MSBAS) for volcano monitoring in two dimensions: opportunities and challenges. Case study Piton de la Fournaise volcano,” *Journal of Volcanology and Geothermal Research*, vol. 344, pp. 21–138, 2017.
- [37] D. C. Helm, “Three-dimensional consolidation theory in terms of the velocity of solids,” *Geotechnique*, vol. 37, no. 3, pp. 369–392, 1987.
- [38] T. J. Burbey, “Storage coefficient revisited: is purely vertical strain a good assumption?,” *Ground Water*, vol. 39, no. 3, pp. 458–464, 2001.
- [39] T. Burbey, “The influence of faults in basin-fill deposits on land subsidence, Las Vegas Valley, Nevada, USA,” *Hydrogeology Journal*, vol. 10, no. 5, pp. 525–538, 2002.

Research Article

Modelling Reservoir Chlorophyll-*a*, TSS, and Turbidity Using Sentinel-2A MSI and Landsat-8 OLI Satellite Sensors with Empirical Multivariate Regression

Yashon O. Ouma ^{1,2}, Kimutai Noor,¹ and Kipkemoi Herbert¹

¹Department of Civil and Structural Engineering, Moi University, 30100 Eldoret, Kenya

²Department of Civil Engineering, University of Botswana, Private Bag 0061, Gaborone, Botswana

Correspondence should be addressed to Yashon O. Ouma; yashon_o@hotmail.com

Received 4 June 2020; Revised 30 July 2020; Accepted 31 August 2020; Published 19 September 2020

Academic Editor: Lei Zhang

Copyright © 2020 Yashon O. Ouma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sentinel-2A/MSI (S2A) and Landsat-8/OLI (L8) data products present a new frontier for the assessment and retrieval of optically active water quality parameters including chlorophyll-*a* (Chl-*a*), suspended particulate matter (TSS), and turbidity in reservoirs. However, because of their differences in spatial and spectral samplings, it is critical to evaluate how well the sensors are suited for the seamless generation of the water quality parameters (WQPs). This study presents results from the retrieval of the WQP in a reservoir from L8 and S2A optical sensors, after atmospheric correction and standardization through band adjustment. An empirical multivariate regression model (EMRM) algorithmic approach is proposed for the estimation of the water quality parameters in correlation with *in situ* laboratory measurements. From the results, both sensors estimated Chl-*a* concentrations with R^2 of greater than 70% from the visible green band for L8 and a combination of green and SWIR-1 bands for S2A. While the NMSE% was nearly the same for both sensors in Chl-*a* estimation, the RMSE was $<10 \mu\text{g/L}$ and $>10 \mu\text{g/L}$ for L8 and S2A estimations of Chl-*a*, respectively. For TSS retrieval, L8 outperformed S2A by 31% in accuracy with $R^2 > 0.9$ from L8's red, blue, and green bands, as compared to $0.47 \leq R^2 \leq 0.61$ from S2A's red and NIR bands. The RMSE were the same as for Chl-*a*, and the NMSE% were both in the same range. Both sensors retrieved turbidity with high and nearly equal accuracy of $R^2 > 70\%$ from the visible and NIR bands, with equal RMSE at $<10\%$ NTU and NMAE% from S2A being higher by more than 30% as compared to L8's NMAE% at 15%. The study concluded that the higher performance accuracy of L8 is attributed to its higher SNR and spectral bandwidth placement as compared to S2A bands. Comparatively, S2A overestimated Chl-*a* and turbidity but performed equally well compared to OLI in the estimation of TSS. The results show that while absolute accuracy of retrieval of the WQPs still requires improvements, the developed algorithms are broadly able to discern the biooptical water quality in reservoirs.

1. Introduction

Despite the fact that there is an increase in the availability of safe drinking water globally, it is still estimated that approximately an eighth of the global population does not have access to potable water [1]. According to [2], nearly half of the world's population will face water shortages by 2030 as the water demand in certain countries will likely to exceed

supply by over 50%. Most of the water sources are within the river and reservoir catchments which are continuously affected by anthropogenic pressure and climate change.

Within a catchment system, the anthropogenic water uses for domestic purposes, agricultural and industrial production, mining, power generation, and forestry practices all result in the deterioration of water quality and water supply. This in effect impacts the aquatic ecosystem and compromises the

access to safe water supply for human consumption. Water quality and water supply are thus linked, although they are often not measured simultaneously. The results of water quality monitoring are important in determining the spatio-temporal trends in surface water and groundwater variabilities. For reservoirs and dams, as the key source of water for cities and rural communities, the *a priori* understanding of the quality of water before treatment is critical not only in understanding the environmental health within the catchment area but also in the minimization of the cost of treatment of the water supply.

To monitor the water quality in natural and artificial reservoirs, the conventional water quality assessment through sampling and laboratory measurement is often employed. The sampling point-based laboratory methods are costly, labor-intensive, and time-consuming and are not able to adequately assess the entire water body [3]. To overcome the limitations in *in situ* water quality monitoring methods, there is a need for regular near real-time [4–7], inexpensive, automated, and noninvasive approaches, with adequate spatial-temporal coverage.

Several studies have investigated the use of different satellite sensors for the assessment of water quality. Particularly, the Landsat sensors have been widely used in the estimation of water quality parameters such as total suspended matter, chlorophyll-*a*, turbidity, Secchi disk depth, total phosphorus, dissolved oxygen, chemical oxygen demand (COD), and biochemical oxygen demand (BOD) as reviewed in [8, 9]. For specific case studies and for the retrieval of water quality parameters (WQPs), the previous studies have developed different correlational algorithms that are based on empirical models, semianalytical models, and matrix inversion models. For semianalytical models, both the biooptical and empirical data are required to describe the relationships between the constituents of a water body and the equivalent surface reflectance that defines the upwelling radiance above the water surface and in the water surface. Based on the same modelling scheme as the semianalytical models, the matrix inversion models require *a priori* information on the water constituents, including the absorption coefficients or absorption slopes [10]. The lack of specific parameters makes the matrix inversion methods more complex and difficult to calibrate. Because of these drawbacks, empirical algorithms are often used for the retrieval and estimation of water quality parameters [3, 11–14].

Though popular, in the use of empirical algorithms, large water quality sample sizes are required and the models are sensitive to local environmental conditions and are therefore not automatically replicable to other case studies or regions. In addition to being case study or region based, the empirical models have been developed more for open sea waters as compared to inland water bodies, due in part to the fact that remote sensing measurements of freshwater resources are far more complex, in terms of surface water spectral reflectivity [15–17].

With the potential of higher temporal resolution of about 2–3 days derived from the synergistic constellation of Landsat-8 and Sentinel-2 [18], it is now possible to synchronize the products from the two satellites' higher temporal

monitoring of aquatic systems [19]. Such frequent revisits are essential for the capture of the dynamics of reservoir water bodies, in terms of surface water quality assessments, considering the effects of seasonal variabilities and atmospheric attenuations. In using Landsat-8/OLI and Sentinel-2A/MSI data for water resource monitoring, recent studies have presented different approaches and results for different case study areas. [20] used Sentinel-2A to study the influence of variations in the concentration of total suspended solids (TSS) and chlorophyll-*a* (Chl-*a*) on the physiological response of oysters and highlighted the use of Sentinel-2A near-infrared (NIR) bands to quantify the total suspended solids. [11] developed empirical-based methods for the retrieval of Chl-*a* in a hypereutrophic reservoir in Brazil and concluded that the Sentinel-2A near-infrared (NIR) bands were significant in Chl-*a* retrieval. Similar studies by [13] presented empirical algorithms using Sentinel-2A data for water quality assessment [21] or bottom mapping [22]. Nonetheless, no attempts have been made to compare and demonstrate the suitability of Sentinel-2A/MSI and Landsat-8/OLI imagery for the retrieval of Chl-*a*, TSS, and turbidity water quality parameters within inland water reservoirs.

In this study, the effectiveness of Sentinel-2A/MSI (S2A) and Landsat-8/OLI (L8) satellite sensors is demonstrated for the estimation of Chl-*a*, turbidity, and TSS water quality parameters in a large reservoir (case 2 water body). Arguably, the medium-spatial resolution satellite sensors, Operational Land Imager (OLI) on board Landsat-8 and Multispectral Imager (MSI) on board Sentinel-2, will be capable of promoting more precise mapping of biooptically active water quality parameters in recent times [11, 19]. However, because of their differences in the spectral and spatial samplings, it is critical to evaluate how well the datasets are suited for the seamless retrieval of water quality parameters. The accuracy of the biooptically active parameters as mapped from different remote sensors is largely dependent on the biooptical equation developed for its retrieval. As observed in the literature review, different case 2 water bodies respond differently to the spectral wavelengths.

Further, for case 2 waters, the concentrations of the water constituents and the corresponding water color are considered to be nonlinear. This implies that for effective measurements in these highly reflective waters, remote sensors with high dynamic spectral ranges and high signal-to-noise ratio (SNR) are needed [17]. It is therefore necessary to develop optimal algorithm(s) for the accurate estimation of biooptical water quality parameters in regional case 2 waters. The current study has two objectives: (1) to identify the most suitable spectral bands (position and width) from the Sentinel-2A/MSI and Landsat-8/OLI sensors for accurate retrieval and estimations of the concentrations of Chl-*a*, TSS, and turbidity and (2) develop, test, and validate empirical multivariate regression model (EMRM) algorithms for the estimation of the water quality parameters in case 2 waters, in correlation with *in situ* laboratory measurements.

To determine the distribution and variability of water quality parameters, ordinary Kriging is used for the spatial mapping and comparison of the distributions of the WQP in the case study reservoir. The rest of the paper is organized

as follows: in Section 2, the study area and the details of the data provided are described. The research methods are outlined in Section 3, and Section 4 presents the study results and discussions which elaborate on the comparison and analyses of the regression modelling and estimation of the WQPs from the compared satellite sensors. The study conclusions and insights are presented in Section 5.

2. Study Area and Data

2.1. Study Site Characterization. The case study reservoir is the Chebara Dam which is located between longitudes $35^{\circ}29'45.6''\text{E}$ and $35^{\circ}30'7.2''\text{E}$ and latitudes $0^{\circ}52'55.2''\text{S}$ and $0^{\circ}53'56.4''\text{S}$ and was constructed in the 1990s to supply water to Eldoret Town, which is situated 31 km away. The 2.1 km long dam has a capacity of 6.24 million cubic meters and is served by River Moiben and other small rivers and streams within the Chebara basin (Figure 1). Most of the rivers and streams flowing into the dam are permanent and flow throughout the year, with least flows experienced during the dry seasons. The climate in the Chebara Dam catchment is warm and temperate, and the area receives significant rainfall even in the driest month, with an annual precipitation average of approximately 1172 mm. Sparsely settled, agriculture is the main economic activity within the basin, with the rest of the land cover being forest, grass, and shrubs (Figures 1(a) and 1(b)). The dam also serves as a source of water to the resident population for domestic use and agricultural irrigation, especially during the dry season or periods.

The sampling stations were selected as shown in Figure 1, with the location and concentration of the sampling stations being based on the depth variations of the dam. Eighteen (18) spatially distributed sampling stations were selected: ten (10) for the calibration of the EMRM algorithm, five (5) for validation, and three (3) for off-season algorithm validation.

2.2. Optical Satellite Sensors

2.2.1. Sentinel-2/MSI Sensor. The Sentinel-2 mission comprises twin polar-orbiting satellites, Sentinel-2A and Sentinel-2B, which were, respectively, launched on 23 June 2015 and 7 March 2017. The sensors are in the same orbit and phased at 180° to each other. The orbit is sun-synchronous at 786 km altitude with an orbit inclination of 98.62° , and the mean local solar time at the descending node is at 10:30 AM. Both the orbit inclination and timing ensure minimum cloud cover and optimal sun illumination. The Sentinel-2 satellites are also aligned with similar satellites, including Landsat and SPOT-5.

Each Sentinel-2 satellite carries a multispectral instrument (MSI), with a swath width of approximately 290 km and spatial high resolutions of 10 m, 20 m, and 60 m, with a revisit time of 10 days with one satellite and 5 days with two satellites at the equator. The Sentinel MSI has 13 spectral bands (430 nm to 2320 nm) in the visible, red-edge, near-infrared, and shortwave infrared regions of the spectrum. With high-temporal resolution and 12-bit radiometric resolution, like the Landsat-8 sensor, the MSI sensor provides

high radiometric dynamics for the observed areas of both the very dark, e.g., water surfaces, and the very bright areas, e.g., ice and snow. It is further observed that the locationing and bandwidth of Landsat OLI are slightly wider than those of the MSI sensor spectral bands, as also indicated in Table 1 for the bandwidth and their spectral locations.

For comparison with the Sentinel-2A MSI, only the similar spectral bands in Landsat OLI are analyzed as presented in Table 1.

2.2.2. Landsat-8/OLI Sensor. The Landsat-8 mission carries the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). OLI collects image data for nine short-wave bands with spatial sampling or resolution of 30 meters (bands 1-7) and the panchromatic band with pixel size of 15 meters (band 8). TIRS collects data for two long-wave thermal bands at 100 meters (band 9) every 16 days which are resampled to 30 meters to match OLI multispectral bands. The OLI sensor is compatible with the earlier Landsat sensors and presents improved measurement capabilities. Compared to the Landsat-7/ETM+, the Landsat-8/OLI has improved radiometric resolution with reduced image noise and spectral heterogeneity. This is observed to be particularly significant in precise water surface extraction and water quality retrieval [23].

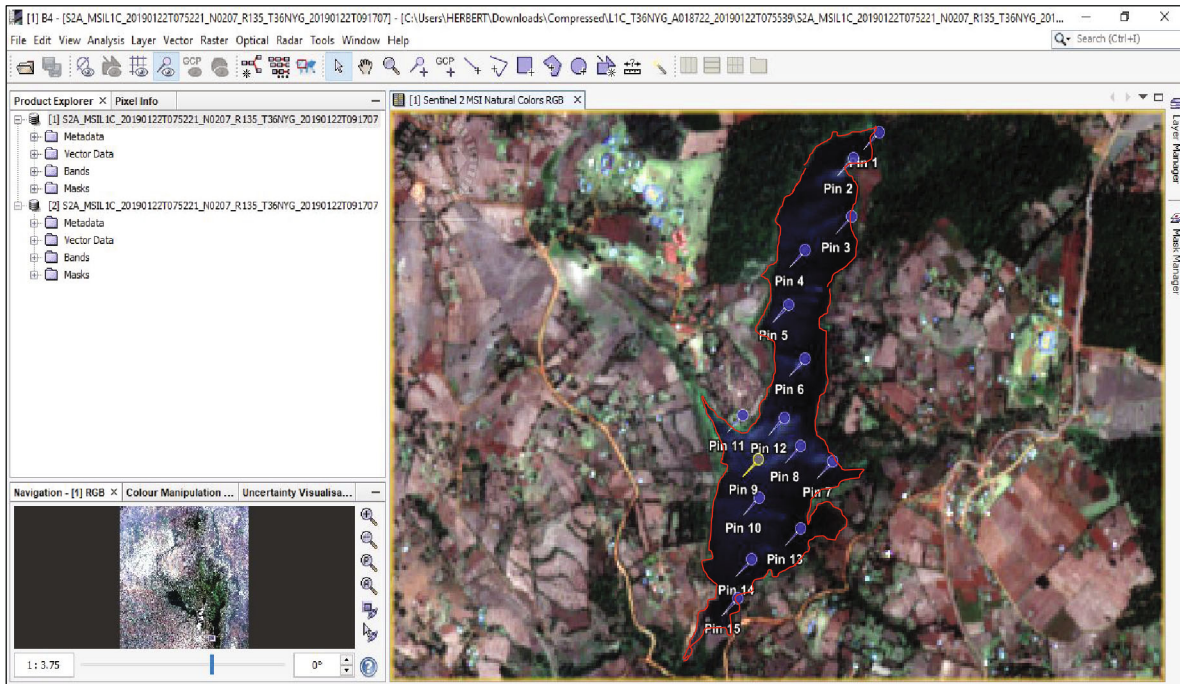
This study utilized the Landsat OLI data acquired on 22 January 2019 and the Sentinel-2A satellite image acquired on 25 January 2019. The fieldwork was carried out during the dry season and coincided with the sensor acquisitions from 23 to 24 January 2019. For the Sentinel-2A, the level 1C is corrected for atmospheric errors using the Sentinel's SNAP algorithm, resulting in the level 2A product of Sentinel-2 from the Sen2Cor processor, which includes scene classification and atmospheric correction.

3. Methods

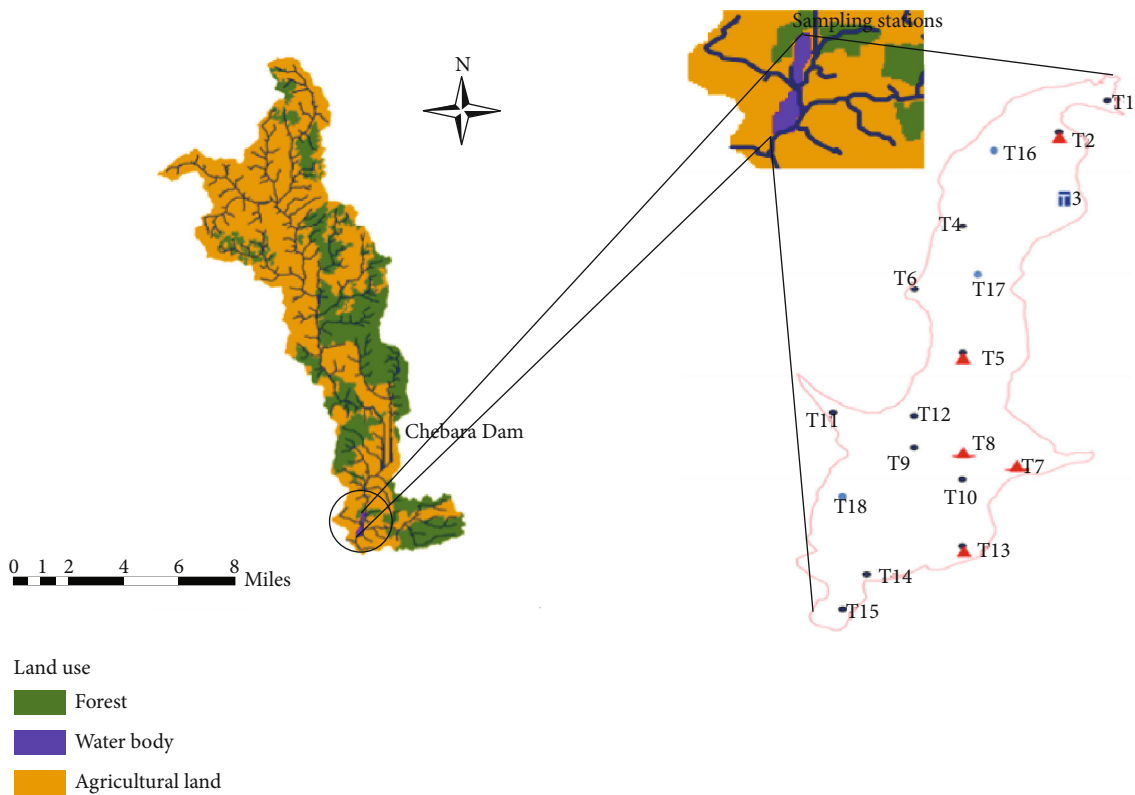
The level 1 data of the Landsat-8 OLI sensor consists of scaled digital number values that are quantized and calibrated. The equivalent level 1 data from Sentinel-2 MSI comprises the top of atmosphere (TOA) reflectance. In this study, the retrieval of the water quality parameters from the satellite sensors over the study region involves the following four steps: (i) deriving the absolute TOA reflectance from scaled DN values for Landsat OLI and scaled TOA reflectance for Sentinel-2 MSI, respectively, for the homologue bands in Table 1; (ii) conversion of TOA reflectance to surface reflectance, which is the actual reflectance originating from the water surface; (iii) conversion of the surface reflectance to the corresponding remote sensing reflectance (R_{rs}) at these bands; and (iv) the retrieval of the water quality parameters from R_{rs} utilizing the proposed empirical multivariate regression model (EMRM).

3.1. Preprocessing of Sentinel-2A and Landsat OLI Datasets

3.1.1. TOA Reflectance Derivation and Signal-to-Noise Ratio Analysis. To derive and compare the TOA from the optical sensors, for Landsat-8 level 1, the digital numbers from each band (B_{L8}) are converted to TOA reflectance using the



(a)



(b)

FIGURE 1: (a) Sentinel-2 near-true color composite of the study dam and surrounding areas and (b) land use/land cover in the Chebara basin and the sampling points (T_i).

Landsat-8 metadata scaling factors and divided by the cosine of the solar zenith angle (see processing steps in [3]). In addition, Landsat per-pixel solar and viewing angles were

calculated using the Landsat-8 Angles Creation Tool provided by USGS (<https://landsat.usgs.gov/solar-illumination-and-sensor-viewing-anglecoefficient-file>) and the angle

TABLE 1: Spectral and spatial samplings of the visible, NIR, and SWIR bands in Landsat OLI and Sentinel-2A.

Band # and spectral range	Landsat-8 OLI			Band # and spectral range	Sentinel-2A MSI		
	Spectral resolution (nm)	Bandwidth (nm)	Spatial resolution (m)		Spectral resolution (nm)	Bandwidth (nm)	Spatial resolution (m)
B1-coastal aerosol	435–451	16	30	B1-coastal aerosol	433–453	21	60
B2-blue	452–512	60	30	B2-blue	458–523	66	10
B3-green	533–590	57	30	B3-green	543–578	36	10
B4-red	636–673	37	30	B4-red	650–680	31	10
B5-NIR	851–879	28	30	B8-NIR	785–900	106	10
				B9-NIR narrow	855–875	21	20
B6-SWIR-1	1566–1651	85	30	B11-SWIR-1	1566–1655	91	20
B7-SWIR-2	2107–2294	187	30	B12-SWIR-2	2100–2280	175	20

coefficient file available with each Collection 1 L1TP file. For the Sentinel-2 MSI, the TOA reflectance and solar view geometry are stored in the metadata for every 5 km.

3.1.2. Surface Reflectance from Sentinel MSI and Landsat OLI. For the TOA reflectance derivation, the datasets from the sensors were converted to the equivalent surface reflectance using the same algorithm in order to minimize any biases in the derivation of the water quality parameters due to derivation of the reflectance. The LaSRC (Land Surface Reflectance Code) algorithm developed for Landsat and based on the 6S radiative transfer code was used and adopted for Sentinel-2A for TOA-to-surface reflectance conversion [24–26]. As validated by [27], the LaSRC has higher accuracy than the original Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm when applied to Landsat-8 OLI data [24, 26].

From the multisensor bandwidth comparisons in Table 1, it is observed that MSI and OLI have variable spatial and spectral sampling resolutions which may result in variabilities in their spectral sensitivities and responses to the same water surface. To determine the significance of the spectral responses, the uniformly corrected sensor bands are compared for the signal-to-noise ratio (SNR) and top of atmosphere (TOA) reflectance over a spatially uniform and clear water body area.

3.1.3. Lambertian BRDF-Adjusted Reflectance. Because most land surfaces do not have a Lambertian Bidirectional Reflectance Distribution Function (LBRDF), the reflectance may vary simply due to factors such as the variations in the solar radiation and viewing angle geometry. Because of BRDF effects, the Landsat and Sentinel-2A swath across the red and near-infrared band reflectances can vary by 0.02–0.06 and 0.06–0.08, respectively [28]. These differences and variations can result in significant noise effects as they are comparable and greater than the sensor calibration errors [29]. For the LBRDF reflectance adjustment, the semiphysical approach developed for Landsat application and demonstrated to also work for Sentinel-2A by [28] was used for the adjustment of the surface reflectance for each comparable band to a 0° nadir view and the observed solar zenith angle [23].

3.1.4. Multisensor Crossband Adjustment. For compatibility and comparability of the Landsat OLI and Sentinel-2A bands, the observed differences in the homologue bands (Table 1) are crossadjusted as discussed in [30, 31]. The need for band adjustment is to minimize the effects of the bandwidth differences as depicted in Table 1. From Table 2, the six bands (blue, green, red, NIR, SWIR-1, and SWIR-2) were adjusted using the crosssensor transformation coefficients (Table 2) as derived from the results by [31]. Approximately the same crosssensor adjustment coefficients were obtained [32] in the harmonization of Landsat OLI and Sentinel-2A MSI sensor data bands.

3.2. Empirical Regression Modelling for Retrieval of Water Quality Parameters. As presented in the introductory and literature review sections, this study proposes the use of EMRM, which follows also from an earlier approach as presented in [3]. Using EMRM, simulations are carried out to establish the multivariate correlations between the sensor band reflectances and the measured *in situ* water quality variables. Table 3 presents the remote sensing reflectance band combination(s) considered for the EMRM analysis of the Sentinel-2A and Landsat OLI data in the estimation of the water quality parameters. In Table 3, i , j , and k refer to the sensor bands from the Landsat OLI and Sentinel-2A datasets (B_{L8} and B_{S2A}), before conversion to remote sensing reflectance $R_{rs}(\lambda)$ and multiband adjustment using the crosssensor band transformations.

The multivariate regression model for estimating the water quality parameters in the reservoir is developed by determining the quantitative relationships between the measured *in situ* water quality parameter and the remote sensing reflectance from the satellite spectral data. The empirical models used in the regression of the *in situ* measurements to the sensor band combinations were comprised of the following model equations [3], where $R_{rs}(\lambda)$ is the corresponding remote sensing reflectance for B_{L8} and B_{S2A} bands and a , b , and c are the regression model constants.

(a) *Linear.* $a * R_{rs}(\lambda) + b$.

(b) *Polynomial.* $a * R_{rs}(\lambda)^2 + b * R_{rs}(\lambda) + c$.

TABLE 2: Crosssensor band transformation coefficients for Landsat-8 [31].

Bands	Landsat-8 OLI sensor		Multisensor adjustment equation
	Intercept	Slope	
Blue	-0.0107	1.0946	$B_{S2A} = \text{slope} * B_{L8} + \text{intercept}$, where B_{S2A} is the Sentinel-2A band and B_{L8} is the Landsat OLI band
Green	+0.0026	1.0043	
Red	-0.0015	1.0524	
NIR	+0.0033	0.8954	
SWIR-1	+0.0065	1.0049	
SWIR-2	+0.0046	1.0002	

TABLE 3: Sentinel-2A MSI and Landsat-8 OLI band combination(s) for water quality parameter retrieval using the proposed EMRM.

Band combination(s)	Sentinel-2A bands	Landsat OLI bands
Single bands	$B_{S2A1} = B_{S2Ai}$	$B_{L81} = B_{L8i}$
Linear band combination	$B_{S2A2} = B_{S2Ai} + B_{S2Aj}$	$B_{L82} = B_{L8i} + B_{L8j}$
Band ratios	$B_{2A3} = B_{S2Ai}/B_{S2Aj}$	$B_{L83} = B_{L8i}/B_{L8j}$
Mixed band combinations	$B_{S2A} = (B_{S2Ai}/B_{S2Aj}) + B_{S2Ak}$	$B_{L84} = (B_{L8i}/B_{L8j}) + B_{L8k}$

B_{L8} and B_{S2A} are, respectively, the Landsat OLI and Sentinel-2A bands, and B_{L81-4} and B_{S2A1-4} are the four band combination models 1-4.

(c) *Logarithmic*. $a * \log_{10} R_{rs}(\lambda) + b$.

(d) *Power*. $a * R_{rs}^b(\lambda)$.

(e) *Exponential*. $a * e^{b * R_{rs}(\lambda)}$.

In the development of the EMRM, ten (10) of the sampling point data were used in the regression modelling in model calibration and five (5) remaining data points (T2, T5, T7, T8, and T13) were used in the validation of the model. Surface water samples from T1 to T15 were sampled in January 2019 during the dry season and those from T16 to T18 were sampled in May after the onset of the raining season. To determine the best-fit model, the correlation of determination R^2 between the predicted model and laboratory-measured water quality parameter is ranked, and the model fit is tested at the entry significance level of $p = 0.05$ and the removal significance level of 0.10. The t -test at the confidence level of 95% is used to accept the determined regressive model. The sampling point-based retrieved WQPs are also spatially mapped to determine the spatial distribution of the concentrations of the WQP using ordinary Kriging [3].

3.3. WQP Retrieval Performance Analysis Metrics. To determine and compare the performance between the sensor-based empirical models in the retrieval of water quality parameter, the regression results were compared with the *in situ* laboratory measurements using the following error matrices: Pearson correlation coefficient R , coefficient of determination R^2 , mean absolute error (MAE), root mean square error (RMSE), and mean absolute percent error (MAPE%) in Table 4. In Table 4, x_i and y_i are, respectively, the laboratory-measured (observed) and the regression model-predicted water quality parameter concentrations at each sample point i and for the n samples.

Figure 2 presents a summary of the methodological approach in the retrieval of the water quality parameters.

4. Results and Discussions

4.1. Signal-to-Noise Ratio (SNR) on Clear Water Bodies. As proposed in Section 3.1.2, the sensors are compared in terms of their responsive SNR indices at specific spectral bandwidths. From spatially uniform and clear water bodies, the Sentinel-2A and Landsat OLI SNR radiometric performances were determined using L1C images for the visible and NIR bands. Notably, for land targets, [33] provided the SNR requirements for the typical radiance. Figure 3 shows the SNR computed from the average of the local window area from the ratio of the area mean to standard deviation ($\text{SNR} = \mu/\sigma$). The SNR is determined using a 3×3 pixel kernel on the mean TOA. The selected water body areas are outlined in Figure 4 (spatial areas A and B).

In the three visible bands blue (490 nm), green (560 nm), and red (665 nm), the Landsat OLI bands exhibit SNR values of 2-3 times higher than the corresponding MSI bands. In the coastal aerosol bands (443 nm), MSI is at approximately 30% higher than OLI in SNR. Theoretically therefore, it can be inferred that since SNR is proportional to the square root of the area of a pixel [34], then the aggregated 20 m-30 m spatial resolution of the MSI bands should be able to offer similar or better radiometric quality in the visible bands as compared to the OLI bands for the clear water areas. The same argument can be applied to the aggregated 60 m OLI coastal aerosol band with 30% higher SNR in the corresponding MSI band. With band adjustment and scaling, the overall radiometric quality of MSI and OLI products can be deduced to be comparable over aquatic systems [35]. The multisensor band adjustment and scaling are presented further in the methodology section.

TABLE 4: Water quality estimation performance analysis metrics.

Error estimator	Error equation
Pearson correlation coefficient and coefficient of determination	$R = \sum_{i=1}^n (y_i - \bar{y}) \cdot (x_i - \bar{x}) / \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$, R^2
Mean absolute error	$MAE = 1/n \sum_{i=1}^n x_i - y_i $
Root mean square error	$RMSE = \sqrt{1/n \sum_{i=1}^n (x_i - y_i)^2}$
Normalized mean absolute error percent	$NMAE\% = \left[1/n \sum_{i=1}^n (x_i - y_i /x_i) \right] * 100$

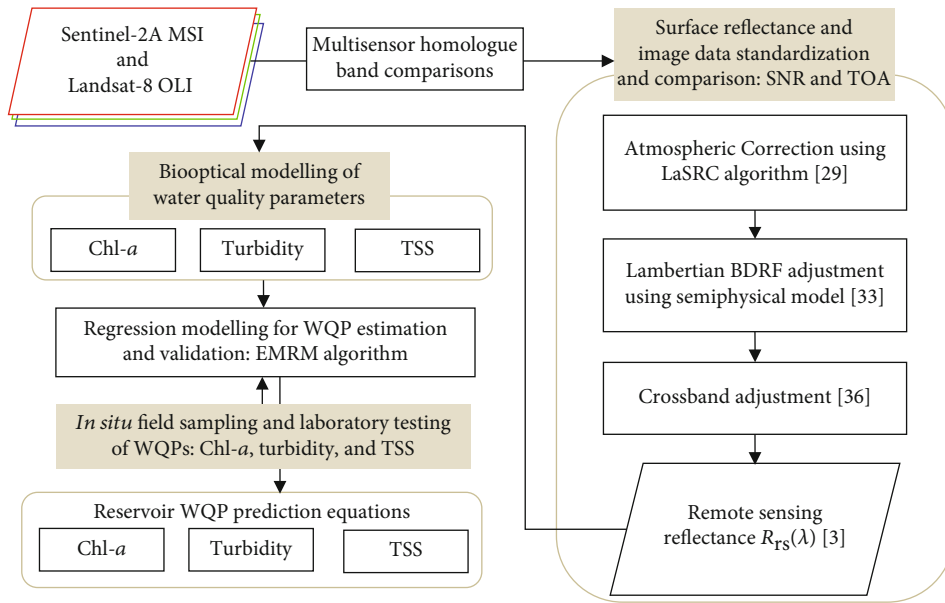


FIGURE 2: Schematic workflow of the approach for the retrieval of water quality parameters from Landsat OLI and Sentinel-2A MSI. WQPs are the water quality parameters, EMRM is empirical multivariate regression modelling algorithmic approach, SNR is the signal-to-noise ratio, and TOA is the top of atmosphere reflectance.

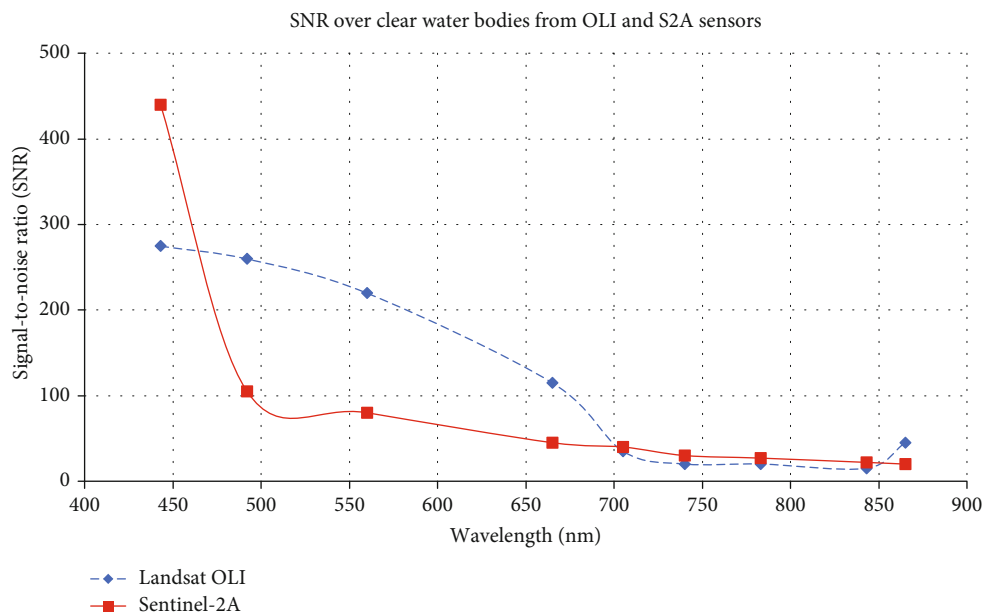


FIGURE 3: SNR on clear water bodies in the visible and NIR bands from Sentinel-2A MSI and Landsat-8 OLI.

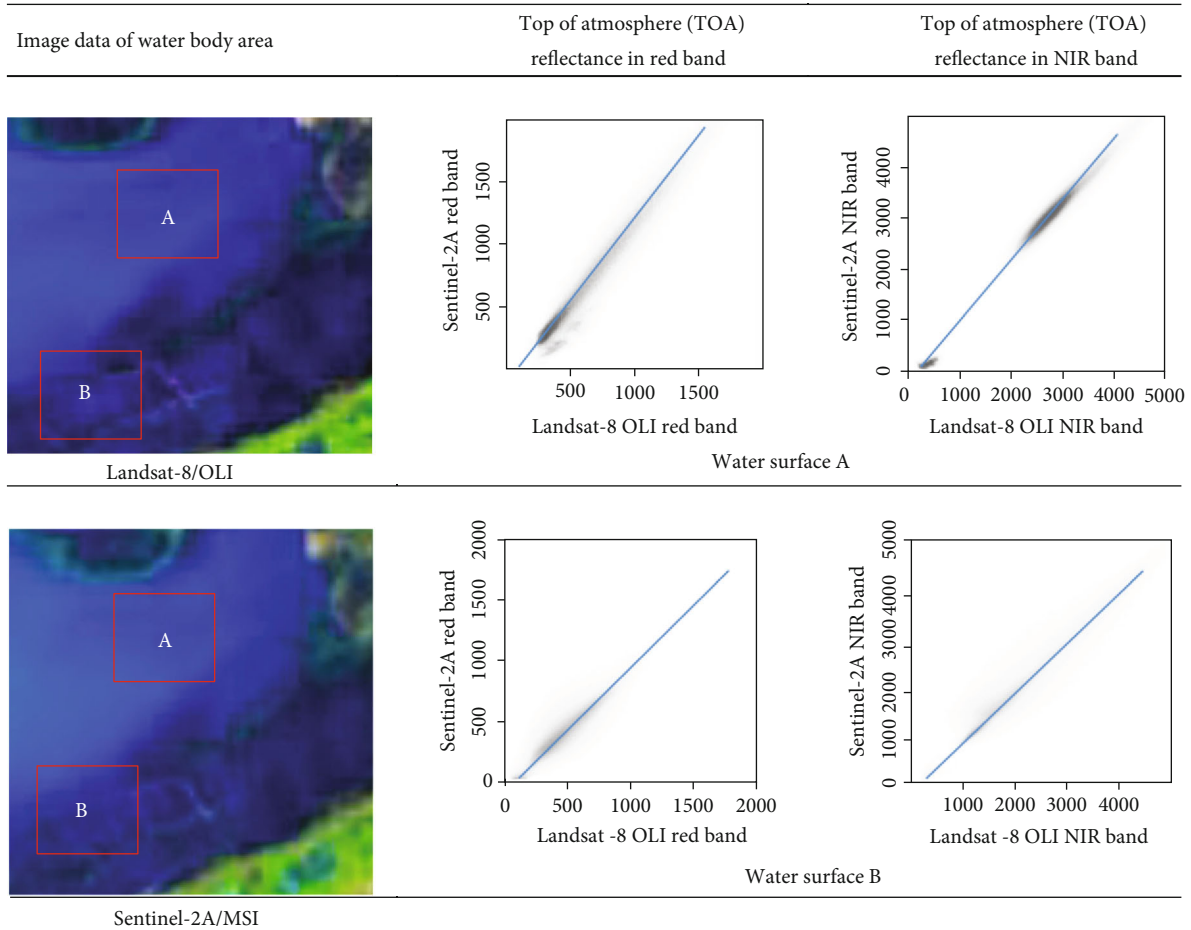


FIGURE 4: The TOA reflectance correlation scatterplots from Landsat OLI and Sentinel-2A reflectance in red and near-infrared bands for clear water surfaces (A and B).

4.2. Sentinel-2 MSI and Landsat-8 OLI TOA Reflectance Comparison. When comparing the TOA and surface reflectance from the two sensors, the spatial subsets (A and B) for the selected water bodies are shown in Figure 4. In terms of the interband spectral responses for the selected water body areas, the scatterplots between the Landsat OLI and Sentinel-2A spectral bands showed a high correlation of $R^2 > 0.876$, for all the seven homologue bands in Table 1. The regression plots in Figure 4 present sample results from the correlations between the red and NIR bands from the spatial subset of the two water body scenes A and B. The red and NIR bands represent the reflectance in the visible and NIR bands, which are commonly used in water quality modelling [3, 36–38]. The regression results indicate that the two sensors contain nearly similar spectral radiance information, despite the marginal difference in the SNR (Figure 3).

Figure 4 shows that for the same scene water body, the sensors have a close crosscorrelation in the red and NIR bands, which is also observed in the TOA reflectance plot in Figure 5 with the TOA from Landsat OLI wavelengths being slightly higher than that from Sentinel-2A bands. In Figure 5, the TOA reflectance values are scaled by a factor of 10,000 and the 10 m Sentinel-2A bands were resampled to the 30 m spatial resolution to match the Landsat OLI.

The moderately higher TOA reflectance in Landsat OLI confirms the higher SNR. The results in Figure 5 show that the Landsat OLI bands present a slightly higher TOA reflectance than the Sentinel-2A bands.

4.3. Comparisons of Sensor Remote Sensing Reflectance $R_{rs}(\lambda)$. A comparison of the remote sensing reflectance from the 15 sampling points from the Sentinel-2A and Landsat OLI is presented in Figure 6, indicating that after sensor band adjustment, the reflectance in the sensor bands ranges from 0.003 to 0.009 Sr^{-1} . For both sensors, the visible and NIR bands show characteristically higher reflectance as compared with the SWIR bands over the reservoir. The sampling stations T2, T6, T9, and T11 depict characteristically higher reflectance than the other sampling points. These sampling points, respectively, represent regions within the dam where there is inflow, the edge of the dam, the deepest and most shallow sections of the dam. In general, the reflectance pattern within the dam follows the same pattern from T1 to T15 especially in Sentinel-2A data, with the deepest point having the highest reflectance of 0.25 Sr^{-1} in Landsat OLI and 0.087 in Sentinel-2A MSI NIR and blue bands, respectively. In Figure 6, OLI shows the highest reflectance in the visible bands (B2–B4) and NIR (B5) bands, while Sentinel-2A shows relatively higher

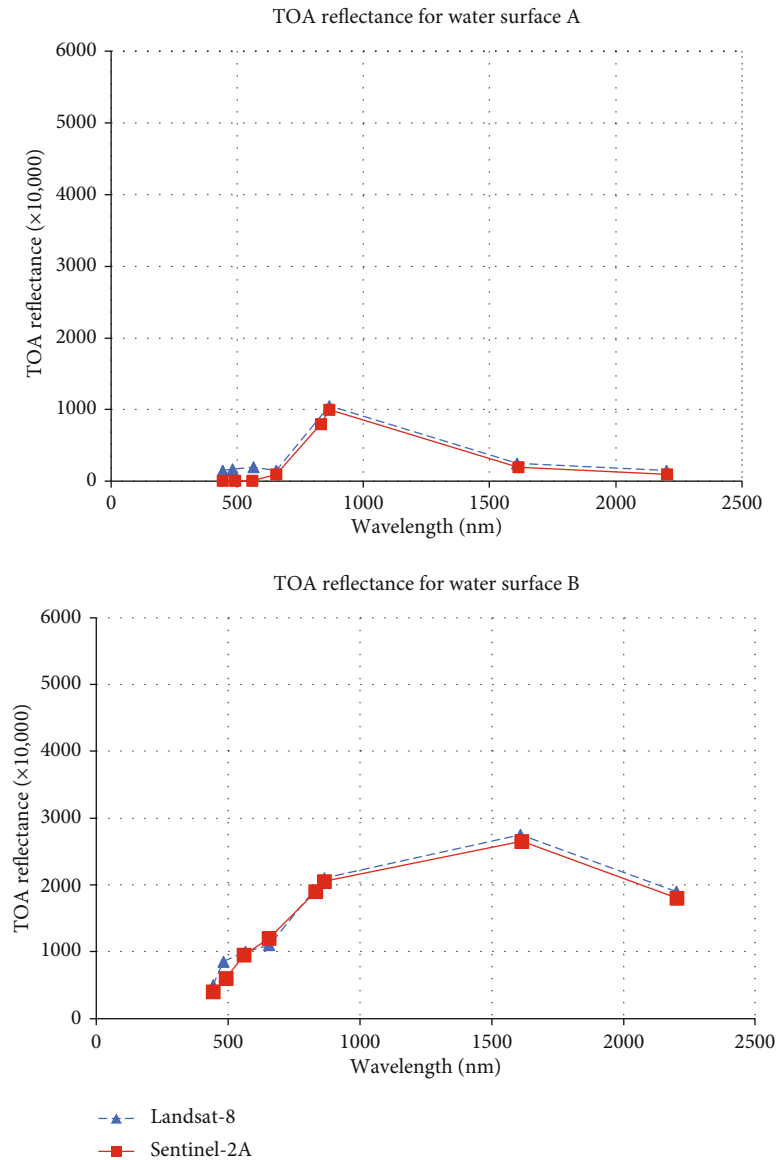


FIGURE 5: Comparison between TOA reflectance from Landsat-8 OLI and Sentinel-2A MSI for clear water surfaces.

TOA reflectance in all the bands at above 0.003 Sr^{-1} . This observation is similar to the observed SNR results in Figure 3, and the near-uniform reflectance from the two sensors can be attributed to the band adjustment using the LaSRC algorithm [27].

4.4. Water Quality Estimations from Sentinel-2A and Landsat OLI

4.4.1. Retrieval of Chlorophyll-*a*. The *in situ* test results determined the concentrations of chlorophyll-*a* to vary considerably within the reservoir with values between $6.23 \mu\text{g/L}$ and $113 \mu\text{g/L}$ and an average of $29.91 \mu\text{g/L}$. Since the reservoir is protected from public use, the observed Chl-*a* concentrations are attributed to the inflow of fertilizer leachate into the reservoir from the surrounding farms. The leachate process carries nutrients which encourage the growth of algae in

the reservoir. This is likely to cause severe eutrophication if not monitored on a continual basis.

From the empirical regression modelling, Figures 7 and 8, respectively, present the best regression models for the estimation of the concentration of Chl-*a* from Sentinel-2A and Landsat OLI. For Sentinel-2A (Figure 7), a second-order polynomial fit was found to be suitable for the retrieval of Chl-*a* using $R_{rs}(\lambda)$ from the difference between the green (B3) and the SWIR-1 (B11) band, with $R^2 = 0.7015$. This was approximately 15% lower in accuracy as compared to the linear model derived from the prediction of Chl-*a* from Landsat OLI with $R^2 = 0.8581$ from the green band (B3) (Figure 8). While the green band is observed to be significant in the detection of Chl-*a* from both sensors, the advantage of Landsat OLI could be because of its wider spectral bandwidth in the green band, i.e., approximately 57 nm as compared to 35 nm in Sentinel-2A (Table 2),

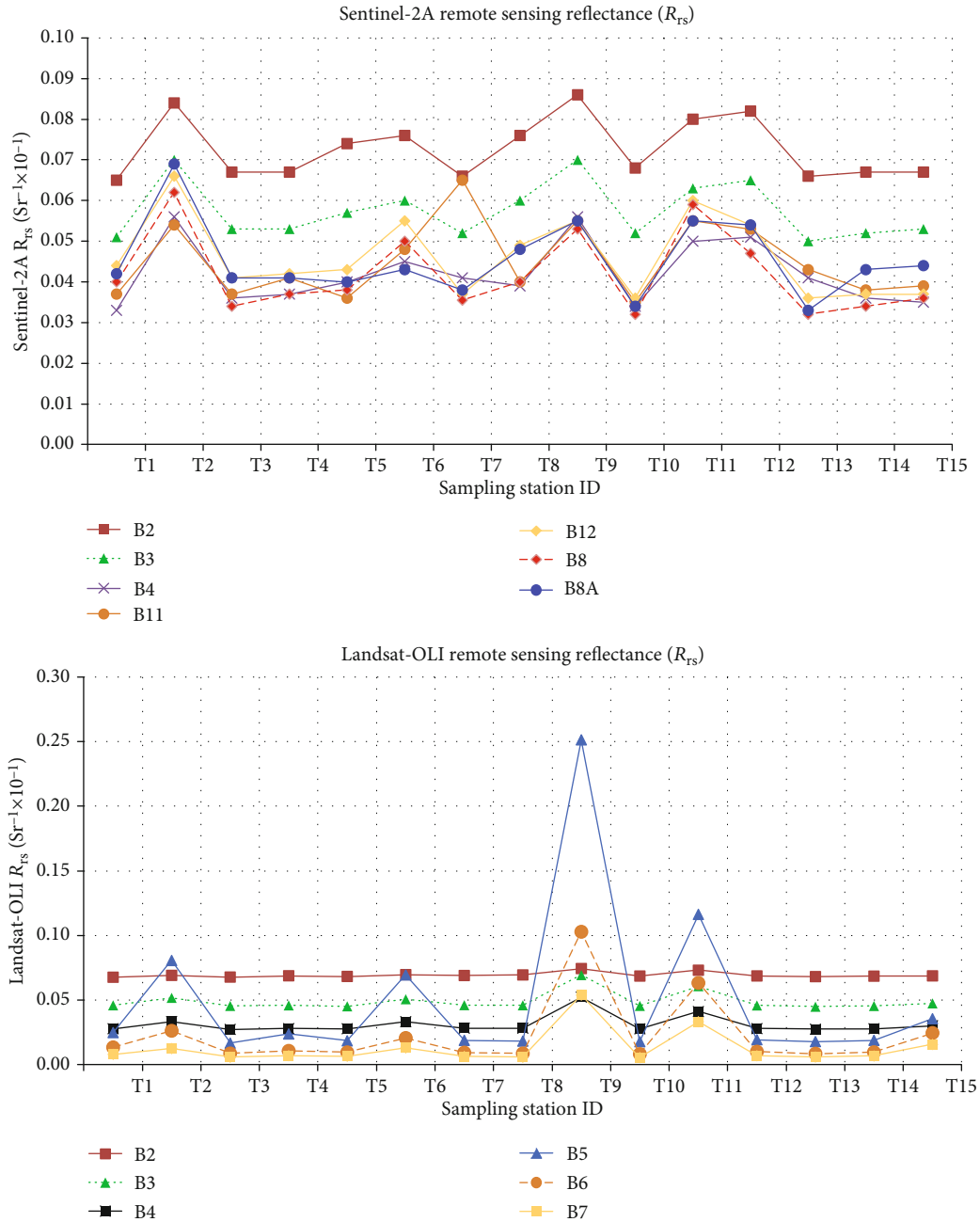


FIGURE 6: Remote sensing reflectance variations over sampling points.

and therefore, the placement of the green band and the SNR in the Landsat OLI is suitable for Chl-*a* retrieval.

The estimated Chl-*a* from Sentinel-2 and Landsat-8 both have a lower RMSE value compared to the mean *in situ*-measured Chl-*a* (29.91 $\mu\text{g/L}$). This confirms the appropriateness of the developed regression model for estimating Chl-*a* in the case study dam at the given acquisition.

A summary of the performance of the two sensors is given in Tables 5 and 6 for the best five fit EMRM results. The results show that Landsat OLI is able to detect the presence of Chl-*a* with R^2 of up to 0.86, as compared to Sentinel-2A with R^2 of up to 0.70. Similarly, the statistical trends from MAE, NMAE%, and RMSE are the same at p value = 0.055,

with lower MAE and RMSE in Landsat OLI, but relatively the same NMAE% error magnitudes for Chl-*a* retrieval from both sensors. The inference from the results is that both the sensors are suitable for the retrieval of Chl-*a*, specifically using the green and blue bands, and thus, as opposed to the results from [39], the red and NIR bands are not informative in the retrieval of water quality in the inland waters in this case study. The green and blue wavelengths are suitable for the detection of algal bloom due to the high reflectance of green algal matter, which is in the bluish color waters.

Compared to previous study results on the estimation of Chl-*a*, [38] compared Landsat-8 L1T and Sentinel-2A L1C with *in situ* measurements for chlorophyll-*a* predictions in

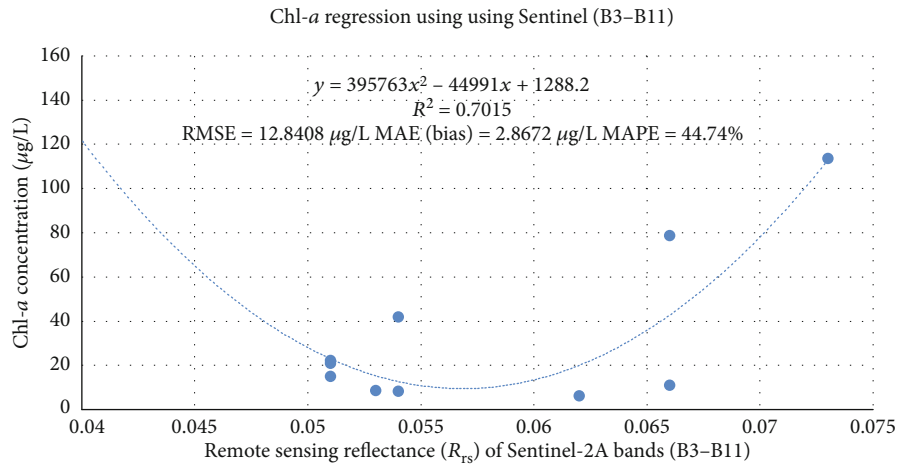


FIGURE 7: *In situ*-derived chlorophyll-*a* and Sentinel-2A band reflectance.

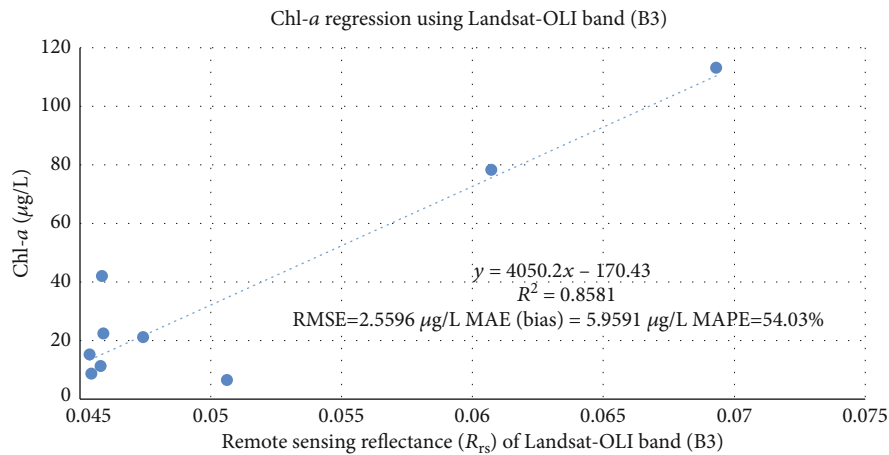


FIGURE 8: Landsat band reflectance and *in situ* Chl-*a* regression.

TABLE 5: Regression model for the retrieval of chlorophyll-*a* from Sentinel-2A bands.

Sentinel-2A regression model equation for Chl- <i>a</i> estimation	Band combination for deriving $R_{rs}(=x)$	R^2	MAE	NMAE%	RMSE
$y = 466683x^2 - 39497x + 835.87$	B3 + (B12-B11)	0.6915	0.1613	8.7155	3.0357
$y = 378524.04x^2 - 30345.60x + 616.37$	B3	0.6300	3.4864	73.1354	27.3437
$y = 395763x^2 - 44991x + 1288.2$	B2-B11	0.7015	2.8672	44.7386	12.8408
$y = 246622x^2 - 38346x + 1497.6$	(B2 + B12)/2	0.6291	6.0437	67.4678	23.2026
$y = 1748431.97x^2 - 126132.79x + 2289.87$	(B2-B11)/2	0.6500	10.2352	73.3885	18.6211

TABLE 6: Regression models for the estimation of chlorophyll-*a* from Landsat OLI bands.

Landsat-8 OLI regression model equation for Chl- <i>a</i> estimation	Band combination for deriving $R_{rs}(=x)$	R^2	MAE	NMSE%	RMSE
$y = 3188.9x - 349.09$	B2 + B3	0.8505	1.4128	59.5924	8.2940
$y = 4050.2x - 170.43$	B3	0.8581	2.5596	54.0317	5.9591
$y = 2749.1x - 539.88$	B2 + B3 + B5	0.8568	0.4808	60.9354	9.9056
$y = 402.73 \ln(x) + 888.88$	B2 + B3	0.8522	1.9835	60.1054	7.7013
$y = 593.38 \ln(x) + 964.48$	B2 + B3 + B4	0.8507	0.8458	61.4433	9.4775

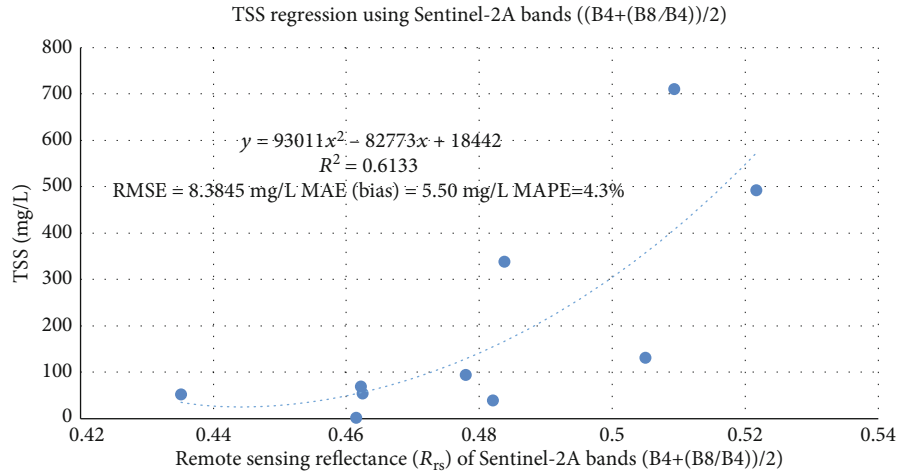


FIGURE 9: Measured TSS plotted against Sentinel-2A reflectance.

the Cassaffouth Reservoir in Argentina. For Landsat, the model with the combination of the blue band and blue/NIR and green/NIR band ratios produced the best results with $R^2 > 0.89$. The best Sentinel MSI model was a two-variable model using both the blue band and the (green/NIR band ratio with $R^2 = 0.880$. [13] also estimated chlorophyll-*a* for Lake Peipsi and Lake Vörtsjärvi using Sentinel MSI with $R^2 = 0.83$, using blue and green bands. In attempts to estimate chlorophyll-*a*, [40] showed that Landsat OLI produced average but better results for chlorophyll-*a* estimations for the case study of Lake Trichonis in Greece, as compared to its predecessors in the Landsat series, using the visible and NIR bands. [41] investigated the potential of deriving chlorophyll-*a* and turbidity using Secchi disk depth in six (6) natural and five (5) artificial lakes in Greece, using Landsat-8's blue, red, and SWIR-2 bands. The regression validation indicated successful correlations with R^2 of 0.75 for chlorophyll-*a* estimations. [11, 39] derived the chlorophyll-*a* for Barra Bonita tropical reservoir using Landsat OLI. The results showed that the NIR-red, NIR-green, and NIR-blue band ratios yielded the best results with R^2 greater than 0.70 for the estimation of Chl-*a* concentrations.

In conclusion and compared to previous studies, the current study further confirms the suitability of the visible and NIR bands in the prediction of chlorophyll-*a* in inland water bodies. The detection of the concentration of chlorophyll-*a* using the developed empirical equations can be an indicator for monitoring the abundance of algae in inland water bodies, since it presents active optical properties in the visible and near-infrared wavelengths.

4.4.2. Estimation of TSS Concentration. The concentrations of suspended particulate matter from the laboratory measurements were low and within acceptable limits ranging between 39 mg/L and 711 mg/L, with a reservoir average of 187 mg/L. The highest concentrations were in the same regions as for Chl-*a* and are attributed to the inflow of water from a river at an area of low depth which leads to the agitation of the sediments from the bottom of the reservoir.

Because the reservoir is deep, the sediments tend to settle faster leading to generally low concentrations of particulate matter. However, the high presence of particulate matter at a given time and region propagates the concentration of algal bloom and hence the observed high concentrations of Chl-*a* in the same reservoir region.

Using the EMRM algorithmic approach, the best results for the estimation of TSS concentration from Sentinel-2A were obtained using a second-order polynomial function relating the red band (B4) and the NIR band (B8) with R^2 of 0.6113 (Figure 9), as compared to $R^2 = 0.9249$ from the Landsat OLI band ratio between the green band (B3) and the blue band (B2) using linear regression modelling (Figure 10). While the mean concentration of TSS from the in situ measurements is at 187 mg/L, the RMSE and bias are averaged and less than 10 mg/L. This confirms the validity of the developed model in estimating TSS within the reservoir.

The overall performance of the satellite sensors in the retrieval of TSS within the reservoir is summarized in Tables 7 and 8 for the best five results. As in the retrieval of Chl-*a*, the visible and NIR bands are the optimal bands in the estimation of TSS. The accuracy of estimation of TSS from Landsat OLI (Table 8) is notably higher than that from Sentinel-2A by at least 30% as measured in terms of R^2 . The MAE, NMAE%, and RMSE are also observed to follow the R^2 trend of higher values from Sentinel-2A predictions as compared to the predictions from Landsat OLI. The results in Tables 7 and 8 show that the placement of the visible bands in Landsat is significant in TSS estimation.

The significance of OLI band 3 and band 2 in TSS estimation is noted and is attributed to the fact that the deeper the water, the more blue light it reflects and hence the higher the reflectance in blue and green bands. It is also known that the particle settlement is higher in deep quiescent water. Therefore, by relation, the higher the blue light detected, the lower the concentration of particles in the water. In the case of the green band, most of the suspended matter in the water was in the form of algae. Therefore, the higher the

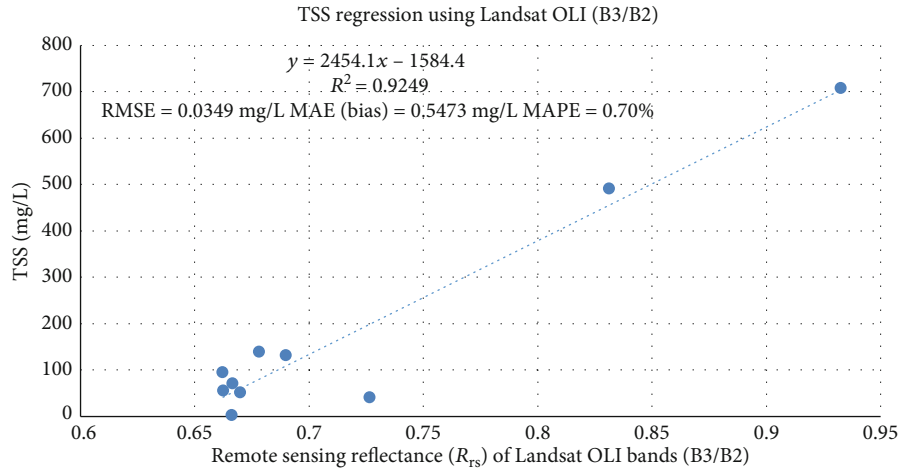
FIGURE 10: Landsat band remote sensing reflectance and *in situ*-measured TSS concentration.

TABLE 7: Regression model for the estimation of TSS from Sentinel-2A MSI.

Sentinel-2A regression model equation for estimation of TSS	Band combination for deriving R_{rs} ($= x$)	R^2	MAE	NMAE%	RMSE
$y = 93011x^2 - 82773x + 18442$	$(B4 + (B8/B4))/2$	0.6133	5.4969	4.3039	8.3845
$y = 89013x^2 - 78936x + 17526$	$(B8A + (B8/B4))/2$	0.5940	9.5691	6.8105	13.4793
$y = 83708x^2 - 73811x + 16299$	$(B8 + (B8/B4))/2$	0.5998	6.6792	7.5058	12.9244
$y = 19420x^2 - 26272x + 8945.4$	B4/B3	0.4846	23.7299	15.7569	24.8370
$y = 46042x^2 - 32585x + 5822.6$	$(B3 + (B4/B3))/2$	0.4722	13.9118	12.9250	21.7248

TABLE 8: Regression model for the estimation of TSS from Landsat-8 OLI.

Landsat OLI regression model equation for estimation of TSS	Band combination for deriving R_{rs} ($= x$)	R^2	MAE	NMAE%	RMSE
$y = 2781.9x - 1393.5$	B2	0.9240	4.7391	12.3842	3.3185
$y = 27176x - 1184.6$	B3	0.9245	6.8510	29.896	12.8425
$y = 2454.1x - 1584.4$	B3/B2	0.9249	0.5473	0.7000	0.0349
$y = 1714.7 \ln(x) + 1171$	B3/B4	0.9161	2.3885	9.3736	2.1329
$y = 1905.3 \ln(x) + 821.01$	B3/B2	0.9159	3.7496	12.1181	4.5714

amount of light reflected in the green wavelength, the higher the concentration of algae and thus the higher the amount of suspended matter.

Compared to previous studies on TSS estimations using satellite data, [36] utilized Landsat OLI for retrieval of suspended sediment concentrations in the Orinoco River in Venezuela. The best performance for TSS estimations was from the red and near-infrared wavelengths, which is similar to the results from Sentinel-2 MSI in this study. [21] also found that for suspended particulate matter predictions for Poyang Lake in China using empirical models, Sentinel-2A MSI band 7 was suitable for sediment-laden waters while Sentinel-2A MSI band 4 was suitable for clear waters. In Playa Colorada Bay, Mexico, [42] generated empirical models for TSS and turbidity estimations using the Landsat imagery. The correlation results between the Landsat reflectance

and the *in situ* measurements indicated that both water quality parameters could be estimated with R^2 of between 0.637 and 0.955, which is comparable with current results using Landsat OLI. [37] also estimated TSS and turbidity for Wular Lake in Kashmir. The parameters were correlated with single Landsat OLI bands and produced R^2 greater than 0.5, while [43] estimated TSS with R^2 of 0.709 using visible bands for Poteran Island waters in East Java, Indonesia.

The results from the current study show the significance of the green, red, and near-infrared wavelengths in the estimations of TSS in inland reservoir water bodies, with specially higher accuracy using Landsat-8 OLI wavelengths. For TSS retrieval from Sentinel-2 MSI, however, the marginal performance could be attributed to interactions of the optical electromagnetic radiations with the atmosphere and resulting low water surface reflectance into the sensor (glint). Thus,

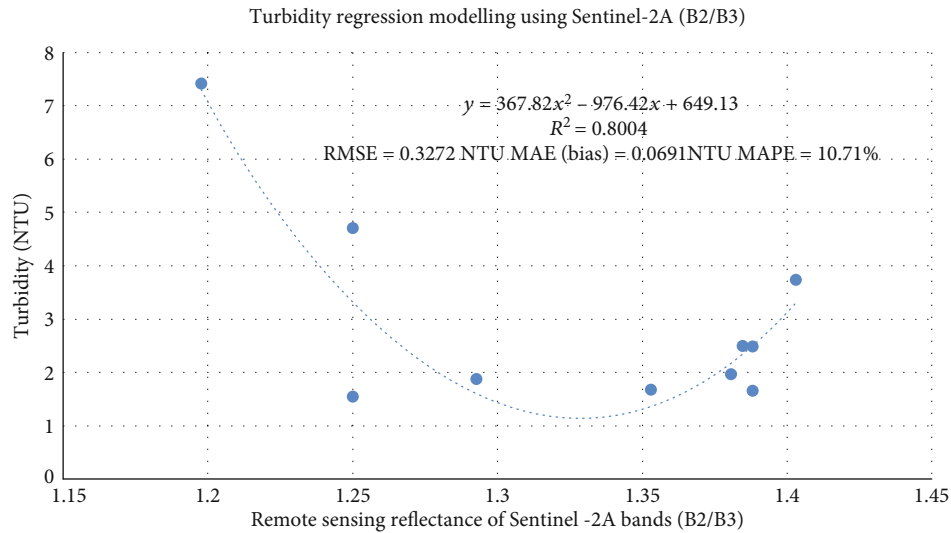


FIGURE 11: Laboratory-measured turbidity and Sentinel-2A band reflectance.

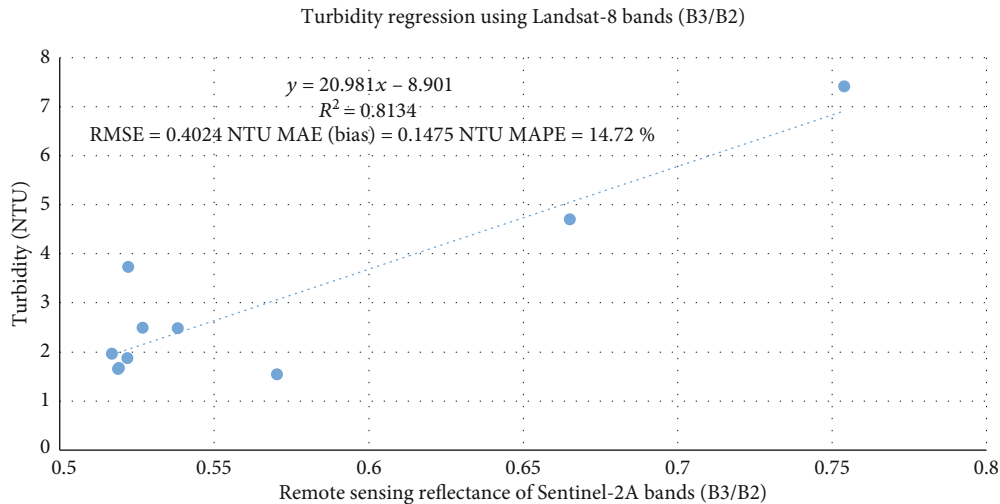


FIGURE 12: Landsat band reflectance and *in situ* turbidity measurements.

further correction of the Sentinel MSI R_{rs} signal is proposed to improve the accuracy in modelling TSS using Sentinel-2 data [13, 38].

4.4.3. Estimation of Turbidity Concentration in Reservoir Waters. With *in situ* turbidity varying from 1.36 NTU to 7.42 NTU and averaging at 2.83 NTU, the reservoir is characterized by low turbid waters. The low turbidities could be attributed to low flows into the reservoir, especially during the period in which the water samples were collected. With minimum inflow of sediment-laden rainwater discharge and the slow velocities of the river water, the concentrations of sediments that would cause reservoir turbidity were low. Further, at a maximum depth of approximately 150 m, the reservoir depth enables the settlement of sediments at the bottom of the reservoir with minimal potential of resuspension by water currents and waves. The higher degree of settling of sediments therefore leads to low turbidity in the reservoir.

From the empirical multivariate regression modelling comparing the satellite reflectance data and the laboratory-measured water quality factor, the study results show that the turbidity was estimated from the two sensors with the same accuracy of $R^2 > 0.7$ (Figures 11 and 12). For both the sensors, the visible bands (blue, green, and red) are observed to be predominant in the retrieval of turbidity within the reservoir. In Figure 11, the best estimate of turbidity using Sentinel-2A is at $R^2 = 0.8004$, which is comparable to turbidity estimates using Landsat OLI with $R^2 = 0.8134$. The models accurately predicted the concentrations of turbidity from both sensors as the RMSE and bias error measures were less than the minimum and average *in situ* turbidity.

Tables 9 and 10 give a summary of the best regression models for the retrieval of turbidity, showing significant contributions from the visible bands from both sensors. The error metrics MAE, NMAE%, and RMSE are, respectively, in the same magnitude range for both sensors. Like in the

TABLE 9: Regression model for the estimation of turbidity from the Sentinel-2A MSI sensor.

Sentinel-2A regression model equation for estimation of turbidity	Band combination for deriving R_{rs} ($= x$)	R^2	MAE	NMAE%	RMSE
$y = 367.82x^2 - 976.42x + 649.13$	B2/B3	0.8004	0.0691	10.7144	0.3272
$y = 971.47x^2 - 1468x + 555.84$	B3/B2	0.7990	0.0746	10.4636	0.3290
$y = 725.32x^2 - 858.52x + 255.91$	B4/B3	0.7353	0.7609	47.9432	1.5866
$y = 118.8x^2 - 401.92x + 341.62$	B2/B4	0.7372	0.7656	45.71415	1.5877
$y = 387.41x^2 - 1103x + 786.37$	B1 + (B1/B2)	0.7592	0.0035	6.9895	0.2192

TABLE 10: Regression model for the estimation of turbidity from the Landsat-8 OLI sensor.

Sentinel-2A regression model equation for estimation of turbidity	Band combination for deriving R_{rs} ($= x$)	R^2	MAE	NMAE%	RMSE
$y = 20.981x - 8.901$	B3/B2	0.8134	0.1475	14.7203	0.4024
$y = 102.56x - 5.5003$	B3 + B4	0.8125	0.1066	9.4289	0.7341
$y = 90.319x - 10.775$	B2 + B3 + B4	0.8123	0.1175	9.9483	0.7836
$y = 20.254 \ln(x) + 46.009$	B2 + B3	0.7978	0.0489	14.5470	0.6069
$y = 14.735 \ln(x) + 30.802$	B2 + B3 + B4	0.7954	0.0865	10.5043	0.7350

TABLE 11: Descriptive statistics of the laboratory-measured and predicted water quality parameters (SD = standard deviation; CV = coefficient of variation; SE = standard error).

Water quality parameter and estimation and data source	Sample (n)	Min	Max	Median	Average	SD	CV (%)	SE	
Chl- a ($\mu\text{g/L}$)	<i>In situ</i> WQP measurements	15	6.24	113.73	15.16	29.91	31.65	105.82	8.17
	Landsat-8 OLI	15	13.29	110.26	15.44	30.77	28.42	92.39	7.34
	Sentinel-2A MSI	15	11.38	221.80	23.04	40.82	56.19	137.65	14.51
Turbidity (NTU)	<i>In situ</i> WQP measurements	15	1.36	7.42	1.97	2.83	1.69	59.64	0.42
	Landsat-8 OLI	15	1.94	6.92	2.05	2.88	1.46	50.77	0.38
	Sentinel-2A MSI	15	1.23	26.25	2.48	4.99	6.51	130.52	1.68
TSS (mg/L)	<i>In situ</i> WQP measurements	15	39.00	711.00	95.00	187.00	197.88	102.23	49.36
	Landsat-8 OLI	15	38.12	702.60	60.23	169.68	195.70	115.34	50.53
	Sentinel-2A MSI	15	35.51	573.74	127.78	177.06	157.98	89.22	40.79

estimation of Chl- a and TSS, the linear and polynomial regression models are the best for the retrieval of the WQP. The turbidity estimation results are particularly important since turbidity is considered one of the most significant water quality parameters [44].

In the estimation of turbidity, the visible and NIR bands are observed to yield the best results. Comparatively, the accuracy of turbidity estimations is nearly equal for the two optical sensors indicating that the reflectance of suspended particulates is much higher in the lower wavelengths, irrespective of the band and or the bandwidth.

For the assessment of the extraction of turbidity as the significant water quality parameter that relates directly to the underwater light penetration that affects productivity, [44] showed from several band combinations that Landsat OLI band 4 (red) has the best correlation with the *in situ*-measured turbidity with a coefficient of determination R^2 of 0.84, scatter index of 0.22, and RMSE of 0.28 NTU. [45] also

recently demonstrated using the multivariate regression model and gene expression programming (GEP) that reservoir turbidity could be extracted with R^2 of greater than 90%, using the combination of visible and NIR bands. While high accuracy has been observed in turbidity estimation, in all the studies, atmospheric correction was found to be a factor in the accuracy of the WQP modelling [13, 22, 38].

4.5. Validation of Water Quality Prediction with In Situ Measurements. The validation of the developed regression models in Section 4.4 was carried out using the five (5) sampling stations (T2, T5, T7, T8, and T13). The validation results are presented in Table 11, including the statistics from the stations which were used in the model calibration. TSS is observed to have the highest variation in concentration, followed by Chl- a and turbidity (least). From the SD, CV, and SE metrics, Landsat OLI tended to underestimate the concentrations of the water quality

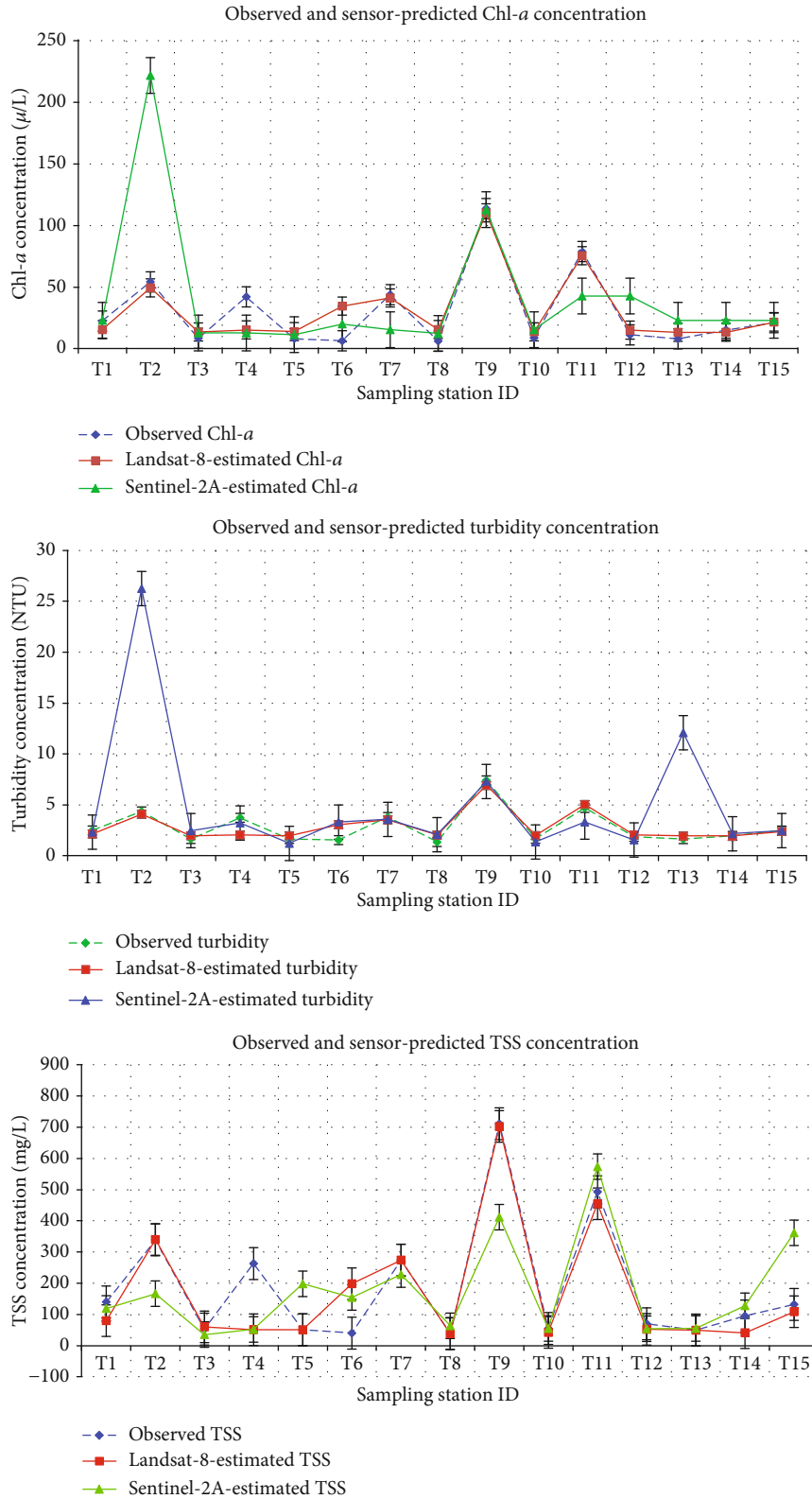


FIGURE 13: Prediction and validation sampling point variations of Chl-*a*, turbidity, and TSS within the reservoir.

parameters, while Sentinel-2A-based models tended to overestimate the same WQP, thus predicting the parameters with a higher coefficient of variation.

The graphical assessment of the validation results in Figure 13 for Chl-*a*, turbidity, and TSS shows that for the five (5) validation sampling stations, the retrieval of Chl-*a* and

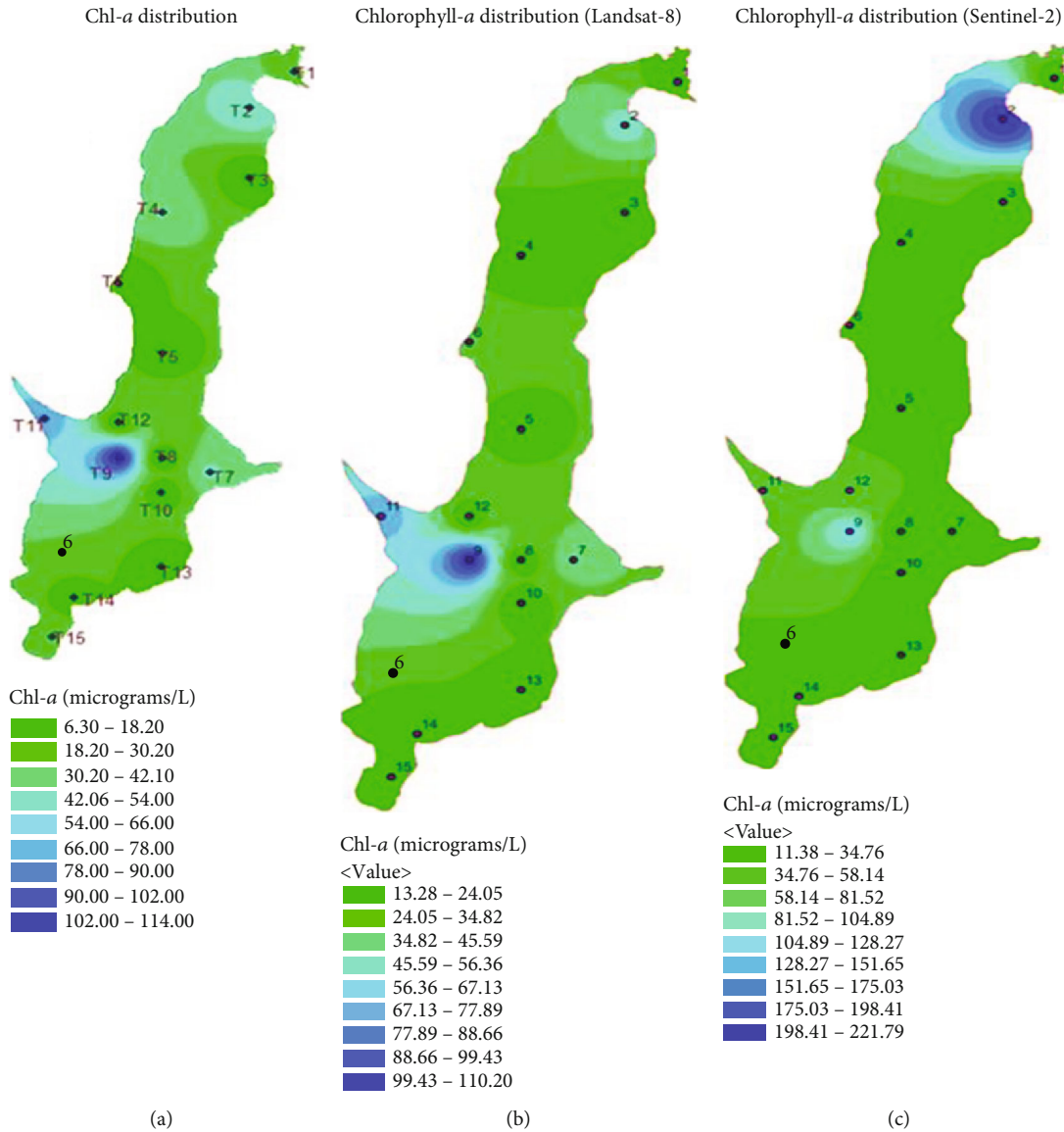


FIGURE 14: Spatial distribution of Chl-*a* concentration from (a) *in situ* measurements and from (b) Landsat OLI and (c) Sentinel-2A.

turbidity matched the observed measurements for both Sentinel-2A and Landsat OLI, except for sampling station T2, where Sentinel-2A grossly overestimated the Chl-*a* and turbidity concentrations by approximately 4 times. For TSS estimation, Table 9 and Figure 13 show lower coincidence between the measured and validated results with standard error at above 40% for both sensors.

The marginally lower performance of the Sentinel-2A in the prediction of the selected WQP may be attributed to the two factors, namely: (i) the sensor’s low spectral resolution in terms of the bandwidths and (ii) the lower signal-to-noise ratio (SNR) as compared to Landsat OLI visible and NIR bands. However, in the four visible channels and the NIR bands, the results of the study show that with adequate radiometric corrections or band adjustments, the OLI and MSI are both capable of monitoring the optically active water quality indicators. In synergy with the Landsat-9, the multi-sensor integration of the Sentinel-2A MSI and Landsat-8 will

result in a temporal revisit time of 2–3 days, making it possible to accurately capture the dynamics of the inland water body characteristics for continuous assessments.

In summary, the results for the prediction of chlorophyll-*a*, TSS, and turbidity demonstrate the potential for cost-effective, large-scale, and high-frequency use of optical satellite remote sensing reflectance data in monitoring optically active water elements. However, it is observed that the effectiveness and accuracy of water quality element determination are dependent on the satellite sensor used, the methodology followed, and also the nature of the waters studied.

4.6. Spatial Distribution of Optically Active Water Quality Parameters

4.6.1. *Chlorophyll-a Distribution.* Using ordinary Kriging as suggested in [3], the spatial interpolation of the concentrations of Chl-*a* observed and estimated from Sentinel-2A

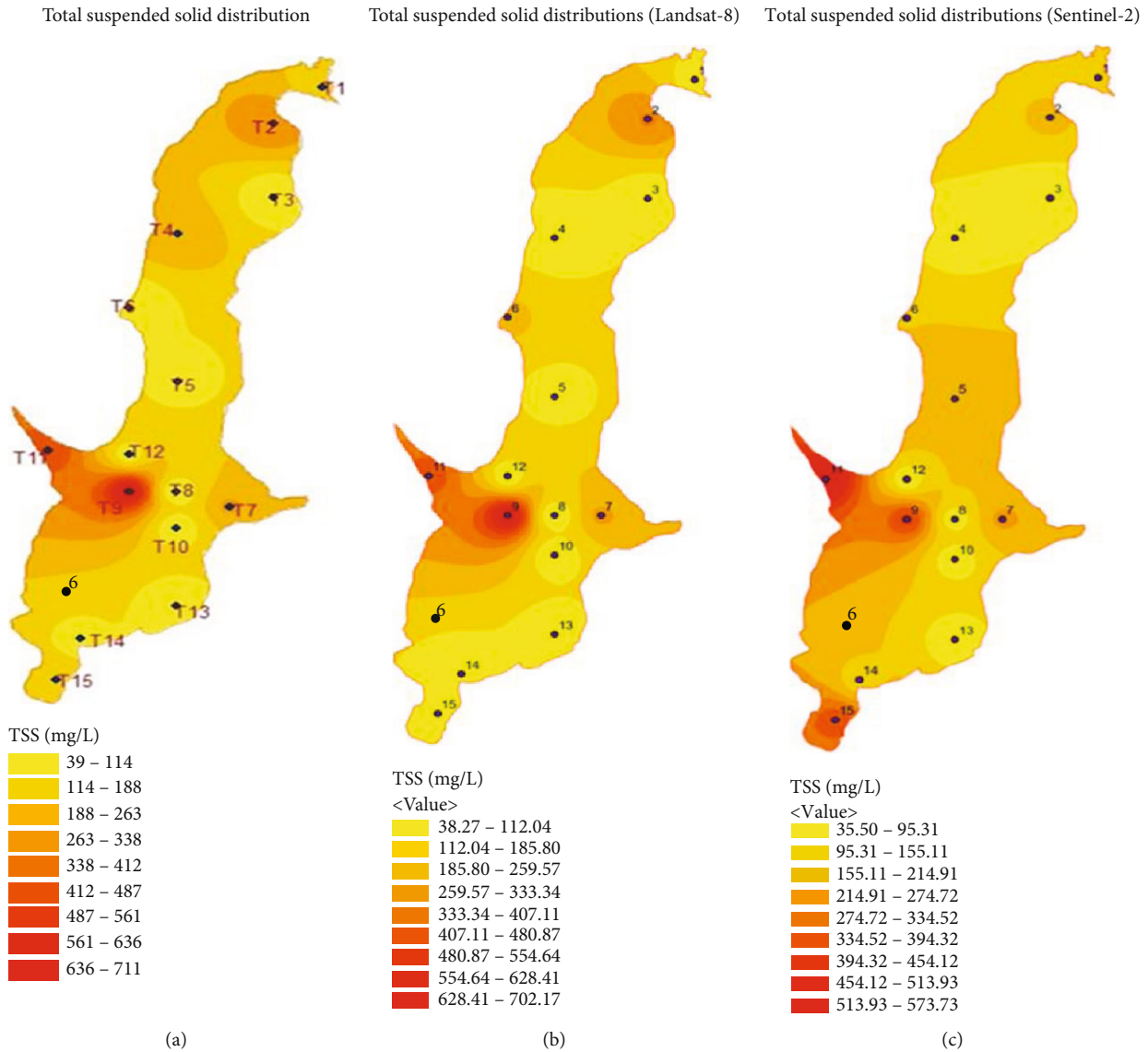


FIGURE 15: Spatial distribution of TSS concentration from (a) *in situ* measurements and (b) Landsat OLI and (c) Sentinel-2A.

and Landsat OLI datasets is presented in Figure 14. It is observed that the chlorophyll distribution trend from the *in situ* measurements and Landsat OLI (Figures 14(a) and 14(b), respectively) is closely correlated with a good coincidence between the high and low chlorophyll-*a* concentration regions. The highest concentrations are observed around the deepest region of the dam at T9, with Landsat-8 OLI slightly underestimating the chlorophyll-*a* concentration (99.43-110.20 $\mu\text{g/L}$), as compared to the laboratory-measured concentrations ranging from 102.00 to 114.00 $\mu\text{g/L}$. On the contrary, for high Chl-*a* concentration regions, Sentinel-2A retrieval does not coincide with that measured and estimated using Landsat OLI as evident in Figure 14(c) with gross overestimation and shift in the spatial location. This difference provides satisfactory evidence as shown in Table 11 and Figure 13, whereby at T2 and T12-T14, the estimation errors of Chl-*a* from Sentinel-2A are marginally higher.

The predominately green band-based Chl-*a* algorithms derived in this study are considered to be suitable for inland

water body systems whereby Chl-*a* is the dominant absorber. For optically complex waters like rivers and polluted lakes, combinations of other water components may result in false Chl-*a* derivations. This can be the case in terrestrially influenced water bodies, where CDOM and nonalgal particles are able to absorb shorter wavelengths in the visible spectrum, thus changing the green reflectance value which may lead to the overestimation of Chl-*a* [46]. For the establishment of the differences between water body systems, the uncertainties and absolute accuracy of the developed algorithms should be considered for Chl-*a*, TSS, and turbidity.

4.6.2. *Distribution of TSS Concentration.* The results in Figure 15 show that the distribution of TSS in the dam was fairly similar between the *in situ* laboratory measurements and the estimations by the Landsat OLI and Sentinel-2A sensors. In the case of TSS and as observed for Chl-*a*, Sentinel-2A tended to slightly overestimate the TSS concentration; however, the peak locations of high and low concentrations

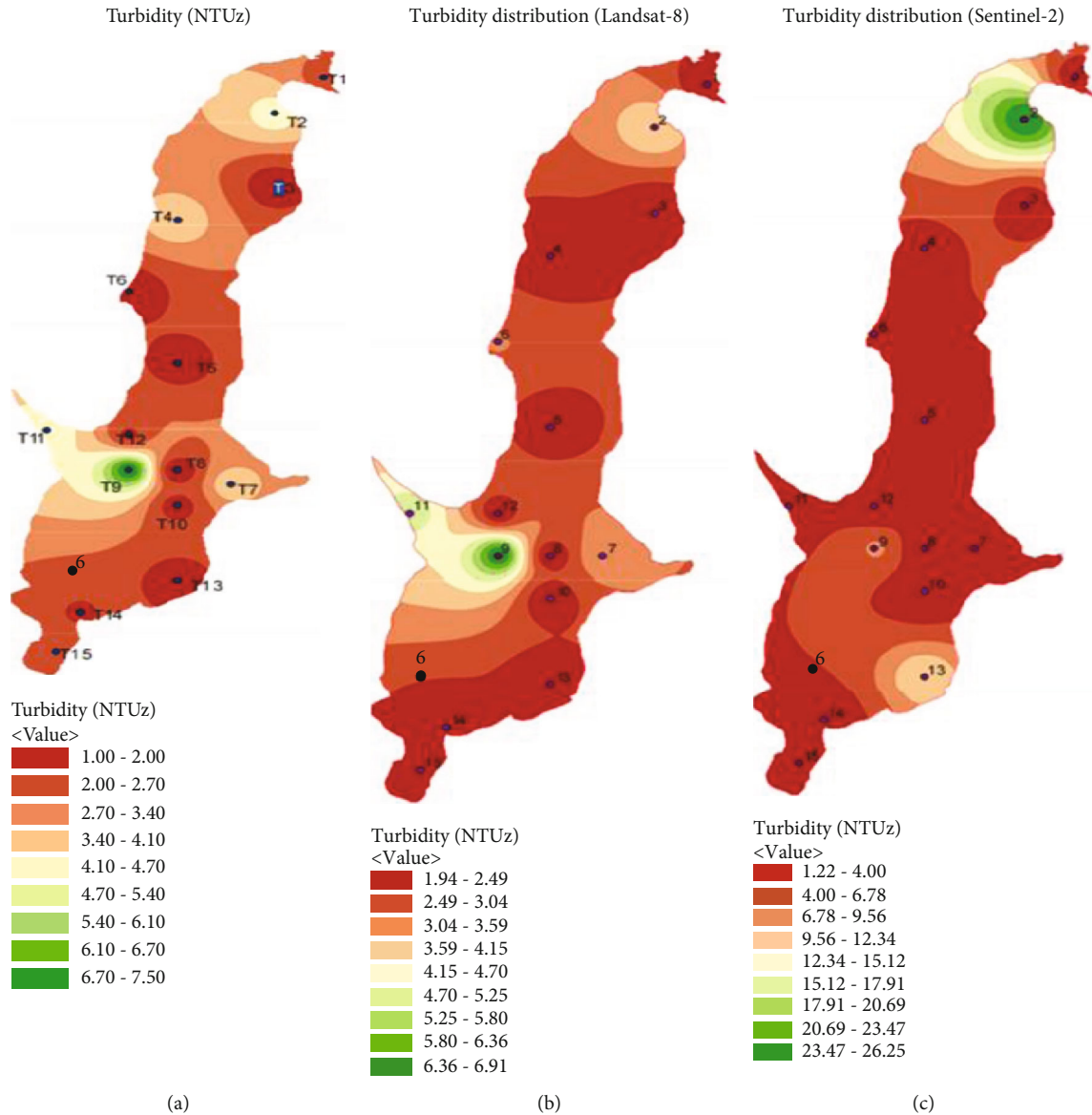


FIGURE 16: Spatial distribution of turbidity concentration from (a) *in situ* measurements and (b) Landsat OLI and (c) Sentinel-2A.

of TSS are in fair coincidence between the satellite-based model estimates and the measured values.

Compared to the statistical analysis results in Section 4.5, for validation, it is observed that the spatial interpolation results depict a better TSS concentration distribution than the discrete analysis in Table 9 and Figure 13. Thus, despite the low statistical accuracy of indications, the two sensors are reliable for the prediction of the general distribution of the TSS within the inland water body.

4.6.3. Distribution of Turbidity Concentration. The distribution of turbidity in Figure 16 is observed to be in coincidence with the trend in Chl-*a* (Figure 14), with Sentinel-2A (Figure 16(c)) overestimating the turbidity concentration, while Landsat OLI estimations (Figure 16(b)) are closely correlated with spatial location and areal distribution with the laboratory-measured turbidity (Figure 16(a)). Notably, though the statistical analysis shows that both sensors

detected the turbidity with accuracy of more than 70%, the spatial distribution maps do not however exhibit the same high accuracy, especially because the Sentinel-2A tended to grossly overestimate turbidity (Figure 16(c)).

The algorithmic models in the current study could be applied to map Chl-*a*, TSS, and turbidity in the Chebara Dam system (Figures 14–16) or in other reservoirs within the same geographic locations. However, the specific coefficients of the algorithm equations should be revised to match the biooptical characteristics of the WQPs present in the target reservoir waters. For replication, the algorithms should be recalibrated to suit the different climatic conditions and seasons and the unique hydrologic characteristics for specific reservoirs. Our future focus is on improving the models and on collecting *in situ* data across a larger extent of the reservoir at multiple seasons for spatial-temporal mapping of WQPs and developing model transfer function models for applications in dams within the same region.

5. Conclusions

This study compared Landsat-8 OLI and Sentinel-2A MSI satellite sensors for the retrieval of chlorophyll-*a*, TSS, and turbidity in an inland water reservoir. For clear water surfaces, Landsat-8 exhibited a higher signal-to-noise ratio (SNR) in the visible bands and marginally higher water surface reflectance in the visible and NIR bands as compared to the corresponding Sentinel-2A bands. Using the visible and NIR bands with the empirical multivariate regression model, the presence of Chl-*a* was retrieved from Sentinel-2A with accuracy of 62-70%, which was 15-25% less than the Chl-*a* estimations from Landsat-8. Using the same multi-spectral bands, TSS was estimated from Sentinel-2A with accuracy of between 47 and 62% as compared to Landsat-8's accuracy of more than 90%. For turbidity estimations, the performance of both sensors was nearly equal with accuracy of more than 75%. The results show the potentials of using both sensors in reservoir water quality retrieval, with Landsat-8 performing better in estimating the tested parameters. To improve the accuracy of water quality estimations in inland water bodies, the study recommends the radiometric calibration and synergistic use of Landsat-8/9 and Sentinel-2A/2B for frequent water quality assessments.

Data Availability

The Landsat data was obtained from the USGS through the EarthExplorer (<https://earthexplorer.usgs.gov/>), and Sentinel-2 was downloaded from the ESA Sentinel Online (<https://sentinel.esa.int/web/sentinel/sentinel-data-access>). The rest of the data used to support the findings of this study are included within the article.

Disclosure

The authors carried out the research while at the Department of Civil and Structural Engineering, Moi University, Kenya.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

- [1] N. Khatri and S. Tyagi, "Influences of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas," *Frontiers in Life Science*, vol. 8, no. 1, pp. 23–39, 2014.
- [2] S. N. Kulshreshtha, "A global outlook for water resources to the year 2025," *Water Resources Management*, vol. 12, no. 3, pp. 167–184, 1998.
- [3] Y. O. Ouma, J. Waga, M. Okech, O. Lavisa, and D. Mbutia, "Estimation of Reservoir Bio-Optical Water Quality Parameters Using Smartphone Sensor Apps and Landsat ETM+: Review and Comparative Experimental Results," *Journal of Sensors*, vol. 2018, Article ID 3490757, 32 pages, 2018.
- [4] J. Boucher, K. C. Weathers, H. Norouzi, and B. Steele, "Assessing the effectiveness of Landsat 8 chlorophyll-*a* retrieval algorithms for regional freshwater monitoring," *Ecological Applications*, vol. 28, no. 4, pp. 1044–1054, 2018.
- [5] V. S. Martins, C. C. F. Barbosa, L. A. S. de Carvalho, D. S. F. Jorge, F. L. Lobo, and E. M. L. M. Novo, "Assessment of atmospheric correction methods for Sentinel-2 MSI images applied to Amazon floodplain lakes," *Remote Sensing*, vol. 9, no. 4, p. 322, 2017.
- [6] T. Kutser, B. Paavel, C. Verpoorter et al., "Remote sensing of black lakes and Using 810 nm reflectance peak for retrieving water quality parameters of optically complex waters," *Remote Sensing*, vol. 8, no. 6, p. 497, 2016.
- [7] Y. O. Ouma, "Advancements in medium and high resolution Earth observation for land-surface imaging: evolutions, future trends and contributions to sustainable development," *Advances in Space Research*, vol. 57, no. 1, pp. 110–126, 2016.
- [8] M. Gholizadeh, A. Melesse, and L. Reddi, "A comprehensive review on water quality parameters estimation using remote sensing techniques," *Sensors*, vol. 16, no. 8, pp. 1298–1306, 2016.
- [9] Y. O. Ouma, C. O. Okuku, and E. N. Njau, "Use of artificial neural networks and multiple linear regression model for the prediction of dissolved oxygen in rivers: case study of hydrographic basin of River Nyando, Kenya," *Complexity*, vol. 2020, Article ID 9570789, 23 pages, 2020.
- [10] W. Zhu, Q. Yu, Y. Q. Tian, B. L. Becker, T. Zheng, and H. J. Carrick, "An assessment of remote sensing algorithms for colored dissolved organic matter in complex freshwater environments," *Remote Sensing of Environment*, vol. 140, pp. 766–778, 2014.
- [11] F. Watanabe, E. Alcântara, T. Rodrigues, N. Imai, C. Barbosa, and L. Rotta, "Estimation of chlorophyll-*a* concentration and the trophic state of the Barra Bonita hydroelectric reservoir using OLI/Landsat-8 images," *International Journal of Environmental Research and Public Health*, vol. 12, no. 9, pp. 10391–10417, 2015.
- [12] M. Potes, G. Rodrigues, A. M. Penha et al., "Use of Sentinel 2 – MSI for water quality monitoring at Alqueva reservoir, Portugal," *Proceedings of the International Associations of Hydrological Sciences (IAHS)*, vol. 380, pp. 73–79, 2018.
- [13] K. Toming, T. Kutser, A. Laas, M. Sepp, B. Paavel, and T. Noges, "First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery," *Remote Sensing*, vol. 8, no. 8, p. 640, 2016.
- [14] W. He, S. Chen, X. Liu, and J. Chen, "Water quality monitoring in a slightly-polluted inland water body through remote sensing - case study of the Guanting Reservoir in Beijing, China," *Frontiers of Environmental Science & Engineering in China*, vol. 2, no. 2, pp. 163–171, 2008.
- [15] M. Potes, M. J. Costa, J. C. B. da Silva, A. M. Silva, and M. Morais, "Remote sensing of water quality parameters over Alqueva reservoir in the south of Portugal," *International Journal of Remote Sensing*, vol. 32, no. 12, pp. 3373–3388, 2011.
- [16] M. Potes, M. J. Costa, and R. Salgado, "Satellite remote sensing of water turbidity in Alqueva reservoir and implications on lake modelling," *Hydrology and Earth System Sciences*, vol. 16, no. 6, pp. 1623–1633, 2012.
- [17] IOCCG, *Remote sensing of ocean colour in coastal, and other optically-complex, waters*, S. Sathyendrath, Ed., Reports of the International Ocean-Colour Coordinating Group, No. 3, IOCCG, Dartmouth, Canada, 2000.
- [18] J. Li and D. Roy, "A global analysis of Sentinel-2A, Sentinel-2B and Landsat-8 data revisit intervals and implications for terrestrial monitoring," *Remote Sensing*, vol. 9, no. 9, p. 902, 2017.

- [19] A. N. Tyler, P. D. Hunter, E. Spyarakos, S. Groom, A. M. Constantinescu, and J. Kitchen, "Developments in earth observation for the assessment and monitoring of inland, transitional, coastal and shelf-sea waters," *Science of the Total Environment*, vol. 572, pp. 1307–1321, 2016.
- [20] P. Gernez, D. Doxaran, and L. Barillé, "Shellfish aquaculture from space: potential of Sentinel2 to monitor tide-driven changes in turbidity, chlorophyll concentration and oyster physiological response at the scale of an oyster farm," *Frontiers in Marine Science*, vol. 4, 2017.
- [21] H. Liu, Q. Li, T. Shi, S. Hu, G. Wu, and Q. Zhou, "Application of Sentinel 2 MSI images to retrieve suspended particulate matter concentrations in Poyang Lake," *Remote Sensing*, vol. 9, no. 7, p. 761, 2017.
- [22] K. Dörnhöfer, A. Göritz, P. Gege, B. Pflug, and N. Oppelt, "Water constituents and water depth retrieval from Sentinel-2A - a first evaluation in an oligotrophic lake," *Remote Sensing*, vol. 8, no. 11, p. 941, 2016.
- [23] U. N. T. Nguyen, L. T. H. Pham, and T. D. Dang, "An automatic water detection approach using Landsat 8 OLI and Google Earth Engine cloud computing to map lakes and reservoirs in New Zealand," *Environmental Monitoring and Assessment*, vol. 191, no. 4, 2019.
- [24] E. F. Vermote, C. Justice, M. Claverie, and B. Franch, "Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product," *Remote Sensing of Environment*, vol. 185, no. Iss 2, pp. 46–56, 2016.
- [25] H. K. Zhang, D. P. Roy, L. Yan et al., "Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences," *Remote Sensing of Environment*, vol. 215, pp. 482–494, 2018.
- [26] E. F. Vermote, D. Tanré, J. L. Deuze, M. Herman, and J. J. Morcette, "Second simulation of the satellite signal in the solar spectrum, 6S: an overview," *IEEE Geoscience and Remote Sensing Letters*, vol. 35, no. 3, pp. 675–686, 1997.
- [27] J. G. Masek, E. F. Vermote, N. E. Saleous et al., "A Landsat surface reflectance dataset for North America, 1990–2000," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 68–72, 2006.
- [28] D. P. Roy, J. Li, H. K. Zhang, L. Yan, H. Huang, and Z. Li, "Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance," *Remote Sensing of Environment*, vol. 199, pp. 25–38, 2017.
- [29] F. Gascon, C. Bouzinac, O. Thépaut et al., "Copernicus Sentinel-2A calibration and products validation status," *Remote Sensing*, vol. 9, no. 6, p. 584, 2017.
- [30] J. A. Barsi, B. Alhammoud, J. Czaplá-Myers et al., "Sentinel-2A MSI and Landsat-8 OLI radiometric cross comparison over desert sites," *European Journal of Remote Sensing*, vol. 51, no. 1, pp. 822–837, 2018.
- [31] R. Chastain, I. Housman, J. Goldstein, M. Finco, and K. Tenneson, "Empirical cross sensor comparison of Sentinel-2A and 2B MSI, Landsat-8 OLI, and Landsat-7 ETM+ top of atmosphere spectral characteristics over the conterminous United States," *Remote Sensing of Environment*, vol. 221, pp. 274–285, 2019.
- [32] R. Shang and Z. Zhu, "Harmonizing Landsat 8 and Sentinel-2: a time-series-based reflectance adjustment approach," *Remote Sensing of Environment*, vol. 235, p. 111439, 2019.
- [33] M. Drusch, U. del Bello, S. Carlier et al., "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sensing of Environment*, vol. 120, pp. 25–36, 2012.
- [34] J. R. Schott, *Remote Sensing the Image Chain Approach*, Oxford University Press, New York, 2nd Ed. edition, 2007.
- [35] B. A. Franz, S. W. Bailey, N. Kuring, and P. J. Werdell, "Ocean color measurements with the Operational Land Imager on Landsat-8: implementation and evaluation in SeaDAS," *Journal of Applied Remote Sensing*, vol. 9, no. 1, article 096070, 2015.
- [36] S. Yezpez, A. Laraque, J. M. Martinez et al., "Retrieval of suspended sediment concentrations using Landsat-8 OLI satellite images in the Orinoco River (Venezuela)," *Comptes Rendus Geoscience*, vol. 350, no. 1-2, pp. 20–30, 2018.
- [37] F. Mushtaq and M. G. Nee Lala, "Remote estimation of water quality parameters of Himalayan Lake (Kashmir) using Landsat 8 OLI imagery," *Geocarto International*, vol. 32, no. 3, pp. 274–285, 2016.
- [38] M. Bonansea, M. Ledesma, C. Rodriguez, and L. Pinotti, "Using new remote sensing satellites for assessing water quality in a reservoir," *Hydrological Sciences Journal*, vol. 64, no. 1, pp. 34–44, 2019.
- [39] F. Watanabe, E. Alcántara, T. Rodrigues, L. Rotta, N. Bernardo, and N. Imai, "Remote sensing of the chlorophyll-a based on OLI/Landsat-8 and MSI/sentinel-2A (Barra Bonita reservoir, Brazil)," *Annals of the Academy of Brazilian Academy of Sciences*, vol. 90, 2 suppl 1, pp. 1987–2000, 2018.
- [40] V. Markogianni, D. Kalivas, G. Petropoulos, and E. Dimitriou, "Analysis on the feasibility of Landsat 8 imagery for water quality parameter assessment in an oligotrophic Mediterranean Lake," *International Journal of Geological and Environmental Engineering*, vol. 11, no. 9, pp. 906–914, 2017.
- [41] E. Kontopoulou, P. Kolokoussis, and K. Karantzalos, "Water quality estimation in Greek lakes from Landsat 8 multispectral satellite data," *European Water*, vol. 58, pp. 191–196, 2017.
- [42] L. C. González-Márquez, F. M. Torres-Bejarano, C. Rodríguez-Cuevas, A. C. Torregróza-Espinosa, and J. A. Sandoval-Romero, "Estimation of water quality parameters using Landsat 8 images: application to Playa Colorada Bay, Sinaloa, Mexico," *Applied Geomatics*, vol. 10, no. 2, pp. 147–158, 2018.
- [43] N. Lailia, F. Arafaha, L. Jaelania et al., "Development of water quality parameter retrieval algorithms for estimating total suspended solids and chlorophyll using Landsat 8 for Poteran Island waters," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-2/W2, 2015 Joint International Geoinformation Conference 2015, pp. 28–30, Kuala Lumpur, Malaysia, October 2015.
- [44] N. H. Quang, J. Sasaki, H. Higa, and N. H. Huan, "Spatiotemporal variation of turbidity based on Landsat 8 OLI in Cam Ranh Bay and Thuy Trieu Lagoon, Vietnam," *Water*, vol. 9, no. 8, p. 570, 2017.
- [45] L.-W. Liu and Y.-M. Wang, "Modelling reservoir turbidity using Landsat 8 satellite imagery by gene expression programming," *Water*, vol. 11, no. 7, p. 1479, 2019.
- [46] C. Roesler, J. Uitz, H. Claustre et al., "Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: a global analysis of WET labs ECO sensors," *Limnology and Oceanography: Methods*, vol. 15, no. 6, pp. 572–585, 2017.

Research Article

Integrating Image and Network-Based Topological Data through Spatial Data Fusion for Indoor Location-Based Services

Dasol Ahn ¹, Alexis Richard C. Claridades ^{1,2} and Jiyeong Lee ¹

¹Department of Geoinformatics, University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, Seoul 02504, Republic of Korea

²Department of Geodetic Engineering, University of the Philippines Diliman, Quezon City 1101, Philippines

Correspondence should be addressed to Jiyeong Lee; jlee@uos.ac.kr

Received 23 April 2020; Accepted 14 July 2020; Published 19 September 2020

Academic Editor: Sang-Hoon Hong

Copyright © 2020 Dasol Ahn et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nowadays, the importance and utilization of spatial information are recognized. Particularly in urban areas, the demand for indoor spatial information draws attention and most commonly requires high-precision 3D data. However accurate, most methodologies present problems in construction cost and ease of updating. Images are accessible and are useful to express indoor space, but pixel data cannot be applied directly to provide indoor services. A network-based topological data gives information about the spatial relationships of the spaces depicted by the image, as well as enables recognition of these spaces and the objects contained within. In this paper, we present a data fusion methodology between image data and a network-based topological data, without the need for data conversion, use of a reference data, or a separate data model. Using the concept of a Spatial Extended Point (SEP), we implement this methodology to establish a correspondence between omnidirectional images and IndoorGML data to provide an indoor spatial service. The proposed algorithm used position information identified by a user in the image to define a 3D region to be used to distinguish correspondence with the IndoorGML and indoor POI data. We experiment with a corridor-type indoor space and construct an indoor navigation platform.

1. Introduction

Most services involving spatial data are available for outdoors, compared to the indoors [1], despite being more crucial in urban areas, where people generally spend more time in a structure [2] or where navigation in cases of evacuation is experiencing more delay [3]. As this concern gains attention, various approaches in representation have been attempted across applications, particularly as this depends on the field and intent of service [4]. Mobile devices have also been increasingly popular and integrated with the daily lives of humans [5].

Multiple studies have shown efforts to represent the indoor environment, such as Light Detection and Ranging (LiDAR) or through Building Information Modeling (BIM) data [6], which require either cumbersome data collection or expensive equipment. The dynamic characteristic of indoor spaces, especially in urban areas, needs constant updating of datasets, and these methodologies pose problems in cost and time. Omnidirectional images taken with a 360°

point of view [7] present an alternative, as this provides cheaper and faster avenues in depicting indoor space. However, these images present difficulty when used in providing services more than visualization, such as identifying spaces or objects within indoor space, which are crucial in applications such as navigation and facility management because they only contain pixel data. For example, we cannot distinguish directly if a room is directly accessible from the hallway, or if it is adjacent to another room from the images alone. In the same way, we can visually see facilities such as fire extinguishers or CCTV cameras. Still, their exact locations are unidentifiable due to a lack of geometric information.

Indoor network-based topological data may provide the information lacking in these omnidirectional images, such as spatial relationships of connectivity between spaces or containment of objects within the spaces. IndoorGML is the standard established by the Open Geospatial Consortium (OGC) for indoor spatial data, geared primarily for representing space for navigation applications [8]. Using

topological data, such as IndoorGML, enables services utilizing query-based analysis on omnidirectional images that alone give only visualization. With this, there is a need to link omnidirectional images that display the spaces and the indoor topological data, which represents the relationships between said spaces, as well as objects contained within.

Different types of data represent various aspects of indoor space, and this variety may be a source of problems in utilization because of compatibility issues [9]. Data fusion is for datasets coming from various sources or formats to produce the same quality of output information or increase understanding of the underlying phenomena represented differently by these datasets to address this problem of variety. In this study, we aim to propose a methodology for the data fusion of image and a network-based topological data without undertaking data conversion, using a separate data model, or reference data. We demonstrate a procedure to establish a relationship between omnidirectional images and IndoorGML data for providing indoor space applications.

This paper is structured as follows. The next section discusses studies on efforts on indoor space expression and methodologies in data fusion. The third section presents the proposed methods for image and topology data fusion. By understanding the relationship between the omnidirectional images and IndoorGML data, we lay out the data requirements that must be satisfied before establishing the fusion. In the following section, we conduct an experimental implementation of the proposed methodology by developing an indoor space service based on navigation, as well as indoor POI display. Finally, the last section concludes with implications and limitations derived from this study and future studies.

2. Related Research

In this section, we review how topological data and data fusion play a role in indoor space representation. We explore various studies regarding methods of expressing indoor space, as well as efforts for data fusion methods to produce indoor spatial data, in the aim to provide various indoor services.

Typical means of expressing indoor space include the use of two-dimensional (2D) or three-dimensional (3D) data. 2D methods include the least amount of information that represents the indoor space [10], such as CAD floor plan drawings. However, if indoor space is in 3D, it is possible to analyze indoor space differently compared when it is in 2D [11] as specific characteristics may seem more apparent. These methods may use existing 3D CAD models or high-precision LiDAR measurement that can accurately represent indoor space but can entail high cost and large file size. These methods may be practical in military applications, in gaming, or in other cases where service is feasible only when indoor space is represented very accurately [4]. On the other hand, it is possible to support various functions such as attribute search and viewing if the data supports the expression of topological relationships of indoor spaces [12]. In the context of indoor navigation and LBS, network-based topological

models are emphasized as necessary [13], especially in the context of the visualization and analysis of the internal composition of as-built structures [14]. Reference [15] has also demonstrated that these types of topological data are more efficient in performing spatial queries.

A typical method of expressing topological relationships among spaces is the Combinatorial Data Model (CDM) based on the Node-Relation Structure (NRS) [16]. The NRS transforms a 3D object into a node, and an edge represents the respective shared boundaries among the rooms, based on the Poincare duality. Hence, in the topological model, the nodes represent the indoor spaces, and the edges represent the topological relationships among connected nodes. Based on the CDM, the Open Geospatial Consortium (OGC), an international standards organization for spatial information, established IndoorGML as a framework to represent topological relationships for indoor spaces and data exchange. Similarly, indoor spaces are defined as nodes (also called states), while edges (or referred to as transitions) express topological relationships [8].

As IndoorGML is capable of representing the indoor spaces, its primary utilization is for the investigation of the usage of indoor space, such as indoor LBS, or indoor route analysis. Architectural components, fixtures, or objects found within the spaces are beyond the scope of this standard. However, to provide indoor spatial services successfully, these objects contained within these spaces, the targets of indoor navigation, must also be represented. Indoor points of interest (indoor POI) expressed geometrically as points represent positions of objects and are used to link their respective attribute information. As this IndoorGML does not directly have specifications about these objects, Jung and Lee [4] utilized the multilayered space model (MLSM) to simultaneously represent indoor topological information through the IndoorGML NRS and the indoor features via indoor POI in an indoor patrol service application. Similarly, Claridades et al. discussed this concept in integrating IndoorGML and indoor POI. In both studies, layers that make up the MLSM divide the space into nonoverlapping layers, and those nodes exist in each layer independently. In turn, interlayer relationships define the relationships among the layers. This definition emphasized an implementation-oriented expression of topological relationships between a node representing a space (for example, a room) and nodes representing objects within said spaces, illustrated in Figure 1 [17].

In many cases, multiple datasets may exist to represent the same geographic features existing in the real world [18]. These may represent various aspects of the feature they represent, or each of them is a representation using different data models that are accordingly suited to each application. This ambiguity may pose problems in data compatibility, in duplication, and in the process of integration. Data fusion is defined as a combination of two or more data sources to provide less expensive, more relevant, or higher quality of information [19]. In this manner, performing data fusion enables overcoming this predicament by linking data from separate sources, collected through different methods, or observing various standards. This approach may also resolve

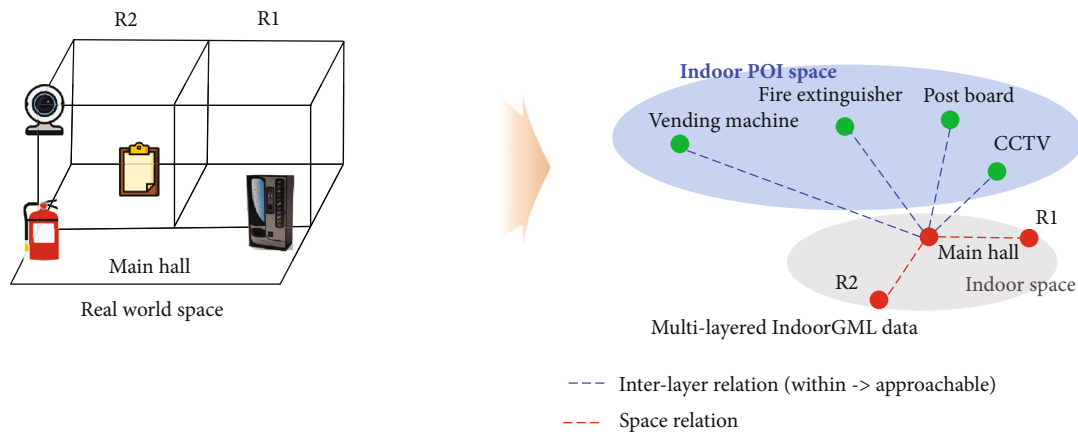


FIGURE 1: Conceptualization of the relationships of spaces and objects.

ambiguities caused by selecting an appropriate data model for an application, since one may be more specific over another in implementing a certain task [20].

Data fusion is especially helpful in GIS because even though geographic datasets are readily accessible across applications, it assists in the combination of features, each having their suitable aspects, to empower geospatial analysis [21, 22]. In the case of spatial datasets accessible through the web through Spatial Data Infrastructures (SDI) of both private and government organizations, data fusion is possible through the concept of linked data, using unique identifiers and standardized web formats to resolve conflicts in data [23, 24]. This technique is aimed at assisting the generation and updating of spatial data [25], as well as the construction of location-aware systems to provide services to record human movement and deliver visual feedback [26].

Stankute and Asche defined their approach by extracting the best-fit geometric data, and the most suitable semantic data from datasets through coordinate matching [21], while other approaches used attribute mapping to achieve feature correspondence [27]. For earth observation data, [28] has applied data fusion to combine multisource data using linguistic quantifiers for environmental status assessment.

As IndoorGML primarily presents topological information, topology-based data fusion approaches are most suitable. This approach is aimed primarily at enriching the dataset, since it contains only the minimum requirements to enable basic indoor spatial modeling. In one of its main application aims of indoor routing, IndoorGML's proponents have suggested extensions for common application domains to be applied to increase utilization [29]. To augment both standards' limitations in representing indoor space [30], an Indoor Spatial Data Model (ISDM) referring to CityGML for the feature model and IndoorGML for the topology model was proposed, by defining additional feature classes. On the other hand, Topological Relation Model (TRM) [20] essentially establishes connections between data through matching geometric data generated from respective models. Since not all datasets may be a source of geometric information, the Topological Relation-based Data Fusion Model approached the problem by generating topological data from surface, network, and volume-based data to estab-

lish matching [18]. Commonly, these approaches determine a correspondence between the features in the data as an approach to data fusion. However, a match among features is not possible in all datasets, such as in the case of images.

Several methodologies have utilized omnidirectional images to create a recreation of a room layout using omnidirectional images, such as RoomNet [31] and LayoutNet [32], which extract predicted room layouts from a single subunit of an interior, so the topological relationships between said subunit and other subunits are indistinguishable. Jung and Lee [4] examined the case for utilizing omnidirectional images and topological information in an indoor patrol application. Online web mapping services such as Google Street View [33] and Kakao Storeview [34] use these images to present a snapshot of the indoor space at the moment of image capture. Also, the method of collection along a corridor-type space was described by determining shooting points or locations where these images may be collected efficiently [35]. In establishing the indoor patrol application, the IndoorGML CellSpace class has an association relationship with the omnidirectional image denoting that each image represents a space situated in each shooting point, which in turn is a node in the IndoorGML NRG. The definition of IndoorGML relationships was extended by defining a within relationship for the objects contained in the spaces. Here, however, the connection between the image and topology data was implemented through the use of a reference data. The algorithm performs a spatial query on polygons, and the coordinate was calculated from a pixel location in the image to identify the containing space and attribute information. Besides, since pixels present in the images only present objects visually and not discretely as in vector data, using exact positions to define topological relationships in this manner may be difficult [4]. Semantically separating an image into different objects has been regarded as a chicken and egg problem—an object's type and shape are essential to determine if a pixel belongs to the object or not, but this object must first be isolated to understand which objects are present in an image [36]. Furthermore, using connectivity relationships, images are loaded discretely in implementing space to space navigation. In effect, this gives the impression of discontinuity in indoor space, especially in the corridors.

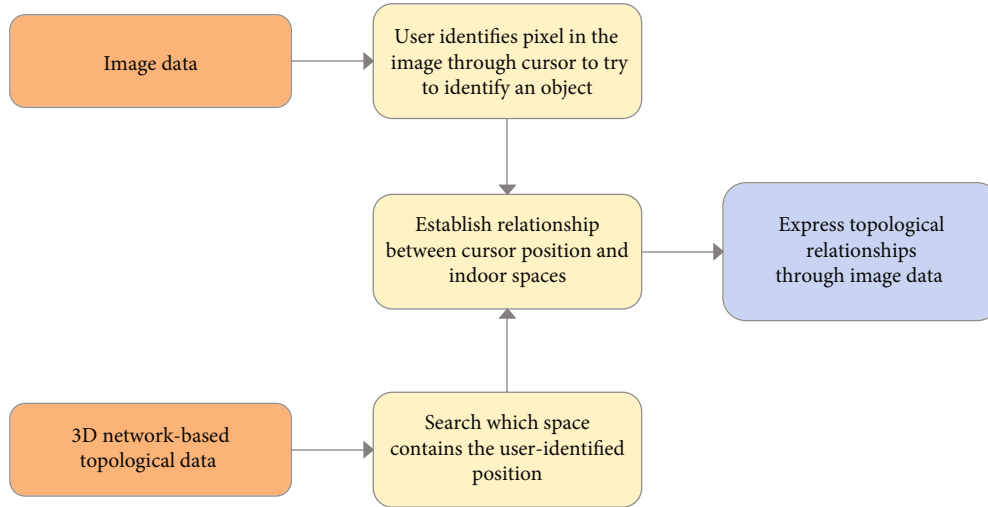


FIGURE 2: Framework of image and topological data fusion.

The concept of spatially extended topology, based on the 9-intersection model [37], is intended to define topological properties of moving objects to provide a concierge service application [38] through defining regions of around points reflecting respective ranges of influence. This concept describes that for an object in a location, a scope of influence of a certain range describes a conceptual area of potential interaction with the said object. It is conceptual, not a physical region, such as the transmission footprint of an antenna or a broadcast range of a WiFi router. This possibility of defining a range of distance provides opportunities for various spatial reasonings about an object.

3. Methodology

In this section, we describe the framework to perform image and topological data fusion. Then, we describe the necessary algorithms to perform a match between locations identified in the image and the nodes that represent the spaces in the topological data. These methods are necessary to perform functionalities to demonstrate how the link of image data and topological data enables spatial analysis in the indoor space.

3.1. Framework for Image and Topology Data Fusion. We utilize the concept of data fusion on spatial data representing the same indoor space, to produce more information than when they are separate. By directly combining information from topological data and image data, this methodology is aimed at providing more relevant information in indoor space, which is especially useful in the context of indoor navigation and visualization.

Because of the high visual content, ease of updating, and relatively small file size, image data is a suitable representation of the indoor environment. However, to provide LBS in indoor space, it must be supplemented by a network-based topological data, in which spaces and their respective relationships are represented directly through nodes and edges, respectively. Images only provide information through

individual pixels, which do not contain enough spatial information that may be further helpful to the LBS. In this data fusion approach, we aim to establish a relationship between the image pixel data and the nodes of the topological dataset. Within the LBS, a user can visualize a space (or object) in the image through the pixels. This pixel's position is the key to perform a query to search for the corresponding node in the topological data. This node represents either the space that contains the position represented by the pixel or an object contained in that space. We illustrate the framework for this approach in Figure 2.

3.2. Establishing the Relationship between the Image and Topological Data. With the method of representing indoor space and objects in these spaces, it is difficult to establish a direct matching to the topology data. Images contain only pixels and do not have geometry, making it difficult to obtain attributes of features or identify the features themselves intuitively. With 3D coordinates, it is possible to establish a 1:1 match towards the nodes of the topological data. Still, a separate method is necessary to recognize the spaces or objects even though they are displayed visually.

Figure 3 illustrates the general method of establishing a relationship between the data. Since a user can visually see an object, the process of space or object identification begins with the user selection of a position in the image. The user only selects a single pixel at this point in the image, and this calculated position defines the relationship with the topological data. We use the image heading (where north is equal to zero), image radius, and horizontal angle of the selected pixel, as well as the vertical angle to calculate X and Y coordinates and image capture height for the Z coordinate. Moving forward, since coordinates of the user-selected pixel are present for the matching, coordinates of the nodes representing either indoor spaces or indoor objects are derived similarly. The function of the node, whether as an indoor space or as an object within the space, is also noted, since this differentiates the topological relationships, say connectivity from space (click) to space (node) or inclusion from space (click) to POI (node).

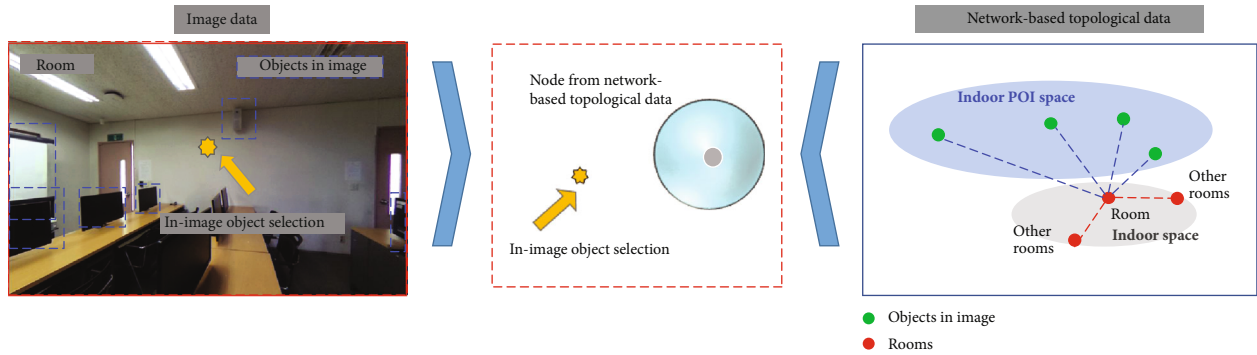
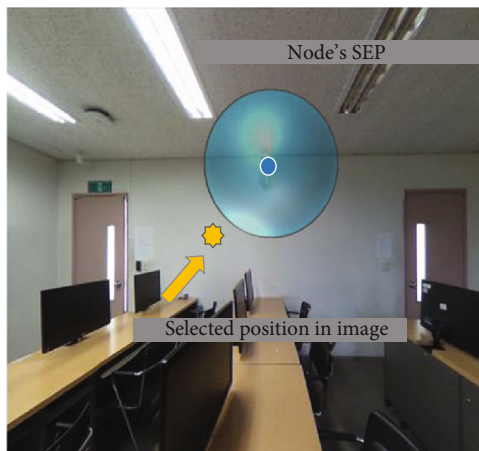


FIGURE 3: Method of linking the nodes and user-identified spaces in image.

TABLE 1: Types of spaces differentiated for allowable ranges.

Topological data		Represented space	Approximate range
Indoor space	Room	Spaces connected with corridors	Average room dimensions
	Hallway	Continuous space along a corridor	Corridor width
POI	Door	Objects that separate between rooms and nonroom spaces	Average door height and width
	Object	Facilities, objects in the room except doors	Average object height and width



Relationship between object (P) and node (R)

- p' : point P
- R° : inside R
- ∂R : boundary of R
- R^- : outside R

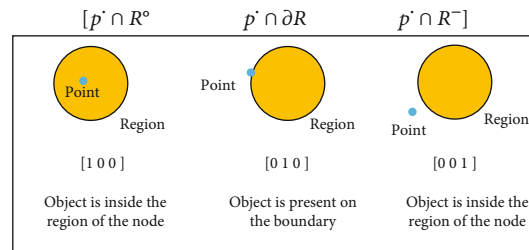


FIGURE 4: Determining topological relationships using the SEP.

In the LBS, the pixel's position defines the link to the topological data by knowing which nodes are in this position's vicinity. With this, a 3D space around the nodes' location can describe vicinities, to investigate which among them contains the pixel. Depending on the type of space or object that the node represents, the size of the region may vary. Table 1 enumerates the differentiation among the objects and space.

The processes described above calculates the position of the user click and the node closest to that click. Their corresponding relationships are determined using the Spatial Extended Point (SEP) approach based on the calculated coordinates, adopted from Lee [38]. In this study, we define the SEP as a region around each of the nodes to represent the area of potential influence or behaviors around the node's point location. Using the SEP, we determine if the user-

selected point is on the exterior, interior, or boundary of this region, easing on the limitation of the lack of geometric information in both image and topology datasets. It follows from this definition that the SEP's interior defines in the region in which the node's influence is present. The SEP's exterior is where there is no influence anymore, and the boundary quantifies the limit of influence and the start of the noninfluencing area. Figure 4 illustrates this procedure.

The interactions between an area and a point described in Figure 4, extended in 3D for a point and region, are representing topological relationships for those entities in 3D space quantitatively as a matrix, referred to as the SEP matrix, as shown in Equation (1). This matrix may take up values of either 1 (satisfied) or 0 (not satisfied) for each element, depending on the conditions. This matrix simplifies the calculation for the topological relationships for two

entities. A generalized expression defines for each observation in each projection of the 3D space in the three Cartesian planes.

$$\begin{bmatrix} P^o \cap R_{xy}^{\circ} & P^o \cap \partial R_{xy} & P^o \cap R_{xy}^{-} \\ P^o \cap R_{yz}^{\circ} & P^o \cap \partial R_{yz} & P^o \cap R_{yz}^{-} \\ P^o \cap R_{xz}^{\circ} & P^o \cap \partial R_{xz} & P^o \cap R_{xz}^{-} \end{bmatrix} \quad (1)$$

where P^o is the selected position in image; R_{xy}° is inside the boundary along the XY plane; R_{yz}° is Inside the boundary along the YZ plane; R_{xz}° is inside the boundary along the XZ plane; ∂R_{xy} lies on the boundary along the XY plane; ∂R_{yz} lies on the boundary along the YZ plane; ∂R_{xz} lies on the boundary along the XZ plane; R_{xy}^{-} is outside the boundary along the XY plane; R_{yz}^{-} is outside the boundary along the YZ plane; and R_{xz}^{-} is outside the boundary along the XZ plane.

The SEP results from calculating the distance between nodes and a user-selected position in the image. These coordinates are calculated along three orthogonal planes in 3D space, to establish values in the matrix given by Equation (1). In this process, there is a smaller threshold value if the node represents an indoor POI compared to when it is representing a space. First, the distance to the user-identified position in the image is calculated along the three orthogonal planes, to determine if there exists an IndoorGML node within the allowable ranges from that point and its type, if this nearby node exists. If the clicked point has a smaller distance, i.e., closer to an IndoorGML node representing a space, the SEP matrix comes from a more significant threshold value. Depending on the type of space of the identified nearby node, the SEP matrix is populated with appropriate values to denote the topological relationship of the user-identified point and the corresponding IndoorGML node.

Moreover, the size of the region defining the SEP is a factor to be considered when linking images describing indoor space and objects to IndoorGML data. Various definitions of this region based on Table 1 defined the allowable ranges when determining the topological relationships using the SEP. This permissible value must be adjusted accordingly, depending on what the IndoorGML node represents. For example, nodes that represent indoor space must have a broader range, compared to nodes that represent objects such as doors. Pseudocode 1 shows the simplified pseudocode.

4. Experimental Implementation

In this section, we demonstrate the proposed methodology for image and topological data fusion using omnidirectional images and IndoorGML, respectively, by building the visualization platform and implementing the algorithms and processes described earlier.

4.1. Datasets for Implementing Image and Topology Data Fusion. In this implementation, omnidirectional images represent and visualize the indoor space and the objects contained in these spaces. Also, IndoorGML is used to describe

topological relationships between spaces. Figure 5 shows the schematic diagram to generate these datasets before data fusion. First, if the omnidirectional images represent indoor space, at the shooting points to be exact, a 360° view of a place is made. This manner of shooting subdivides one continuous space into subspaces [35]. The image headings (or where the image direction is numerically set to zero and referred to) vary from image to image because of inconsistencies in capture. The irregularity in image capture may cause inconsistencies when calculating positions of spaces in the images, so this must be corrected.

In this study, we express the spatial relationships through IndoorGML as with Jung and Lee [4]. The NRS is the basis of expression of topological relationships of adjacency and connectivity among indoor spaces, and the relationships of objects represented as POI are expressed similarly through a within relationship. Since connectivity relationships of spaces allow the use of an image-based topology authoring tool that produces an XML file structure, indexing each image into a scene is possible [4, 35]. We construct the IndoorGML data as an XML database integrated into the image XML file. For each scene in the XML data representing an image, IndoorGML topological relationships such as connected and adjacent spaces or even POI are contained in that particular space.

4.2. Study Area and Experimental Environment. The target area for this study is the 6th floor of the 21st Century Building, University of Seoul. Image capture was carried out in this location using a DLSR camera and Ricoh Theta S, equipped with a rotary rotator. We matched the captured images using PTGui 10.0.12 to generate the omnidirectional images. Table 2 summarizes these tools.

Based on the IndoorGML data, these images were connected using PanoTour Pro 2.5, an image-based topology authoring tool, to establish the connectivity relationships of each image and to build HTML and XML files used in the service. As PanoTour indexes the image data in XML format and links each to a scene, Krpano scene call scripts display the images. Accordingly, the topology data based on IndoorGML was constructed as an XML database and integrated into the image XML files.

Bitnami was used to build the server to handle the image and network-based topology data. We used web-based languages such as HTML, XML, and JavaScript to implement the algorithms discussed in the previous section. Functions used in the pseudocode described in Pseudocode 1 were assisted by Krpano JavaScript functions, particularly in obtaining coordinates from the images using `getCoordinate`. The output is a web browser-operated platform where the user can pan, zoom, and use scroll around a single point defined omnidirectional image capture location, corresponding to an IndoorGML node. The following section discusses the implementation of the algorithms within this platform to demonstrate the data fusion between omnidirectional images and IndoorGML.

4.3. Applying Data Fusion to Omnidirectional Images and IndoorGML Data. The integration of the image and

```

calculateSEP ( $x_i, y_i, z_i, x_j, y_j, z_j, \text{Node\_part}, \text{Tolerance}$ )
Step 0. Initialize constants
     $r \leftarrow$  image radius
     $\text{offset} \leftarrow$  camera height
     $\text{Node\_part} \leftarrow$  role of Node in scene (type of space)
     $\text{Tolerance} \leftarrow$  allowable range of the SEP
Step 1. Check  $\text{click\_mouse}$  if FALSE
    Step 1.1 Read topological data and obtain node parameters
         $\text{Node\_ath}, \text{Node\_atv} \leftarrow$  position of Node in scene
         $\text{Node\_part} \leftarrow$  part of Node in scene
Step 2. Define function to obtain coordinates from user input
    getCoordinate ( $\text{mouse\_ath}, \text{mouse\_atv}, r, \text{offset}$ ) {
        If camera direction is North, set  $\text{Hd} \leftarrow 0$ , increasing clockwise
         $\text{Angle\_H} \leftarrow (\text{Hd} + \text{mouse\_ath}) * \text{PI}/180$ 
         $\text{Angle\_V} \leftarrow \text{mouse\_atv} * \text{PI}/180$ 
         $x_i \leftarrow r * \cos(\text{Angle\_H})$ 
         $y_i \leftarrow r * \sin(\text{Angle\_H})$ 
         $z_i \leftarrow \text{offset} + r * \sin(\text{Angle\_V})$ 
    }
Step 3. Set  $\text{click\_mouse}$  as TRUE
    Step 3.1 Obtain  $\text{mouse\_ath}, \text{mouse\_atv}$  from click
    Step 3.2 Calculate coordinates
        Step 3.3.1 Obtain coordinate of user identified point
             $x_i, y_i, z_i \leftarrow$  getCoordinate ( $\text{mouse\_ath}, \text{mouse\_atv}, r, \text{offset}$ )
        Step 3.3.2 If  $\text{Node\_part}$  is Indoor Space
             $x_j, y_j, z_j \leftarrow$  getCoordinate ( $\text{mouse\_ath}, \text{mouse\_atv}, r, \text{offset}$ )
        Step 3.3.3 If  $\text{Node\_part}$  is Indoor Object
             $x_j, y_j, z_j \leftarrow$  getCoordinate ( $\text{mouse\_ath}, \text{mouse\_atv}, r, \text{offset}$ )
Step 4. Calculate SEP matrix
    Step 4.1 Initialize: blank 3 x 3 SEP Matrix  $\leftarrow 0$ 
    Step 4.2 Set Tolerance depending on type of space or type of object
        Step 4.2.1 Calculate SEP Matrix values along XY plane
            IF XY distance is less than Tolerance,  $\text{SEP\_Matrix}[0][0] \leftarrow 1$ 
            IF XY distance is same as Tolerance,  $\text{SEP\_Matrix}[0][1] \leftarrow 1$ 
            IF XY distance is greater than Tolerance,  $\text{SEP\_Matrix}[0][2] \leftarrow 1$ 
        Step 4.2.2 Calculate SEP Matrix values along YZ plane
            IF YZ distance is less than Tolerance,  $\text{SEP\_Matrix}[1][0] \leftarrow 1$ 
            IF YZ distance is same as Tolerance,  $\text{SEP\_Matrix}[1][1] \leftarrow 1$ 
            IF YZ distance is greater than Tolerance,  $\text{SEP\_Matrix}[1][2] \leftarrow 1$ 
        Step 4.2.3 Calculate SEP Matrix values along XZ plane
            IF XZ distance is less than Tolerance,  $\text{SEP\_Matrix}[2][0] \leftarrow 1$ 
            IF XZ distance is same as Tolerance,  $\text{SEP\_Matrix}[2][1] \leftarrow 1$ 
            IF XZ distance is greater than Tolerance,  $\text{SEP\_Matrix}[2][2] \leftarrow 1$ 

```

PSEUDOCODE 1 Identifying objects or spaces in images using the SEP matrix.

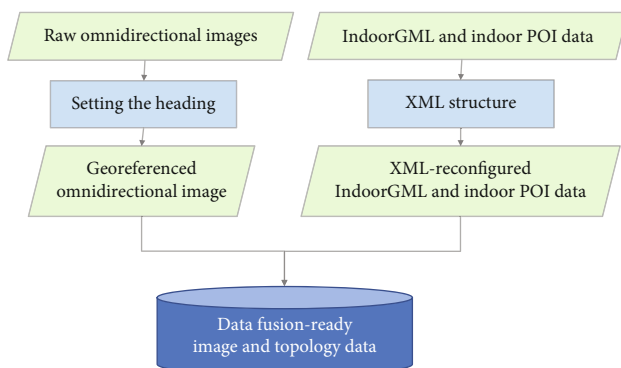


FIGURE 5: Schematic diagram for the method of building omnidirectional image and indoor topological data.

topological data defined using algorithms described in the previous section demonstrates understanding the relationship between the nodes of the IndoorGML data and the omnidirectional images. In this section, we build an interface upon actions utilizing these algorithms through user-initiated actions, such as double-clicking and long-pressing.

First, when a user double-clicks on an omnidirectional image, the previously calculated SEP matrices of the location of the selected pixel is used to determine if an image or an object is present. In either case, a double-click would link the image's or object's information to the current image, and if it is an image, it would eventually load it into display. Figure 6 illustrates this process.

To further demonstrate the relationship between the IndoorGML data and omnidirectional image, a long click

TABLE 2: The experimental environment.

Imaging equipment	35 mm DSLR camera (36.4 million pixels) 180° fisheye lens Rotary rotator Ricoh Theta S (12 million pixels)
Omnidirectional image-matching program	PTGui 10.0.12
Image-based topology authoring program	Kolor PanoTour Pro 2.5
Webserver	Bitnami (Apache)
Server script	XML, HTML, JavaScript
JavaScript plug-in	Krpano

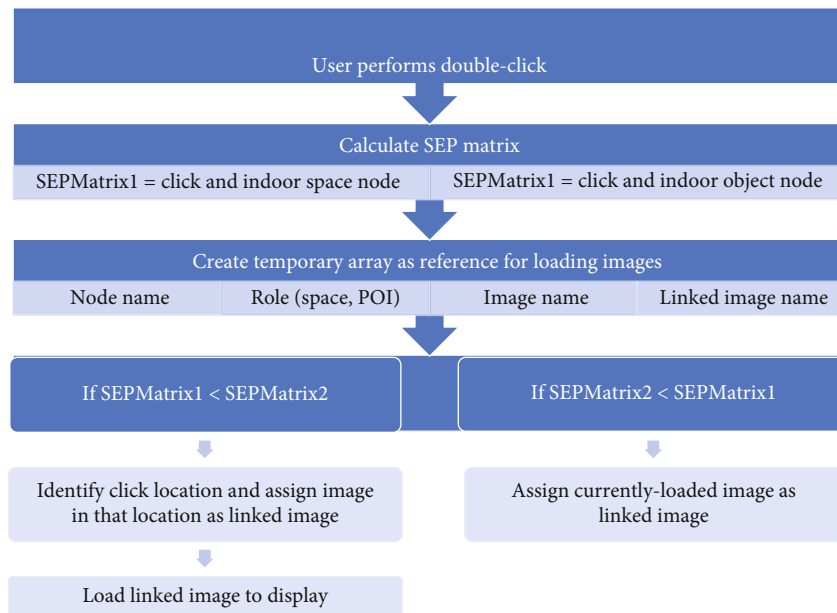


FIGURE 6: Using user double-clicks to identify objects and display image scenes.

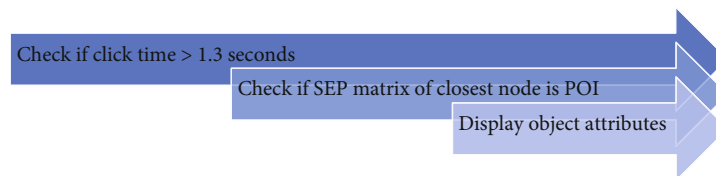


FIGURE 7: Using user long clicks to display object attributes.

on a pixel expresses attributes of objects. In a similar process of identifying objects in the image, the algorithm checks the SEP if it contains doors and objects, and attribute data is only displayed if the object is present in or on the boundary of this SEP. This algorithm presents a method to display information about rooms and facilities, using the properties of the IndoorGML nodes. In other words, the user can see attributes of these items, visible as pixels but not as discrete objects in the images through a long click, which triggers the calculation of the SEP matrix to indicate the topological relationship of the identified position and the positions of each node in the interior space. Figure 7 illustrates this process.

The above operations derive information about the features and spaces using existing attributes present in the IndoorGML data, linked to the omnidirectional images that provide the visual interface to the user. To further illustrate the ability to portray spatial information through the omnidirectional image, we implement an image-based indoor navigation. The user inputs then names of the origin and destination locations, and these names are located within the list of the names of all omnidirectional images. The algorithm identifies the path from the start location towards the entered destination using the IndoorGML data. For a particular image, it draws the path by identifying the linked image

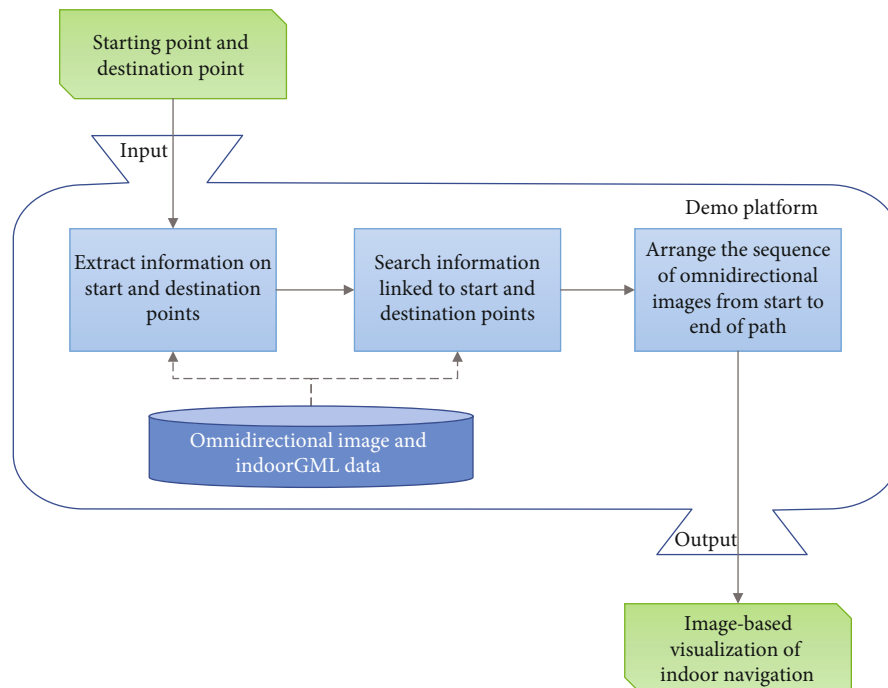


FIGURE 8: Implementation of image-based visualization of indoor navigation.



FIGURE 9: Recognition of space in an omnidirectional image using SEP.

attribute, which represents the successive image for that path, repetitively, until the destination. The sequence of images is loaded with appropriate delays and transitions to visualize a smooth and realistic visualization of navigating through the spaces represented by the images. Figure 8 illustrates this procedure.

4.4. Resulting Platform for Omnidirectional Images and IndoorGML Data Fusion. We implemented the process of identifying objects and spaces in the omnidirectional images using the SEP matrix, which defined the relationships of the

objects in the indoor space and the nodes from the indoor topological data, as shown in Figure 9. When a user double-clicks on the image, the algorithm identifies the node which SEP contains the click position, and the corresponding linked image to that node is displayed. This linked image is present in the location represented by that node. This double-clicking action is an implementation of moving from one position to another, as visualized by successive displaying of omnidirectional images, and demonstrates that topological information, connectivity relationship, for instance, can be obtained directly from the images.

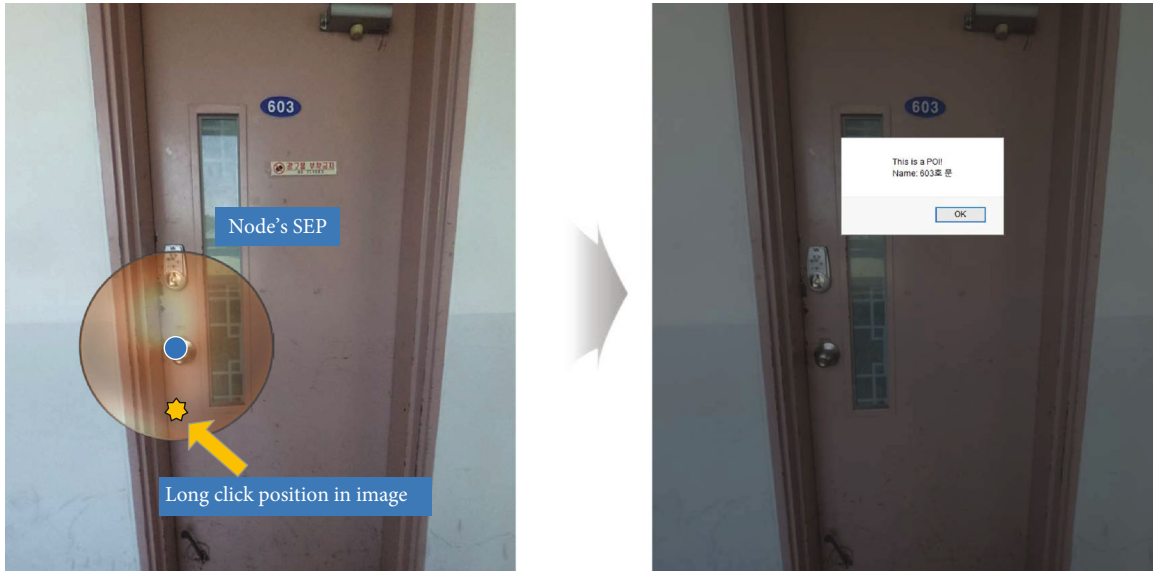


FIGURE 10: Showing object attributes in an omnidirectional image using SEP.



FIGURE 11: Result of image-based visualization of indoor navigation.

Similarly, by using the established relationship of the omnidirectional images and IndoorGML data, users can display attributes of objects using a long click. As shown in Figure 10, the long click identifies the node in which SEP contains the click position, and information about the node is only displayed if that node is representing an object or facility located indoors and not an indoor space. Similar to the previously demonstrated function, this shows that topological information can be obtained directly from the images and, in this case, attributes of the IndoorGML nodes.

In this application, the link between the image and topology data is also used by visualizing the navigation from a starting location to a destination location, even if the starting location is not the currently loaded scene. The user is made to enter names of the desired start and endpoints, and these

names are searched in the image data attributes, after which the path between these points is established using the topology data. Each omnidirectional image along the established path is arranged from start to end, and each one is loaded with appropriate directions of turn, transitions, and delays to achieve a smooth visualization of navigation. The result of this process is illustrated in Figure 11.

Figure 11 illustrates a sample result of the visualization of navigation from one room to another, including the spaces along the path of movement. The user is prompted to input the names of the rooms, in this case from a starting point, Room 605 to the destination Room 607. The visualization of navigation commences with the image displaying that Room 605 is loaded, then rotates towards the direction of the closest door to the destination, loads the next image

displaying the corridor, transitions to the next image near the door of Room 607, then transitions to the image displaying the interior of the destination. The continuous visualization of the path from one location to another demonstrates the continuity of indoor space, despite being represented discretely with subsampled nodes in the topological data and separate omnidirectional images in each image capture location.

5. Conclusions and Future Studies

Indoor space has been expressed in various ways in previous studies, with each method differing in method of collection and generation, emphasized aspect of space, and applications. Geometric datasets such as LiDAR provide realistic and accurate visualizations, but omnidirectional images provide comparable results in this aspect despite it being easier and much cheaper to collect, process, and update. In addition, while it is important to visualize indoor space in three dimensions, studies show that the capability for spatial analysis provided by topological datasets is necessary in providing indoor spatial services such as in the case of LBS—an aspect where omnidirectional images lack. With this, this study proposes a data fusion method between image data and topological data, implemented with omnidirectional images and IndoorGML.

Indoor spaces are expressed using topological data given by IndoorGML, an international standard established by OGC, where they are abstracted through a zero-dimensional node, and the respective spatial relationships are expressed as one-dimensional edges. Currently, IndoorGML does not explicitly support the representation of objects or facilities in these spaces, so the concept of the Point of Interest (POI) was implemented to expand IndoorGML's definitions of topological relationships from defining those of between spaces, towards the objects contained in these spaces as well.

The image and topological data are used in this study together to recognize objects and spaces in the images through the concept of the SEP, where user-identified pixels are related to the nodes of the IndoorGML data. The SEP signifies a region of influence for each node, and it enables a simplified solution to represent topological relationships between positions in the image and the nodes that represent spaces. In our experimental implementation, we collected omnidirectional images in the interior of a building, and various functions are implemented based on this established relationship, from an image-to-image movement visualization to the visualization of continuous indoor navigation. Also, using the expanded conceptualizations of the topological relationships in IndoorGML, objects within the indoor space are portrayed as indoor POI, also represented as nodes. These indoor POI are not just visualized through the images, but the data fusion method through the SEP has also enabled spatial analysis such as displaying attributes of facilities.

In our paper, the indoor topological data and the images are representing corridor types of indoor space. Because of this, there may be differences in the manner of generating IndoorGML data for other indoor environments. In addition, there may be differences in how the acceptable ranges for

using the SEP will be applied. Considering these factors, however, it can still be expected that indoor topological data can be linked to omnidirectional images to achieve similar results. Also, there is no officially established data model on how to represent indoor POI yet, and if such is created, this may help formalize how objects and facilities are to be represented along the spaces they are contained in.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

This paper is based on the first author's master's thesis.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by a grant (20NSIP-B135746-04) from the National Spatial Information Research Program (NSIP) funded by Ministry of Land, Infrastructure and Transport of Korean government.

References

- [1] K. Kim and K. Lee, "Handling Points of Interest (POIs) on a Mobile Web Map Service Linked to Indoor Geospatial Objects: A Case Study," *ISPRS International Journal of Geo-Information*, vol. 7, no. 6, p. 216, 2018.
- [2] N. E. Klepeis, W. C. Nelson, W. R. Ott et al., "The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants," *Journal of Exposure Analysis and Environmental Epidemiology*, vol. 11, no. 3, pp. 231–252, 2001.
- [3] M. P. Kwan and J. Lee, "Emergency response after 9/11: the potential of real-time 3D GIS for quick emergency response in micro-spatial environments," *Computers, Environment and Urban Systems*, vol. 29, no. 2, pp. 93–113, 2005.
- [4] H. Jung and J. Lee, "Development of an Omnidirectional-Image-Based Data Model through Extending the IndoorGML Concept to an Indoor Patrol Service," *Journal of Sensors*, vol. 2017, Article ID 5379106, 14 pages, 2017.
- [5] Z. Zheng, Y. Chen, S. Chen, L. Sun, and D. Chen, "Location-aware POI recommendation for indoor space by exploiting WiFi logs," *Mobile Information Systems*, vol. 2017, Article ID 9601404, 16 pages, 2017.
- [6] M. Gunduz, U. Isikdag, and M. Basaraner, "A Review of Recent Research in Indoor Modelling & Mapping," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B4, pp. 289–294, 2016.
- [7] S.-J. Oh and I.-P. Lee, "Georeferencing of Indoor Omnidirectional Images Acquired by a Rotating Line Camera," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 30, no. 2, pp. 211–221, 2012.

- [8] OGC (Open Geospatial Consortium), "IndoorGML v. 1.0.3," 2018, <http://docs.opengeospatial.org/is/14-005r4/14-005r4.html>.
- [9] D. Ahn, J. Park, and J. Lee, "Defining Geospatial Data Fusion Methods Based on Topological Relationships," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-4/W9, pp. 317–319, 2018.
- [10] H. Y. Kang and J. Lee, "A Study on the LOD(Level of Detail) Model for Applications based on Indoor Space Data," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 32, no. 2, pp. 143–151, 2014.
- [11] A. Königer and S. Bartel, "3d-Gis for Urban Purposes," *GeoInformatica*, vol. 2, no. 1, pp. 79–103, 1998.
- [12] S.-J. Lee and J.-Y. Lee, "Finding Isolated Zones through Connectivity Relationship Analysis in Indoor Space," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 30, no. 3, pp. 229–240, 2012.
- [13] N. A. Giudice, L. A. Walton, and M. Worboys, "The informatics of indoor and outdoor space: a research agenda," in *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness - ISA '10*, pp. 47–53, New York, NY, USA, 2010.
- [14] J. Lee and M. P. Kwan, "A combinatorial data model for representing topological relations among 3D geographical features in micro-spatial environments," *International Journal of Geographical Information Science*, vol. 19, no. 10, pp. 1039–1056, 2005.
- [15] S. Lee and J. Lee, "Validation of Efficient Topological Data Model for 3D Spatial Queries," *Journal of Korea Spatial Information Society*, vol. 19, no. 1, pp. 93–105, 2011.
- [16] J. Lee, "A Spatial Access-Oriented Implementation of a 3-D GIS Topological Data Model for Urban Entities," *GeoInformatica*, vol. 8, no. 3, pp. 237–264, 2004.
- [17] A. R. Claridades, I. Park, and J. Lee, "Integrating IndoorGML and Indoor POI Data for Navigation Applications in Indoor Space," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 37, no. 5, pp. 359–366, 2019.
- [18] J. Park, D. Ahn, and J. Lee, "Development of Data Fusion Method Based on Topological Relationships Using IndoorGML Core Module," *Journal of Sensors*, vol. 2018, Article ID 4094235, 14 pages, 2018.
- [19] F. Castanedo, "A Review of Data Fusion Techniques," *The Scientific World Journal*, vol. 2013, Article ID 704504, 19 pages, 2013.
- [20] J. Lee, H. Y. Kang, and Y. J. Kim, "Developing Data Fusion Method for Indoor Space Modeling based on IndoorGML Core Module," *Journal of Korea Spatial Information Society*, vol. 22, no. 2, pp. 31–44, 2014.
- [21] S. Stankute and H. Asche, "An Integrative Approach to Geospatial Data Fusion," *Computational Science and Its Applications - ICCSA 2009*, pp. 490–504, 2009.
- [22] G. Hammann, A. Marchetti, and K. Navulur, "Data fusion empowers geospatial analysis," *Earth Imaging Journal*, 2013, <https://eijournal.com/print/articles/data-fusion-empowers-geospatial-analysis> (accessed Apr. 13, 2020).
- [23] S. Wiemann, *Data Fusion in Spatial Data Infrastructures*, Dresden University of Technology, 2017.
- [24] S. Wiemann and L. Bernard, "Spatial data fusion in Spatial Data Infrastructures using Linked Data," *International Journal of Geographical Information Science*, vol. 30, no. 4, pp. 613–636, 2015.
- [25] H. Chen, Q. Sun, L. Xu, and Z. Xiong, "Application of data fusion in the production and updating of spatial data," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-7/W1, pp. 20–22, 2013.
- [26] N. Schmidt, C. Sas, and S. Barnett-Cormack, "Spatial data fusion with visual feedback," in *Proceedings of the British HCI Workshop on Visualisation and Interaction*, pp. 3–8, London, UK, 2008, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.1982&rep=rep1&type=pdf>.
- [27] J. Ok, W. Hee, and Y. Kiyun, "Matching Method of Digital Map and POI for Geospatial Web Platform," *Journal of Korean Society for Geospatial Information System*, vol. 17, no. 4, pp. 23–29, 2009.
- [28] P. Carrara, G. Bordogna, M. Boschetti, P. A. Brivio, A. Nelson, and D. Stroppiana, "A flexible multi-source spatial-data fusion system for environmental status assessment at continental scale," *International Journal of Geographical Information Science*, vol. 22, no. 7, pp. 781–799, 2008.
- [29] H. K. Kang and K. J. Li, "A Standard Indoor Spatial Data Model—OGC IndoorGML and Implementation Approaches," *ISPRS International Journal of Geo-Information*, vol. 6, no. 4, p. 116, 2017.
- [30] Y. J. Kim, H. Y. Kang, and J. Lee, "Development of indoor spatial data model using CityGML ADE," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-2/W2, pp. 41–45, 2013.
- [31] C. Y. Lee, V. Badrinarayanan, T. Malisiewicz, and A. Rabinovich, "RoomNet: end-to-end room layout estimation," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4875–4884, Venice, Italy, 2017.
- [32] C. Zou, A. Colburn, Q. Shan, and D. Hoiem, "LayoutNet: Reconstructing the 3D Room Layout from a Single RGB Image," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2051–2059, Salt Lake City, Utah, USA, 2018.
- [33] L. Google, "Google Street View," October 2019 <https://www.google.com/streetview/understand/>.
- [34] Kakao, "Kakao Maps," October 2019 <https://map.kakao.com/>.
- [35] A. R. Claridades, J. Lee, and A. Blanco, "Using Omnidirectional Images for Semi-Automatically Generating IndoorGML Data," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 36, no. 5, pp. 319–333, 2018.
- [36] K. Chen, Y. K. Lai, and S. M. Hu, "3D indoor scene modeling from RGB-D data: a survey," *Computational Visual Media*, vol. 1, no. 4, pp. 267–278, 2015.
- [37] E. Clementini, J. Sharma, and M. Egenhofer, "Modelling topological spatial relations: Strategies for query processing," *Computers and Graphics*, vol. 18, no. 6, pp. 815–822, 1994.
- [38] B. Lee, "Location Based Concierge Service with Spatially Extended Topology for Moving Objects," *Spatial Information Research*, vol. 17, no. 4, pp. 445–454, 2009.

Research Article

Characteristics of the Reanalysis and Satellite-Based Surface Net Radiation Data in the Arctic

Minji Seo ¹, Hyun-Cheol Kim ², Kyeong-Sang Lee ¹, Noh-Hun Seong ¹,
Eunkyung Lee ³, Jinsoo Kim ¹ and Kyung-Soo Han ¹

¹Division of Earth Environmental System Science (Major of Spatial Information Engineering), Pukyong National University, Busan 48513, Republic of Korea

²Unit of Arctic Sea-Ice Prediction, Korea Polar Research Institute, Incheon 21990, Republic of Korea

³Korea Ocean Satellite Center, Korea Institute of Ocean Science & Technology, Busan 49111, Republic of Korea

Correspondence should be addressed to Kyung-Soo Han; kyung-soo.han@pknu.ac.kr

Received 25 June 2020; Revised 18 August 2020; Accepted 28 August 2020; Published 17 September 2020

Academic Editor: Lei Zhang

Copyright © 2020 Minji Seo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this study, we compared four net radiation products: the fifth generation of European Centre for Medium-Range Weather Forecasts atmospheric reanalysis of the global climate (ERA5), National Centers for Environmental Prediction (NCEP), Clouds and the Earth's Radiant Energy System Energy Balanced and Filled (EBAF), and Global Energy and Water Exchanges (GEWEX), based on ground observation data and intercomparison data. ERA5 showed the highest accuracy, followed by EBAF, GEWEX, and NCEP. When analyzing the validation grid, ERA5 showed the most similar data distribution to ground observation data. Different characteristics were observed between the reanalysis data and satellite data. In the case of satellite-based data, the net radiation value tended to increase at high latitudes. Compared with the reanalysis data, Greenland and the central Arctic appeared to be overestimated. All data were highly correlated, with a difference of 6–21 W/m² among the products examined in this study. Error was attributed mainly to difficulties in predicting long-term climate change and having to combine net radiation data from several sources. This study highlights criteria that may be helpful in selecting data for future climate research models of this region.

1. Introduction

The Arctic has a complex climate system with atmosphere–ocean–land interactions and lower-latitude forcing that occur on various temporal and spatial scales [1]; thus, this area is particularly vulnerable to climate change [2, 3]. The surface radiation budget (SRB) is a key parameter for understanding the polar regions, which have a high albedo and emissivity [4–7]. SRB can be used to characterize the net radiation, specifically, the balance between downward and upward shortwave solar radiation and longwave thermal radiation from the atmosphere and various surfaces [8]. It is effectively the remainder of the radiative energy on the Earth's surface. Surface net radiation plays an important role in not only the energy cycle but also the ecosystem and carbon and water cycles [8–11]. Remote sensing of the radiative

flux can provide information on a global scale, as well as point-scale data from ground observations [12]. Due to these advantages, numerous radiative flux products have been derived from satellite and multispectral sensor data over the last decade [12]. Recently, various radiative flux products of the Arctic from satellite data and reanalysis data have been made available.

Satellite-based data, such as Clouds and the Earth's Radiant Energy System (CERES) [13], Global Energy and Water Exchanges Project (GEWEX) SRB [14], CloudSat [15], and Climate Monitoring Satellite Application Facility cLOUD, Albedo and surface Radiation dataset from Advanced Very High Resolution Radiometer (AVHRR) data-Edition 2 (CMSAF CLARA-A2) [16], provide information on radiative flux and radiative surface exchange processes. Reanalysis-based data, provided as absorbed solar radiation, includes

the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim) [17]; fifth-generation ERA (ERA5) [18]; National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) [19]; Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) [20, 21]; and the Japan Meteorological Agency's Japanese 55-Year Reanalysis (JRA-55) [22].

Numerous studies have attempted to evaluate these radiation products; however, most were conducted at low to mid-latitudes and only a few focused on the Arctic [23–29]. Riihelä et al. performed an intercomparison of radiative components on the Arctic during the spring and summer of 2007 [29]. Seo et al. compared CERES synoptic TOA and surface fluxes and clouds (SYN) and GEWEX SRB Arctic data from March 2003 to December 2007 [30]. Both of these studies analyzed only satellite data. Research has been conducted to develop the Arctic Observation and Reanalysis Integrated System (ArORIS) that combines satellite, reanalysis, and in situ products in the Arctic [28]. Studies evaluating the Arctic's net radiation have mostly been short term, on the order of 3 years or less, or are based on single-source data (satellite-based, reanalysis, or model data). Despite the many radiative products produced, comparative analysis by data type is lacking or has been carried out on accumulated data from the past. The analysis of recently updated data is still in progress.

In this study, we used data from the Arctic that has frequently been used in past research with recently updated radiation products. Among several reanalyzed data, ERA5 and NCEP data were used in this study. NCEP is a traditional reanalysis data developed in the 1990s, and the rest of the reanalysis data have been developed after the 2000s. ERA5 was selected because it is the most recently produced data and is calculated using various observation-based products as input data compared to past reanalysis data.

The purpose of this study was to assess the net radiation products over the Arctic using validation and intercomparison analyses. This paper is divided into two parts for validation and intercomparison analysis. The intercomparison analysis used four net radiation products: two reanalysis products, ERA5 and NCEP/NCAR, and two satellite-based products, CERES and GEWEX, for diversity. ERA5 and CERES Energy Balanced and Filled (EBAF) Edition 4.0 have been updated recently; NCEP/NCAR and GEWEX have been used for a longer period of time. This research is expected to provide some clarity regarding the accuracy of net radiation products and the relationship between product characteristics and comparative data from the Arctic for future modeling.

2. Materials and Methods

2.1. Data

2.1.1. ERA5. ERA5 is the fifth-generation ECMWF atmospheric reanalysis product of the global climate. ERA5 data are produced by Copernicus Climate Change Service (C3S) and provide the most recent atmospheric reanalysis of ECMWF. ERA5 is based on the new numerical weather

product (IFS Cycle 41r2, 2016 version) [18]. Compared to ERA-Interim, ERA5 offers improved spatial resolution, from 79 km to 31 km, and a higher output frequency, from every 6 hours (h) to hourly sampling from 1979 [31]. Therefore, ERA5 is a higher-resolution model, as a replacement for ERA-Interim [32, 33]. Atmospheric data assimilation has been upgraded to a 12-hour 4D-Var ensemble, and advanced climate data are used for model input [33]. Observation data such as satellite data, weather station, buoy, and radiosonde have been used on an average of 0.75 million per day in 1979, when data assimilation begins, and 24 million in recent years.

C3S provides ERA5 products with a spatial resolution of $0.25^\circ \times 0.25^\circ$ in spherical coordinates. We used the monthly averaged ERA5 radiative flux data to match the temporal resolution provided by EBAF. Monthly average data refers to the entire month and is generated from all hourly data for the month. ERA5 is provided in joule per square meter, not in a common unit; it is necessary to convert the unit for comparing with other radiation products. In the case of monthly mean data, it is converted into watt per square meter by dividing 86400 seconds (24 hours).

2.1.2. NCEP. The NCEP/NCAR is a reanalysis product [19]. The spatial resolution of NCEP comprises a T62 Gaussian grid ($\sim 1.91^\circ$), and temporal resolution is available at several levels, e.g., daily and monthly [34]. We selected monthly mean data to match the temporal resolution of EBAF data for the comparison. Radiative transfer schemes are based on the method of Lacis and Hansen [35] in the shortwave range and that of Schwarzkopf and Fels [36] in the longwave range [37].

2.1.3. EBAF. CERES EBAF is an all-sky surface flux product. Monthly data are presented as the monthly mean, with a spatial resolution of 1° on a uniform grid. The EBAF data used in this study were acquired over the coverage period from March 2000 to March 2018, using the Surface Radiation Edition 4.0 data product (May 2017 release). The EBAF product is calculated using EBAF top-of-atmosphere (TOA) flux, temperature/humidity profiles based on CERES SYN1deg-Month/SYN1deg-3hour, and cloud vertical profiles obtained from CALIPSO/CloudSat [38]. The EBAF product includes monthly surface upward/downward shortwave and longwave irradiances constrained by CERES-derived TOA irradiance [39]. CERES EBAF TOA Edition 4.0 was computed by Loeb et al. [40]. Kato et al. [39] determined the uncertainty in downward and upward shortwave irradiances of the North Pole to be 14 and 16 W/m^2 , respectively, using this product; both downward and upward longwave irradiances were 12 W/m^2 .

2.1.4. GEWEX. GEWEX SRB data provide several temporal resolutions of 1-hour, 3-hour, daily, and the monthly mean. We used monthly mean data to better coincide with the temporal resolution of EBAF for the comparison. The spatial resolution is 1° for uniform grid data. The coverage period of the data was from July 1983 to December 2007. The satellite-based data are calculated from readings obtained from the

TABLE 1: Summary of radiative flux products used in this study.

Dataset	Coverage	Temporal resolution	Spatial resolution	Based on	Version
ERA5	1979.01-2019.03	Monthly mean	$0.25^\circ \times 0.25^\circ$	Reanalysis	Monthly averaged reanalysis
NCEP	1948.01-2018.12	Monthly mean	T62 Gaussian grid	Reanalysis	Mean monthly NMC reanalysis
GEWEX	1983.07-2007.12	Monthly mean	$1^\circ \times 1^\circ$	Satellite	LW: rel3.1 SW: rel3.0
EBAF	2000.03-2018.03	Monthly mean	$1^\circ \times 1^\circ$	Satellite	Surface Ed4.0

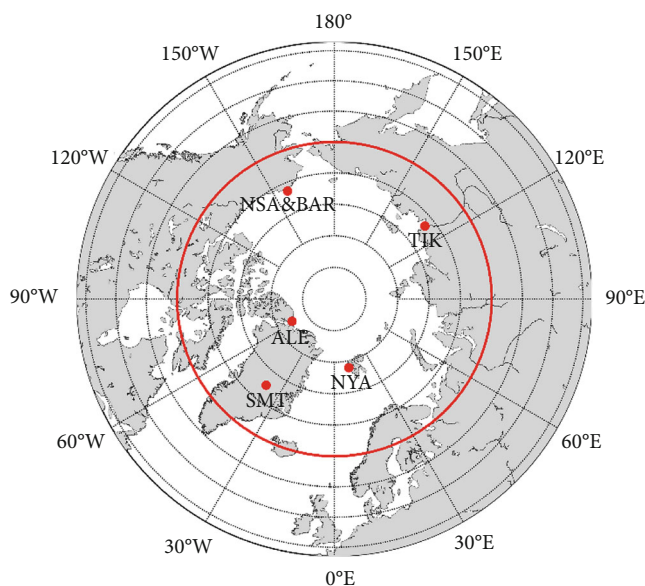


FIGURE 1: Locations of the six ground observation sites: Alert (ALE), Ny Alesund (NYA), Tiksi (TIK), Barrow (BAR), ETH/Greenland Summit (SMT), and North Slope of Alaska (NSA). The red circle indicates the Arctic Circle (poleward 65°N).

AVHRR sensor mounted on the National Oceanic and Atmospheric Administration satellite [41]. Radiative transfer schemes use the Pinker and Laszlo [42] method in the short-wave range and that of Fu et al. [43] in the longwave range. Commonly used cloud information was obtained from the International Satellite Cloud Climatology Project [29]. GEWEX monthly mean data for shortwave irradiance shows a root mean square (RMS) accuracy of 23.34 W/m^2 and a bias of -5.22 W/m^2 . In the longwave range, the RMS accuracy was 11.1 W/m^2 , with a bias of 0.9 W/m^2 from 1998 to 2007 [44, 45]. All accuracy values were calculated on a global scale.

Table 1 provides a summary of the spatial radiative flux data used in this study.

2.1.5. Ground Observation Data. Data from six ground observation sites monitored by three institutions were selected for this study, as discussed below. These data are used by the National Aeronautics and Space Administration for validation of CERES surface Edition 4.0 data and are provided by “EBAF Surface Ed4.0 Validation Subsetting and Browsing” and “CERES/ARM Validation Experiment (CAVE)” [46]. We used monthly averaged data with the lowest temporal resolution. The data provided by CAVE averages 1 min data into 15 min data; if there are any gaps, then the gaps are filled

via interpolation of the 15 min data. Hourly, daily, and monthly averaging was performed using interpolated data.

The sources for the ground observation data included the following institutions: the Baseline Surface Radiation Network (BSRN), the Swiss Federal Institute, and the United States Department of Energy Atmospheric Research Measurement (ARM) Program. The six ground data location sites (Figure 1) are as follows: ALE, NYA, TIK, BAR, SMT, and NSA. BSRN provided Alert (ALE), Ny Alesund (NYA), Tiksi (TIK), and Barrow (BAR) data. BSRN, initiated in 1992 through the World Climate Research Programme (WCRP), was designed in 2004 to be a global baseline network of surface radiant energy for the Global Climate Observing System (GCOS). It also contributes to the Global Atmosphere Watch program by providing high-quality data from a variety of climatic zones [47–49]. The Swiss Federal Institute of Technology (ETH) sources ETH/Greenland Summit (SMT) data that belongs to the Integrated Characterization of Energy, Clouds, Atmospheric State, and Precipitation at Summit (ICECAPS) project. ETH maintains the data, consisting of broadband radiative flux measurements at Summit Station [50]. North Slope of Alaska (NSA) data on the physical conditions of the Arctic atmosphere are provided by ARM [51]. Table 2 shows a detailed information on the six ground observation sites.

2.2. Method. This research is divided into two parts: (1) validation of the net radiation products using ground observation data and (2) an intercomparison among the products in an attempt to better understand the characteristics of the data (Figure 2). In the research flow chart, the yellow square box is the validation part and the green box is the intercomparison part. Validation is performed through three analyses after collocation with measured data using spatial data. And it analyzes the characteristics of the differences of each data through five intercomparison analyses.

In the validation part of the study, the four net radiation products were compared using ground observation data. Collocation of the spatial data was carried out using a minimum distance method. The method of the minimum distance found the nearest location using the great-circle distance. The distance between each point is obtained through

$$d = r \Delta\sigma$$

$$\Delta\sigma = 2 \arcsin \left(\sqrt{\sin^2 \left(\frac{\Delta lat}{2} \right) + \cos lat_1 \cos lat_2 \left(\frac{\Delta lon}{2} \right)} \right). \quad (1)$$

TABLE 2: Overview of the ground observation data.

Site	Location (latitude, longitude)	Observation site data source	Data periods
TIK	71.59, 128.92, Elev: 48 m	WCRP/GEWEX Baseline Surface Radiation Network (BSRN)	2011.04–2016.12
SMT	72.60, 321.58, Elev: 3216 m	Swiss Federal Institute, ETH broadband radiometer measurements	2011.01–2014.06
NYA	78.93, 11.93, Elev: 11 m	WCRP/GEWEX Baseline Surface Radiation Network (BSRN)	2000.03–2018.03
NSA	71.32, 203.39, Elev: 8 m	DOE Atmospheric Research Measurement (ARM) Program	2000.03–2018.03
ALE	82.45, 297.49, Elev: 200 m	WCRP/GEWEX Baseline Surface Radiation Network (BSRN)	2004.07–2014.03
BAR	71.32, 203.39, Elev: 8 m	WCRP/GEWEX Baseline Surface Radiation Network (BSRN)	2000.03–2017.08

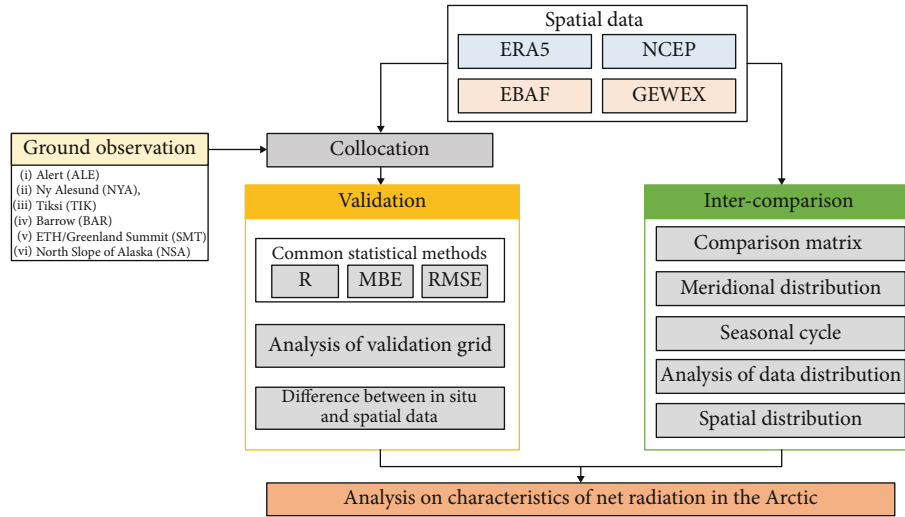


FIGURE 2: Flow chart to analyze the characteristics of the Arctic net radiation.

d refers the distance between ground observation data and spatial product. Find and compare the point where this distance is the minimum. r refers to the radius of sphere, Δ σ refers to the central angle, and lat_1 and lat_2 refer to the latitude at ground observation and net radiation products to be compared. The absolute difference between longitude and latitude between two points is denoted by Δlat and Δlon .

Three common statistical methods were used for the validation [11, 52]: the correlation coefficient (R), root mean square error/difference (RMSE/RMSD), and mean bias error (MBE). The R -value here indicates the correlation between observation data and spatial data of the net radiation products: closer to 1 (−1) indicates a strong positive (negative) relationship with ground observation or comparison data. $R > 0$ indicates that there is a positive correlation, $R < 0$ indicates a negative correlation, and $R = 0$ indicates that the data being compared are uncorrelated. RMSE is associated with the weight of the maximum error [53]; this is sometimes referred to as RMSD between values being modeled. MBE indicates systematic differences [52]. Positive (negative) MBE represents an overestimation (underestimation). The statistical characteristics of ground observation data were analyzed. In addition, the validation grid of net radiation data was analyzed using a box plot owing to the different spatial resolutions. The plot represents the information as a univariate distribution with the advantages of an approximate range and the ability to accommodate greater variability.

The second part of the study involved an intercomparison of the spatial net radiation products over the Arctic Circle, to analyze the differences among products used in actual research in the area. The study period was designated as March 2000 to December 2007, where all four datasets existed, thus over an approximately 8-year period. The intercomparison part of the study was further divided into four parts. First, the entire Arctic Circle was compared and analyzed as a large region from 65°N poleward. Second, a correlation analysis was conducted among the products. Third, the R -value, MBE, and RMSD were calculated. Finally, the results were configured in matrix form to intuitively analyze the correlation relationships among the radiation products for the Arctic Circle region. Individual data points were compared using the average net radiation in the Arctic Circle.

In the second part, the meridional distribution and seasonal cycle parts were analyzed separately. For analysis of the meridional distribution, we calculated the zonal mean at 2° intervals, as this corresponds to the spatial resolution of NCEP (with a resolution greater than 1°). The distribution of net radiation for each datum was analyzed by averaging every 2° from 60°N to 90°N. We also examined the seasonal cycle. In the third part, the distributions of actual pixel values were analyzed using histogram data and scatter density plots of the Arctic Circle region. Histogram analysis was conducted to confirm the data distribution over the comparison period by analyzing the maximum and minimum peak

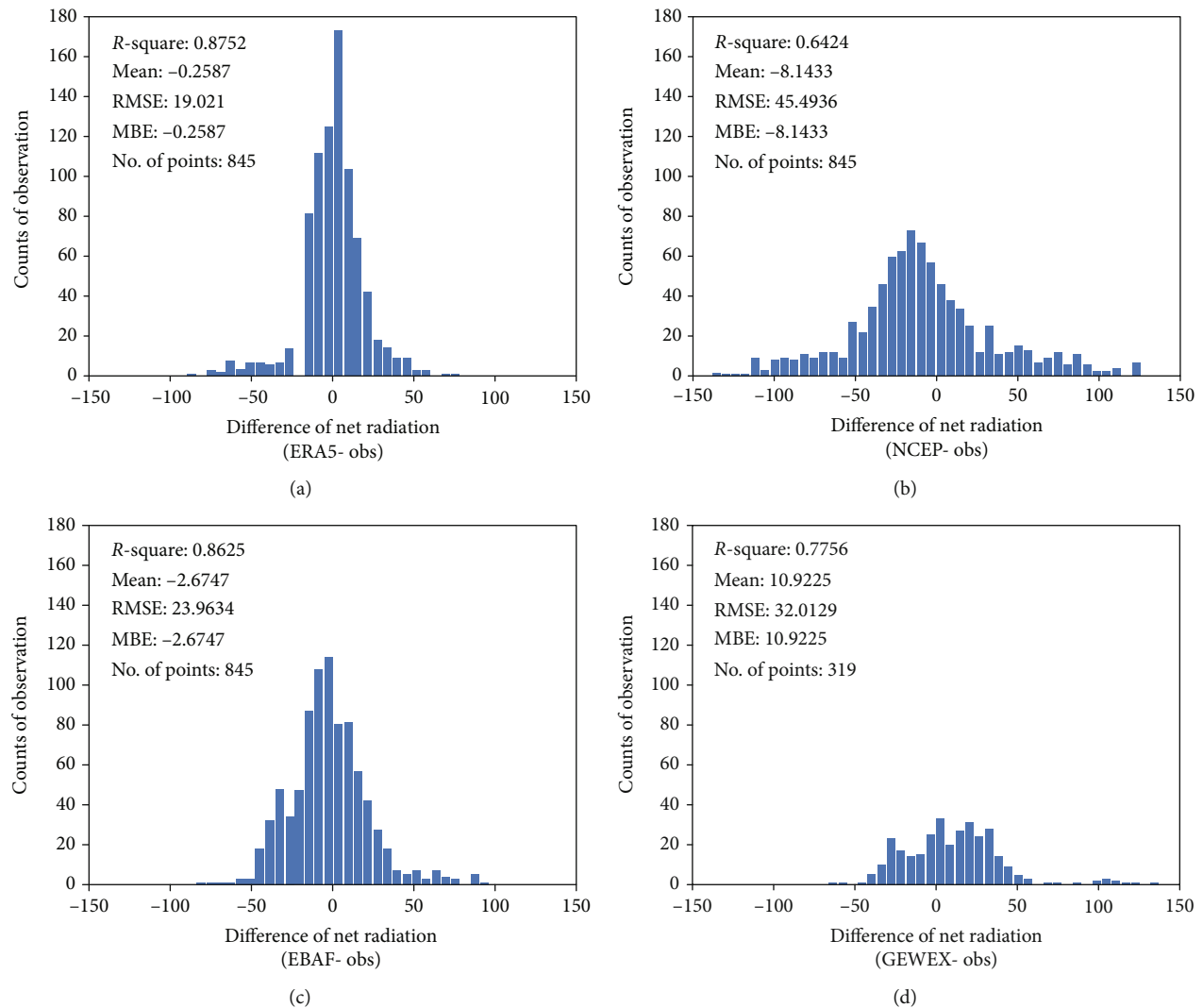


FIGURE 3: Histograms showing the difference between ground observation data and net radiation data from (a) the fifth generation of European Centre for Medium-Range Weather Forecasts atmospheric reanalysis of the global climate (ERA5), (b) National Centers for Environmental Prediction (NCEP), (c) Earth's Radiant Energy System Energy Balanced and Filled (EBAF), and (d) Global Energy and Water Exchanges (GEWEX) products.

distributions and the range of values. Scatter density plot analysis was performed with high accuracy, in connection with the validation part. Finally, we analyzed the differences in spatial distributions in matrix form to identify the characteristics of the products with respect to the Arctic region.

3. Results and Discussion

3.1. Validation with Ground Observation Data

3.1.1. Overview of the Validation. Figure 3 shows histogram plots of the differences between ground observation data and individual net radiation products. ERA5 data were more accurate than the other products, with a coefficient of determination (R^2) of 0.88, an RMSE of 19.02 W/m^2 , and a low MBE of $-0.26 W/m^2$. Compared to the other products, the difference was concentrated at 0, indicating a low variance.

In contrast, NCEP was observed to have the lowest accuracy among the products, with an R^2 value of 0.64, an RMSE of 45.49 W/m^2 , and an MBE of $-8.14 W/m^2$; in this case, the data appeared to be widely scattered, as evidenced by the high MBE and variance (standard deviation: 44.76 W/m^2). EBAF showed an R^2 of 0.86, an RMSE of 23.96 W/m^2 , and an MBE of $-2.37 W/m^2$, with a standard deviation of 23.81 W/m^2 as the second most accurate product. GEWEX was only available until 2007. Thus, the number of samples was small compared to the other data. Only the mean of the difference was observed in GEWEX, with an R^2 value of 0.78, an RMSE of 32.01 W/m^2 , an MBE of 10.92 W/m^2 , and a standard deviation of 30.09 W/m^2 . Overall, the most accurate model was ERA5, followed by EBAF, GEWEX, and NCEP.

The validation results for ground observation data are shown as Table 3. TIK and SMT locations have data going

TABLE 3: Validation of the four net radiation products by ground observation site (six sites total).

Site		ERA5	NCEP	EBAF	GEWEX
TIK	R	0.96	0.86	0.97	—
	RMSE (W/m^2)	14.92	36.77	22.22	—
	MBE (W/m^2)	5.22	5.05	-4.41	—
SMT	R	0.94	0.88	0.63	—
	RMSE (W/m^2)	8.45	23.25	20.78	—
	MBE (W/m^2)	-5.76	-10.47	18.61	—
NYA	R	0.94	0.91	0.93	0.91
	RMSE (W/m^2)	26.95	65.62	35.99	46.64
	MBE (W/m^2)	-18.52	-4.03	-7.11	12.27
NSA	R	0.97	0.83	0.96	0.94
	RMSE (W/m^2)	14.39	37.59	16.30	22.21
	MBE (W/m^2)	6.61	-10.15	-2.51	10.73
ALE	R	0.94	0.84	0.96	0.92
	RMSE (W/m^2)	21.41	28.94	17.64	21.63
	MBE (W/m^2)	13.79	-9.34	4.79	10.09
BAR	R	0.97	0.81	0.96	0.93
	RMSE (W/m^2)	14.52	41.29	19.1	25.07
	MBE (W/m^2)	2.83	-13.59	-6.13	10.10

back to 2007, preventing a comparison with GEWEX. EBAF showed a significantly lower correlation at SMT than other sites. ERA5 indicated high accuracy with a strong correlation, low RMSE, and an MBE approaching 0.

3.1.2. Analysis of the Net Radiation Distribution. The validation pixel of each data point was analyzed due to the different grid sizes of each product. The statistical distribution of the grid used for validation was confirmed, as shown in Figure 4; observation data are represented by a black box in the figure. Overall, ERA5 showed a denser distribution than NCEP, for which the distribution was scattered considerably. However, the interquartile range (IQR) for the two products was similar. Thus, the values of the net radiation were similarly distributed. In addition, all data exhibited a negative skewness with respect to their distributions, as the mean was higher than the median value. The distribution of data was observed differently depending on the surface conditions. In SMT, which is a homogeneous surface, the IQR of most data was observed to be small. However, in the case of heterogeneous surface condition as NYA, the IQR was determined according to the grid to validation. The NCEP with the lowest spatial resolution was observed to have the largest IQR, and the IQR of ERA5 with the high spatial resolution was observed to be small.

The statistical characteristics of each observation site are as follows. SMT showed an average net radiation of $-12.45 W/m^2$ (standard deviation: $11.73 (25.72\%) W/m^2$). This site had a short-duration coverage of just 3 years. SMT is located in the center of Greenland, and it has a homogeneous surface compared with the other sites. The data varia-

tion was also smaller at SMT compared with the other sites. BAR showed an average net radiation of $21.58 W/m^2$ (standard deviation: $60.72 (25.57\%) W/m^2$). Therefore, the data variation was larger than in other regions. The other ground observation sites, TIK, NYA, NSA, and ALE, showed average net radiation values of $19.34, 5.41, 16.16,$ and $-1.99, W/m^2$, respectively, with standard deviations of $51.58 (30.09\%), 48.28 (26.38\%), 56.87 (24.29\%),$ and $46.53 (25.43\%) W/m^2$.

3.2. Intercomparison of Net Radiation Products

3.2.1. Comparison Matrix of the Arctic Circle. In the intercomparison of four net radiation products, a confusion matrix was used to highlight the differences among data expressed in terms of the R -value, RMSD, and MBE (Figure 5). All products showed strong relationships with the other three products, with correlation coefficients exceeding 0.9. The MBE was positive, based on x -axis data. GEWEX appeared to overestimate the net radiation compared with the other products. The RMSD varied over a wide range, from 6.16 to $21.19 W/m^2$. The lowest RMSD corresponded to a comparison between EBAF and ERA5 but was highest between GEWEX and NCEP. In the reanalysis data, the secondary low in the RMSD was $9.60 W/m^2$; the difference from the largest to the smallest was in the following order: EBAF/NCEP, GEWEX/EBAF, and GEWEX/ERA5. GEWEX showed the largest difference in comparison with satellite-derived data, such as EBAF. Thus, all data showed a high correlation; however, the MBE and RMSD indicated high variation, despite similar patterns in the data. Thus, the differences associated with the variation in data and the range of

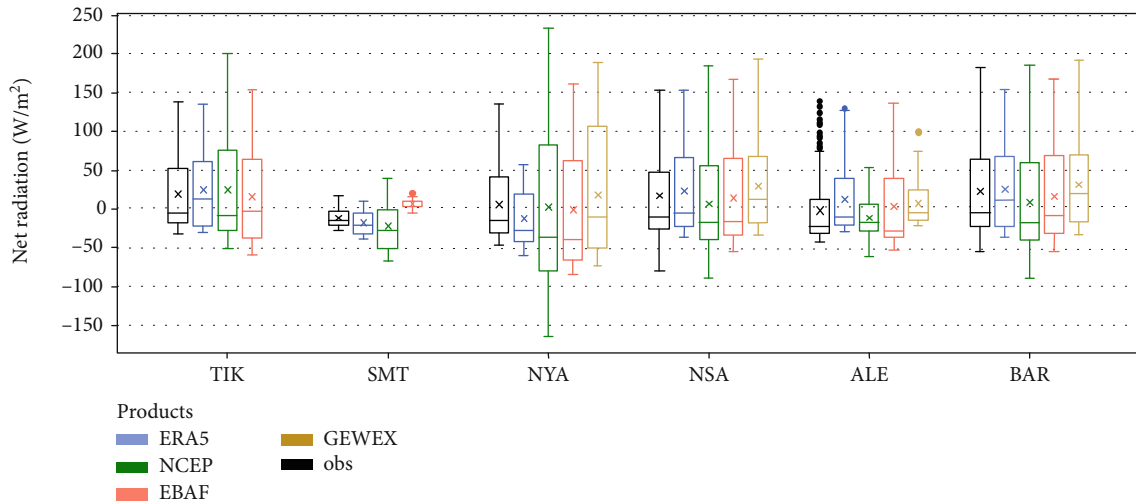


FIGURE 4: Distributions of the validation grid at ground observation sites; the x marker indicates the average. The reanalysis data of ERA5 and NCEP are indicated by blue and green, respectively, and satellite-based data for EBAF and GEWEX are indicated by red and yellow, respectively, with individual observations shown as a black dot.

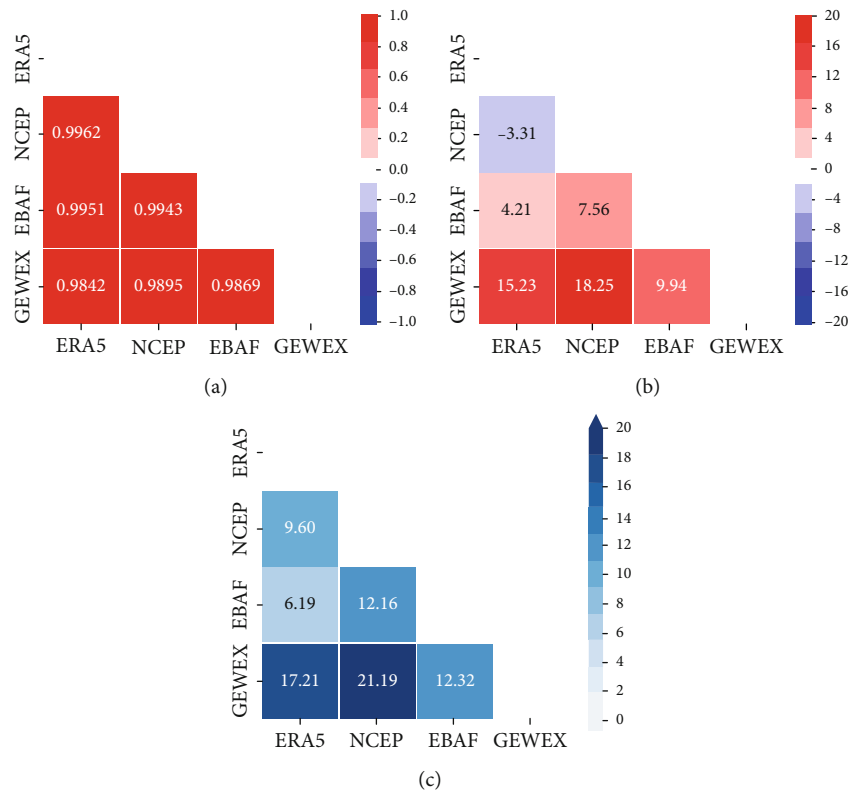


FIGURE 5: Confusion matrices for the intercomparison of all-grid average values from the four radiation products in the Arctic Circle: (a) correlation coefficients (R), (b) mean bias error (MBE), and (c) root-mean square difference (RMSD).

observing values appeared to strongly affect the agreement among radiation products.

3.2.2. Comparisons of Meridional Distribution and Seasonal Cycle. To understand the characteristics of net radiation data for latitude, we analyzed the 2° zonal mean for each of the four products; this was chosen based on the spatial resolution

of NCEP data. In a previous study, the same method was used to analyze meridional characteristics [54]. Figure 6(a) shows the net radiation distribution according to latitude. GEWEX data were $5\sim 10\text{ W/m}^2$ higher than other data. Net radiation was highest in the order of GEWEX, EBAF, ERA5, and NCEP. All data showed a decreasing pattern between latitudes 60°N and 78°N . However, above 78°N ,

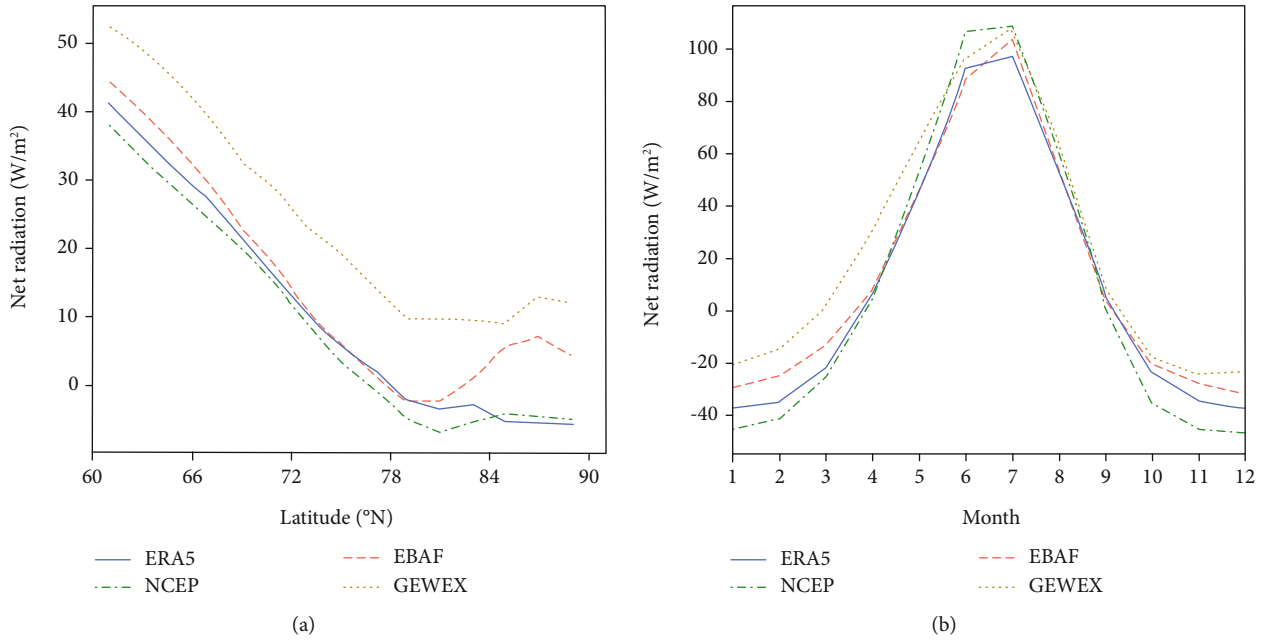


FIGURE 6: Two-degree zonal means and seasonal cycle in the Arctic: (a) distribution of the 2° zonal mean by latitude; (b) seasonal cycle of net radiation averaged in the Arctic Circle. In both graphs (a, b), the reanalysis data of ERA5 and NCEP are indicated by blue and green lines, respectively, and satellite-based data as EBAF and GEWEX are indicated by red and yellow lines, respectively.

differences among the distributions emerged. ERA5 showed a poleward decrease in net radiation. In NCEP net radiation data, a decreasing pattern was observed initially closer to 60° ; however, an increasing pattern developed at latitude $80^\circ N$. The net radiation distribution of EBAF data by latitude was divided into three ranges. In first range from $60^\circ N$ to $78^\circ N$, the net radiation of EBAF showed a decreasing trend, similar to ERA5 and NCEP. However, the EBAF distribution rose from latitude $78^\circ N$ to $86^\circ N$. Above $86^\circ N$, the EBAF net radiation pattern again showed a reduction. Similarly, the GEWEX net radiation exhibited a decreasing pattern up to latitude $78^\circ N$. From $78^\circ N$ to $84^\circ N$, the net radiation for GEWEX remained relatively the same, regardless of latitude.

Figure 6(b) shows the seasonal cycle distribution. Except for summer, GEWEX indicate overestimation compared to for net radiation products, which was consistent with MBE in the confusion matrix and the 2° zonal mean by latitude. In all data, the maximum value of net radiation corresponded to summer in July, and the minimum value coincided with winter in December. This pattern is similar to the variation in temperature and the number of hours of daylight [55]; thus, this characteristic is caused by the polar night, because incoming solar radiation is weak or nonexistent in the autumn and winter seasons. Thus, the seasonal variation of the net radiation can be divided into two periods: from April to September and from October to March. The period from April to September includes the melting season and summer. During this time, differences in the data were reduced. Notably, EBAF and ERA5 distributions were similar. The second period from October to March included the freezing season and winter. During this period, the differences among the radiation products increased and the net

radiation recorded increased in the order from the largest to the smallest of GEWEX, EBAF, ERA5, and NCEP.

3.2.3. Comparison with Distribution of Data per Grid. The confusion matrix analysis was performed by averaging the Arctic Circle region as one region. The Arctic has a highly seasonal cycle, as indicated in all four radiation products. However, using the 2° zonal mean, we confirmed differences in latitude patterns within the Arctic Circle. Figure 7 shows a histogram of the four net radiation products. ERA5 showed an average of $4.74 W/m^2$ and a standard deviation of $51.39 W/m^2$ for a total of 13,392,000 samples. In this case, the standard deviation accounted for 14.68% of the total value. NCEP had an average of $1.77 W/m^2$ and a standard deviation of $58.58 W/m^2$ for a total of 214,272 samples; here, the standard deviation accounted for 14.44% of the total value. EBAF produced an average of $10.07 W/m^2$ and a standard deviation of $50.51 W/m^2$ for a total of 837,000 samples, with the standard deviation accounting for 14.68% of the total value. ERA5 produced an average of $20.01 W/m^2$ and a standard deviation of $51.08 W/m^2$ for a total of 837,000 samples; in this case, the standard deviation accounted for 15.04% of the total value. The NCEP product had the widest data distribution, whereas GEWEX was more concentrated. Satellite data also indicated a higher mean compared to that of reanalysis data.

Figure 8 shows the monthly density scatter of net radiation for NCEP, EBAF, and GEWEX based on ERA5; here, the x -axis represents ERA5 and the y -axis represents comparative data in the figure. Most of the net radiation values were distributed over a low range ($<30 W/m^2$). All data had a R^2 above 0.9, indicating a strong relationship with ERA5.

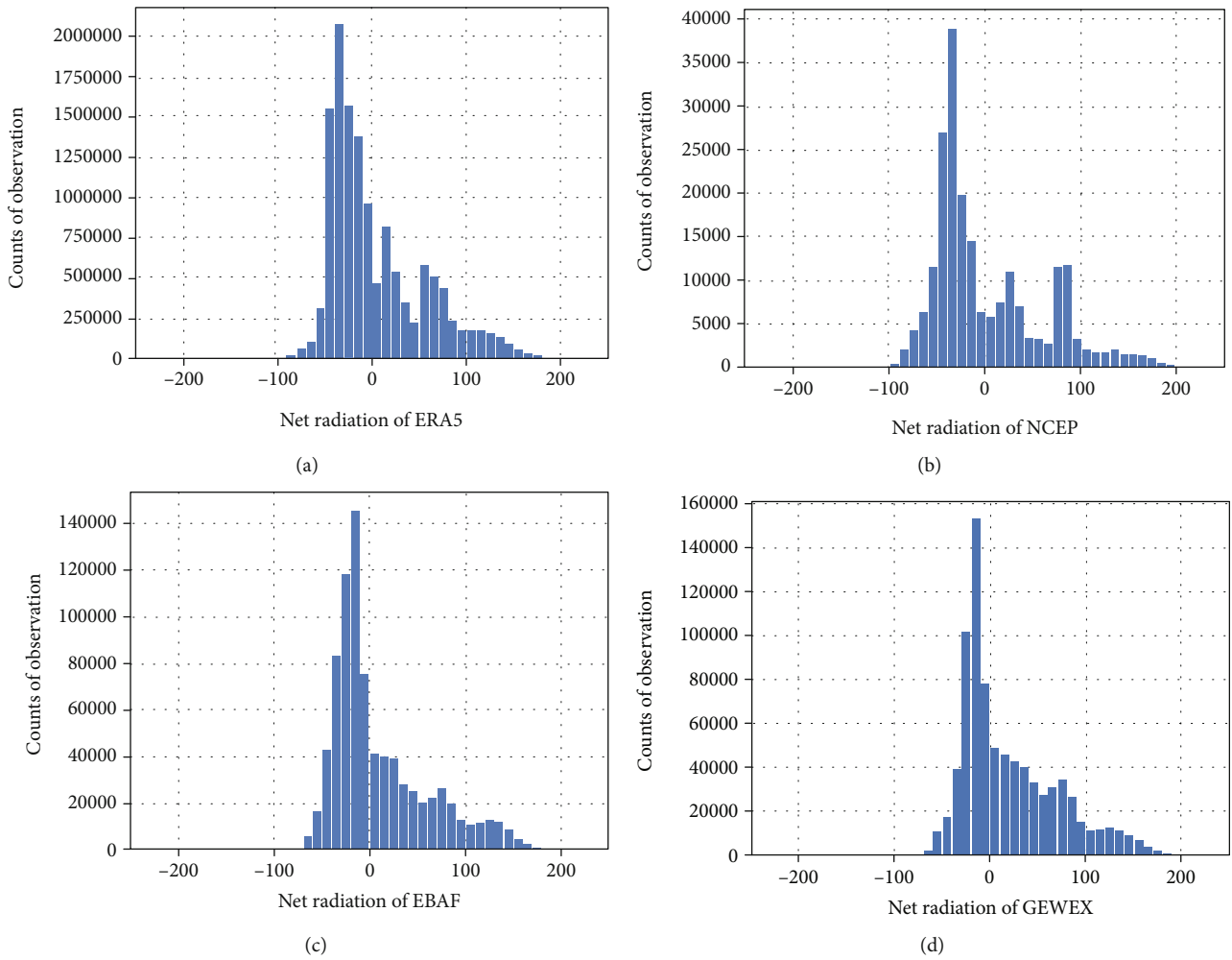


FIGURE 7: Histogram of net radiation in the Arctic Circle from March 2000 to December 2007: (a) ERA5, (b) NCEP, (c) EBAF, and (d) GEWEX.

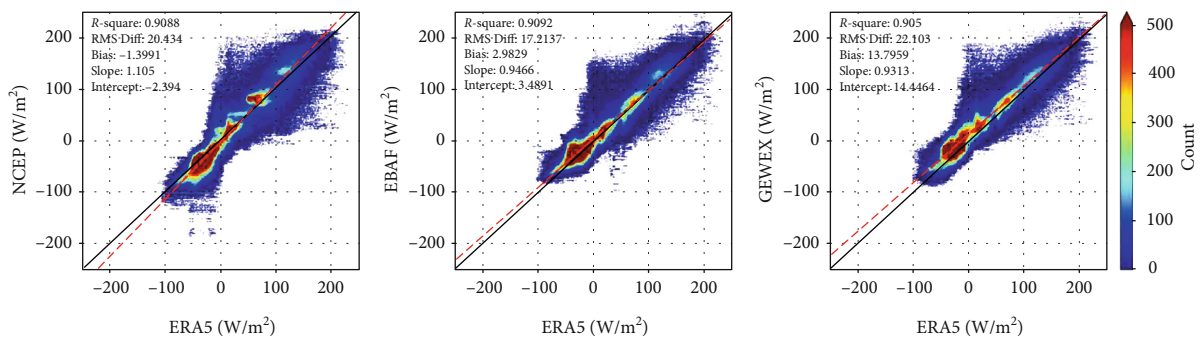


FIGURE 8: Density scatter between ERA5 and the other net radiation products. The x-axis represents ERA5 data, and the y-axis corresponds to (a) NCEP, (b) EBAF, and (c) GEWEX data.

With respect to ERA5, the differences among NCEP, EBAF, and GEWEX data were -1.40 ± 20.43 , 2.98 ± 17.21 , and $13.80 \pm 22.10 \text{ W/m}^2$, respectively. NCEP data were lower than the 1 : 1 line in range, where values below zero were distributed. In addition, the data indicated a more widespread distribution of net radiation. EBAF data were similar to ERA5 data, as indicated by the regression line fit. The regres-

sion line of GEWEX was generally higher than the 1 : 1 line; thus, GEWEX data tended to show higher values than ERA5. The density scatter distribution of GEWEX-ERA5 was more similar to EBAF data than NCEP.

3.2.4. Differences in Spatial Distribution. We analyzed the spatial differences in net radiation between NCEP, EBAF,

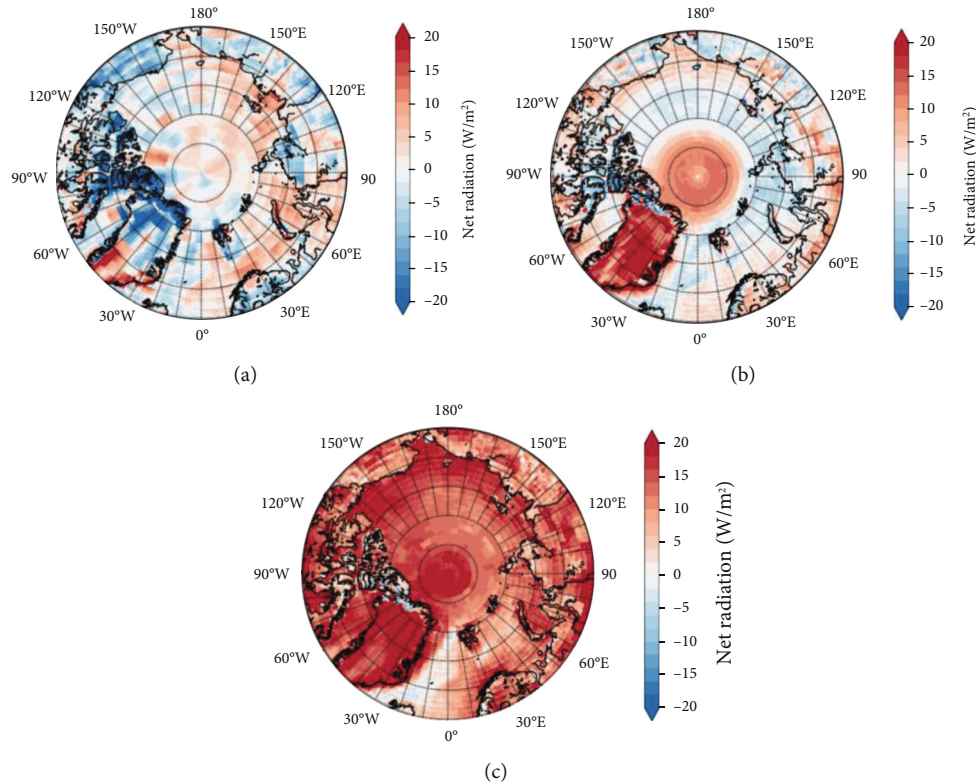


FIGURE 9: Distribution of spatial differences in net radiation based on ERA5 from March 2000 to December 2007. The red shading indicates an overestimation of the compared net radiation product, and blue shading indicates an underestimation. (a) NCEP–ERA5, (b) EBAF–ERA5, and (c) GEWEX–ERA5.

and GEWEX in the Arctic Circle, based on ERA5 data, as shown in Figure 9; the red shading indicates an overestimation of the corresponding data, and the blue shading indicates underestimation. In the case of NCEP, a mixture of overestimation and underestimation is shown in the figure. The net radiation differences were smaller over oceans than continents, with overestimation in the east and underestimation in the west. Regarding land areas, Canada and North Greenland showed a strong underestimation in net radiation, whereas South Greenland was overestimated. In the case of EBAF, less of a difference was observed compared with the other two datasets. Thus, substantial discrepancies were evident among the radiation products in Greenland. Also, the central Arctic area exhibited an overestimation of the net radiation, which was consistent with the 2° zonal mean by latitude. EBAF data indicated an increase in net radiation from 78°N to 86°N , whereas in the ocean adjoining the Central Arctic, a small underestimation was evident. In the case of GEWEX, the entire area was overestimated, with the exception of a few areas. The north coastline of Greenland and East Greenland Sea showed an underestimation in net radiation. These areas were underestimated in all comparisons; thus, ERA5 always estimated a high net radiation value in the area. These regional characteristics should be noted when using ERA5 data. Different regional characteristics were observed in GEWEX and EBAF. The two products are the same satellite-based data, but there is a difference in the radiative transfer scheme between the two products. The dif-

ference in the net radiated energy of the two data may occur primarily due to the difference in the method of calculating the radiation data. There are also differences in the cloud data used. CERES EBAF uses MODIS data and SAR-based CALIPSO/CloudSat data, and GEWEX uses ISCCP DX data. In Figures 9(b) and 9(c), both EBAF and GEWEX data are overestimated compared to ERA5 data in the area above about 80°N , which is the area where CALIPSO/CloudSat observations are missing. The two satellite-based data show the same characteristics in the region where CALIPSO/CloudSat data is excluded. Therefore, differences in satellite-based net radiation may occur due to differences in input cloud data.

4. Summary and Conclusions

In this study, we evaluated the surface net radiation from reanalysis (ERA5 and NCEP) and satellite-based (EBAF and GEWEX) products in the Arctic region. The maximum validation period was from March 2000 to March 2018, and the product comparison period was from March 2000 to December 2007, during which all data overlapped. The overall results for the validation process indicated ERA5 as having the highest accuracy of the four products. The quantitative accuracies of the four products were as follows: ERA5, -0.26 ± 19.02 ; EBAF, -2.37 ± 23.96 ; GEWEX, 10.92 ± 32.01 ; and NCEP, $-8.14 \pm 45.49 \text{ W/m}^2$. ERA5 and EBAF showed similar accuracies with respect to CERES EBAF (RMSE: 20.93, MBE: 2.40), with higher uncertainty at higher latitudes

[11]. The difference in accuracy may be due to spatial resolution. ERA5, the smallest valid grid, showed a distribution of values similar to the observation data. In the case of NYA, this particular observation site is located on an island (Svalbard, a Norwegian island); here, the grid size was larger, resulting in a larger IQR. The reason for the lower accuracy in EBAF than ERA5 can be attributed to the low accuracy of the CERES inversion for the snow surface and the uncertainty in the matchup near the coastal line due to the low spatial resolution [56].

The reason for focusing on these differences is that the Arctic is covered in ice; thus, the ice albedo effect is important. In particular, net radiation is connected to the ice-albedo feedback mechanism, making this feedback a key parameter of climate change in the Arctic [6, 57]. When analyzing climate change, differences in model accuracy and patterns of change are due to differences in the net radiation. Thus, it is essential to monitor and improve the assessments of surface energy budgets to better understand and predict climate change [28]. Continuous research in this area is essential for a process-level connection between the model and observations [58].

The purpose of this study was to attempt to better understand the accuracy of net radiation products in high-latitude areas and the differences among products. This helps when selecting a product for climate change modeling. Based on the results from this study, it is possible to analyze long-term energy changes in the Arctic by selecting the optimized net radiation energy.

The validation results were limited to ground observation sites covered with ice. Since the Arctic Circle is mainly composed of oceans, validation in sea ice is an essential part. Therefore, it is necessary to add the net radiation validation of sea ice using drift data in the future. In this study, differences occurred between net radiated products and the characteristics of the differences were analyzed. However, it was not possible to find out what factors caused such differences. Therefore, future research will need to consider the factors responsible for creating differences in the net radiation products. There are many different things that cause differences in net radiation. It is necessary to study what factors cause the difference in net radiation from the results of this study.

The radiative energy component is expected to play a key role in projecting accurate representations and climate change trends for the Arctic region. Changes in net radiation can help predict variables that cause physical processes in the Arctic such as Arctic sea ice. Based on the results of this study, it can be extended and utilized in long-term energy change observation, sea ice prediction model, and climate change modeling.

It is necessary to analyze which data is most appropriate for analyzing climate change and how the difference between the radiative products has an impact on climate change analysis. A radiative flux such as net radiation can be used as a key parameter to determine the energy balance of the Arctic. These results will enable accurate observation and prediction of Arctic climate and help to identify the causes of recent abnormal climates in connection with the Arctic midlatitude region.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the Korea Polar Research Institute (KOPRI), grant number PE20080. ARM data is made available through the U.S. Department of Energy as part of the Atmospheric Radiation Measurement Program.

References

- [1] M. C. Serreze, A. P. Barrett, A. G. Slater, M. Steele, J. Zhang, and K. E. Trenberth, "The large-scale energy budget of the Arctic," *Journal of Geophysical Research: Atmospheres*, vol. 112, no. D11, 2007.
- [2] IPCC, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, 2013.
- [3] Y. Huang, X. Dong, B. Xi, E. K. Dolinar, R. E. Stanfield, and S. Qiu, "Quantifying the uncertainties of reanalyzed Arctic cloud and radiation properties using satellite surface observations," *Journal of Climate*, vol. 30, no. 19, pp. 8007–8029, 2017.
- [4] S. C. Tsay, K. Stamnes, and K. Jayaweera, "Radiative energy budget in the cloudy and hazy Arctic," *Journal of the Atmospheric Sciences*, vol. 46, no. 7, pp. 1002–1018, 1989.
- [5] J. Cohen and D. Rind, "The effect of snow cover on the climate," *Journal of Climate*, vol. 4, no. 7, pp. 689–706, 1991.
- [6] M. Seo, H.-C. Kim, M. Huh et al., "Long-term variability of surface albedo and its correlation with climatic variables over Antarctica," *Remote Sensing*, vol. 8, no. 12, 2016.
- [7] K. S. Lee, D. Jin, J. M. Yeom et al., "New approach for snow cover detection through spectral pattern recognition with MODIS data," *Journal of Sensors*, vol. 2017, Article ID 4820905, 15 pages, 2017.
- [8] S. Liang, K. Wang, X. Zhang, and M. Wild, "Review on estimation of land surface radiation and energy budgets from ground measurement, remote sensing and model simulations," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 3, no. 3, pp. 225–240, 2010.
- [9] V. C. P. J. Ramanathan, P. J. Crutzen, J. T. Kiehl, and D. Rosenfeld, "Aerosols, climate, and the hydrological cycle," *Science*, vol. 294, no. 5549, pp. 2119–2124, 2001.
- [10] M. Wild, D. Folini, C. Schär, N. Loeb, E. G. Dutton, and G. König-Langlo, "The global energy balance from a surface perspective," *Climate Dynamics*, vol. 40, no. 11–12, pp. 3107–3134, 2013.
- [11] A. Jia, S. Liang, B. Jiang, X. Zhang, and G. Wang, "Comprehensive assessment of global surface net radiation products and uncertainty analysis," *Journal of Geophysical Research: Atmospheres*, vol. 123, no. 4, pp. 1970–1989, 2018.
- [12] J. M. Ramírez-Cuesta, D. Vanella, S. Consoli, A. Motisi, and M. Minacapilli, "A satellite stand-alone procedure for deriving

- net radiation by using SEVIRI and MODIS products,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 73, pp. 786–799, 2018.
- [13] B. A. Wielicki, B. R. Barkstrom, B. A. Baum et al., “Clouds and the Earth’s Radiant Energy System (CERES): algorithm overview,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 4, pp. 1127–1141, 1998.
- [14] S. J. Cox, AS&M, Hampton, VA, P. W. Stackhouse et al., “The NASA/GEWEX surface radiation budget project: overview and analysis,” in *12th Conference on Atmospheric Radiation*, Madison, Wis, 2006, July.
- [15] G. L. Stephens, D. G. Vane, S. Tanelli et al., “CloudSat mission: performance and early science after the first year of operation,” *Journal of Geophysical Research: Atmospheres*, vol. 113, no. D8, 2008.
- [16] K. G. Karlsson, K. Anttila, J. Trentmann et al., “CLARA-A2: the second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data,” *Atmospheric Chemistry and Physics*, vol. 17, no. 9, pp. 5809–5828, 2017.
- [17] D. P. Dee, S. M. Uppala, A. J. Simmons et al., “The ERA-Interim reanalysis: configuration and performance of the data assimilation system,” *Quarterly Journal of the Royal Meteorological Society*, vol. 137, no. 656, pp. 553–597, 2011.
- [18] F. Johannsen, S. Ermida, J. P. A. Martins, Trigo, Nogueira, and Dutra, “Cold bias of ERA5 summertime daily maximum land surface temperature over Iberian Peninsula,” *Remote Sensing*, vol. 11, no. 21, 2019.
- [19] E. Kalnay, M. Kanamitsu, R. Kistler et al., “The NCEP/NCAR 40-year reanalysis project,” *Bulletin of the American Meteorological Society*, vol. 77, no. 3, pp. 437–471, 1996.
- [20] M. M. Rienecker, M. J. Suarez, R. Gelaro et al., “MERRA: NASA’s Modern-Era Retrospective Analysis for Research and Applications,” *Journal of Climate*, vol. 24, no. 14, pp. 3624–3648, 2011.
- [21] M. G. Bosilovich, S. Akella, L. Coy et al., “MERRA-2: initial evaluation of the climate,” Technical Report Series on Global Modeling and Data Assimilation, NASA Tech. Memo. NASA/TM-2015-104606, 2015, (<https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich803.pdf>).
- [22] S. Kobayashi, Y. Ota, Y. Harada et al., “The JRA-55 reanalysis: general specifications and basic characteristics,” *Journal of the Meteorological Society of Japan. Ser. II*, vol. 93, no. 1, pp. 5–48, 2015.
- [23] J. Maslanik, J. C. Key, W. Fowler, T. Nguyen, and X. Wang, “Spatial and temporal variability of satellite-derived cloud and surface characteristics during FIRE-ACE,” *Journal of Geophysical Research: Atmospheres*, vol. 106, no. D14, pp. 15233–15249, 2001.
- [24] M. Chiacchio, J. Francis, and P. Stackhouse Jr., “Evaluation of methods to estimate the surface downwelling longwave flux during Arctic winter,” *Journal of Applied Meteorology*, vol. 41, no. 3, pp. 306–318, 2002.
- [25] J. Liu, J. A. Curry, W. B. Rossow, J. R. Key, and X. Wang, “Comparison of surface radiative flux data sets over the Arctic Ocean,” *Journal of Geophysical Research: Oceans*, vol. 110, no. C2, 2005.
- [26] J. E. Kay and T. L’Ecuyer, “Observational constraints on Arctic Ocean clouds and radiative fluxes during the early 21st century,” *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 13, pp. 7219–7236, 2013.
- [27] X. Dong, B. Xi, S. Qiu, P. Minnis, S. Sun-Mack, and F. Rose, “A radiation closure study of Arctic stratus cloud microphysical properties using the collocated satellite-surface data and Fu-Liou radiative transfer model,” *Journal of Geophysical Research: Atmospheres*, vol. 121, no. 17, pp. 10,175–10,198, 2016.
- [28] M. W. Christensen, A. Behrangi, T. S. L’ecuyer, N. B. Wood, M. D. Lebsock, and G. L. Stephens, “Arctic observation and reanalysis integrated system: a new data product for validation and climate study,” *Bulletin of the American Meteorological Society*, vol. 97, no. 6, pp. 907–916, 2016.
- [29] A. Riihelä, J. R. Key, J. F. Meirink, P. Kuipers Munneke, T. Palo, and K. G. Karlsson, “An intercomparison and validation of satellite-based surface radiative energy flux estimates over the Arctic,” *Journal of Geophysical Research: Atmospheres*, vol. 122, no. 9, pp. 4829–4848, 2017.
- [30] M. Seo, E. Lee, K.-s. Lee et al., “Comparative analysis of radiative flux based on satellite over Arctic,” *Korean Journal of Remote Sensing*, vol. 34, no. 6, pp. 1193–1202, 2018.
- [31] F. Al Senafi, A. Anis, and V. Menezes, “Surface heat fluxes over the Northern Arabian Gulf and the Northern Red Sea: evaluation of ECMWF-ERA5 and NASA-MERRA2 reanalyses,” *Atmosphere*, vol. 10, no. 9, 2019.
- [32] R. Urraca, T. Huld, A. Gracia-Amillo, F. J. Martinez-de-Pison, F. Kaspar, and A. Sanz-Garcia, “Evaluation of global horizontal irradiance estimates from ERA5 and COSMO-REA6 reanalyses using ground and satellite-based data,” *Solar Energy*, vol. 164, pp. 339–354, 2018.
- [33] H. Hersbach, B. Bell, P. Berrisford et al., “Global reanalysis: goodbye ERA-Interim, hello ERA5,” *ECMWF Newsletter*, vol. 159, pp. 17–24, 2019.
- [34] E. A. Nussbaumer and R. T. Pinker, “Estimating surface longwave radiative fluxes at global scale,” *Quarterly Journal of the Royal Meteorological Society*, vol. 138, no. 665, pp. 1083–1093, 2012.
- [35] A. A. Lacis and J. Hansen, “A parameterization for the absorption of solar radiation in the earth’s atmosphere,” *Journal of the Atmospheric Sciences*, vol. 31, no. 1, pp. 118–133, 1974.
- [36] M. D. Schwarzkopf and S. B. Fels, “The simplified exchange method revisited: an accurate, rapid method for computation of infrared cooling rates and fluxes,” *Journal of Geophysical Research*, vol. 96, no. D5, pp. 9075–9096, 1991.
- [37] S. Bony, Y. Sud, K. M. Lau, J. Susskind, and S. Saha, “Comparison and satellite assessment of NASA/DAO and NCEP-NCAR reanalyses over tropical ocean: atmospheric hydrology and radiation,” *Journal of Climate*, vol. 10, no. 6, pp. 1441–1462, 1997.
- [38] CERES, “EBAF-surface Ed4.0 data quality summary (CERES_EBAF-Surface_Ed4.0_DQS; June 27, 2018),” https://ceres.larc.nasa.gov/documents/DQ_summaries/CERES_EBAF-Surface_Ed4.0_DQS.pdf (accessed on 15 May 2020).
- [39] S. Kato, F. G. Rose, D. A. Rutan et al., “Surface irradiances of Edition 4.0 Clouds and the Earth’s Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) data product,” *Journal of Climate*, vol. 31, no. 11, pp. 4501–4527, 2018.
- [40] N. G. Loeb, D. R. Doelling, H. Wang et al., “Clouds and the Earth’s Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) top-of-atmosphere (TOA) Edition-4.0 data product,” *Journal of Climate*, vol. 31, no. 2, pp. 895–918, 2018.
- [41] A. A. Sörensson and R. C. Ruscica, “Intercomparison and uncertainty assessment of nine evapotranspiration estimates over South America,” *Water Resources Research*, vol. 54, no. 4, pp. 2891–2908, 2018.

- [42] R. T. Pinker and I. Laszlo, "Modeling surface solar irradiance for satellite applications on a global scale," *Journal of Applied Meteorology*, vol. 31, no. 2, pp. 194–211, 1992.
- [43] Q. Fu, K. N. Liou, M. C. Cribb, T. P. Charlock, and A. Grossman, "Multiple scattering parameterization in thermal infrared radiative transfer," *Journal of the Atmospheric Sciences*, vol. 54, no. 24, pp. 2799–2812, 1997.
- [44] T. Zhang, P. W. Stackhouse Jr., S. K. Gupta, S. J. Cox, J. Colleen Mikovitz, and L. M. Hinkelman, "The validation of the GEWEX SRB surface shortwave flux data products using BSRN measurements: a systematic quality control, production and application approach," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 122, pp. 127–140, 2013.
- [45] T. Zhang, P. W. Stackhouse Jr., S. K. Gupta, S. J. Cox, and J. C. Mikovitz, "The validation of the GEWEX SRB surface longwave flux data products using BSRN measurements," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 150, pp. 134–147, 2015.
- [46] D. A. Rutan, F. G. Rose, N. M. Smith, and T. P. Charlock, "Validation data set for CERES surface and atmospheric radiation budget (SARB)," *WCRP/GEWEX Newsletter*, vol. 11, no. 1, pp. 11–12, 2001.
- [47] A. Ohmura, H. Gilgen, H. Hegner et al., "Baseline surface radiation network (BSRN/WCRP): new precision radiometry for climate research," *Bulletin of the American Meteorological Society*, vol. 79, no. 10, pp. 2115–2136, 1998.
- [48] F. Feng and K. Wang, "Merging satellite retrievals and reanalyses to produce global long-term and consistent surface incident solar radiation datasets," *Remote Sensing*, vol. 10, no. 1, 2018.
- [49] W. Zhang, X. Zhang, W. Li et al., "Evaluation of Bayesian multimodel estimation in surface incident shortwave radiation simulation over high latitude areas," *Remote Sensing*, vol. 11, no. 15, 2019.
- [50] N. B. Miller, M. D. Shupe, C. J. Cox, V. P. Walden, D. D. Turner, and K. Steffen, "Cloud radiative forcing at Summit, Greenland," *Journal of Climate*, vol. 28, no. 15, pp. 6267–6280, 2015.
- [51] X. Dong and G. G. Mace, "Arctic stratus cloud properties and radiative forcing derived from ground-based data collected at Barrow, Alaska," *Journal of Climate*, vol. 16, no. 3, pp. 445–461, 2003.
- [52] A. Loew, W. Bell, L. Brocca et al., "Validation practices for satellite-based Earth observation data across communities," *Reviews of Geophysics*, vol. 55, no. 3, pp. 779–817, 2017.
- [53] T. Kato, "Prediction of photovoltaic power generation output and network operation," in *Integration of Distributed Energy Resources in Power Systems*, pp. 77–108, Academic Press, 2016.
- [54] L. Yu, Z. Zhang, S. Zhong et al., "An inter-comparison of six latent and sensible heat flux products over the Southern Ocean," *Polar Research*, vol. 30, no. 1, 2011.
- [55] M. Pidwirny, "Earth-sun relationships and insolation," in *Fundamentals of Physical Geography*, University of British Columbia, 2nd edition, 2006, Date Viewed. <http://www.physicalgeography.net/fundamentals/6i.html> (accessed on 15 May 2020).
- [56] A. Jia, B. Jiang, S. Liang, X. Zhang, and H. Ma, "Validation and spatiotemporal analysis of CERES surface net radiation product," *Remote Sensing*, vol. 8, no. 2, 2016.
- [57] D. K. Perovich, B. Light, H. Eicken, K. F. Jones, K. Runciman, and S. V. Nghiem, "Increasing solar heating of the Arctic Ocean and adjacent seas, 1979–2005: Attribution and role in the ice-albedo feedback," *Geophysical Research Letters*, vol. 34, no. 19, 2007.
- [58] J. E. Kay, T. L'Ecuyer, H. Chepfer, N. Loeb, A. Morrison, and G. Cesana, "Recent advances in Arctic cloud and climate research," *Current Climate Change Reports*, vol. 2, no. 4, pp. 159–169, 2016.

Research Article

Land Cover Classification Using SegNet with Slope, Aspect, and Multidirectional Shaded Relief Images Derived from Digital Surface Model

Dae Geon Lee,¹ Young Ha Shin,² and Dong-Cheon Lee ²

¹Technology Department, Geo & Co., Ltd., Goyang-si 10495, Republic of Korea

²Department of Environment, Energy & Geoinformatics, Sejong University, Seoul 05006, Republic of Korea

Correspondence should be addressed to Dong-Cheon Lee; dclee@sejong.ac.kr

Received 28 July 2020; Revised 20 August 2020; Accepted 28 August 2020; Published 12 September 2020

Academic Editor: Sang-Hoon Hong

Copyright © 2020 Dae Geon Lee et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Most object detection, recognition, and classification are performed using optical imagery. Images are unable to fully represent the real-world due to the limited range of the visible light spectrum reflected light from the surfaces of the objects. In this regard, physical and geometrical information from other data sources would compensate for the limitation of the optical imagery and bring a synergistic effect for training deep learning (DL) models. In this paper, we propose to classify terrain features using convolutional neural network (CNN) based SegNet model by utilizing 3D geospatial data including infrared (IR) orthoimages, digital surface model (DSM), and derived information. The slope, aspect, and shaded relief images (SRIs) were derived from the DSM and were used as training data for the DL model. The experiments were carried out using the Vaihingen and Potsdam dataset provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) through the International Society for Photogrammetry and Remote Sensing (ISPRS). The dataset includes IR orthoimages, DSM, airborne LiDAR data, and label data. The motivation of utilizing 3D data and derived information for training the DL model is that real-world objects are 3D features. The experimental results demonstrate that the proposed approach of utilizing and integrating various informative feature data could improve the performance of the DL for semantic segmentation. In particular, the accuracy of building classification is higher compared with other natural objects because derived information could provide geometric characteristics. Intersection-of-union (IoU) of the buildings for the test data and the new unseen data with combining all derived data were 84.90% and 52.45%, respectively.

1. Introduction

The field of DL has grown significantly over the past decade coupled with rapid improvements in computer performance. Since McCulloch and Pitts introduced artificial neuron that is a computational model of the neural networks to mimic the human brain in 1943 [1], DL as a branch of machine learning has evolved steadily to this day. In recent years, advances in image processing computer vision, information and communication technology, and geoinformatics have accelerated the development and use of the DL. Many DL tasks involve visual information processing such as object recognition and identification from imagery [2–4]. It is true that the optical images provide rich, diverse, and explicit information.

“Seeing is believing,” i.e., human beings have been relying on the visual information to understand the real world than through any other media. Images obtained from optical sensors are formed by recording reflected light, in the visible spectral range, from the surfaces of the terrain objects. In this aspect, it is not sufficient to reveal real-world features by utilizing image alone.

Performance of the DL depends on the training data that sometimes leads to the issue of overfitting. Various schemes have been suggested to avoid overfitting such as drop-out, early-stopping, regularization, cross-validation, and hyperparameter tuning. However, such efforts might not be the fundamental solutions to prevent overfitting. Extracting intrinsic and characteristic information from the original

data to utilize training data would make DL more robust. Recent deep learning researches with a similar concept are found. Maltezos et al. proposed building detection method by training the CNN model with multidimensional features that include entropy, height variation, intensity, normalized height, planarity, and standard deviation extracted from light detection and ranging (LiDAR) data. Each feature reflects the unique physical property of the objects [5]. Audebert et al. introduced the SegNet DL model with multimodal and multiscale data for semantic labeling. Optical IR images, DSM created from LiDAR data, normalized DSM (nDSM), and normalized difference vegetation index (NDVI) derived from multispectral data are used to train the DL model [6]. Zhou et al. proposed a CNN-based AlexNet DL model with fusion of point cloud data and high-resolution images for land cover classification. Specifically, a stratified segmentation scheme with grey-level cooccurrence matrix (GLCM) [7, 8] was introduced to improve segmentation efficiency for increasing classification accuracy. Mean, variance, homogeneity, dissimilarity, and entropy from GLCM features were DL model training process [9]. Alidoost and Arefi applied LeNet-5 [10] for automatic recognition of various roof types by training features extracted from both LiDAR data and orthoimages [11]. Pibre et al. presented multiple source data fusion strategies to detection trees [12]. Two different modes were applied to integrate heterogeneous data for training DL models that is a similar idea to [6]. In the study, early fusion and late fusion of the IR aerial images and NDVI and DSM created from LiDAR data were performed. Better results were obtained with early fusion when both NDVI and DSM were used [12, 13].

The recent researches of DL have focused on efforts to improve training performance by utilizing multisource data and/or creating information from the raw data. In the field of geoinformatics and remote sensing, the major data sources are optical and laser sensors. Point cloud data that consists of 3D coordinates can be obtained directly from laser sensors (i.e., LiDAR data) or indirectly by stereo image matching. In addition, derived information from the original raw data are utilized. Examples of the derived information are NDVI and DSM created from multispectral imagery and LiDAR data, respectively. Combining multisource data (e.g., optical and multispectral imagery, point cloud data, and DSM) with derived information from original raw data (e.g., NDVI, cooccurrence features, and surface orientation) provide more reliable results.

As for the DL models, CNN is one of the most extensively used models and has been successfully applied to high-level image processing (or late-vision in computer vision) tasks such as image recognition, scene analysis and description, classification, and medical image computing. Well-known CNN models are LeNet, AlexNet, GoogLeNet, VGGNet, ResNet, Mask R-CNN, SegNet, etc. Most of them are winners of the Image Net Large Scale Visual Recognition Challenge (ILSVCR) [14]. One of the major applications of DL is object classification from images [15, 16]. The goal of this paper is to perform semantic segmentation for land cover classification using CNN based SegNet model. The training datasets are composed of IR images and DSM with derived informa-

tion from the DSM. The derived information includes surface orientation (i.e., slope and aspect of each DSM grid cell) and multidirectional SRIs. DSM is a 3D representation of the terrain surface features, including natural and man-made objects, that are formed with dense point clouds of the 3D coordinates (X, Y, Z). In the field of geoinformatics, the point clouds with 3D coordinates can be obtained directly from laser sensors (i.e., LiDAR data) or indirectly by stereo image matching. Since images are 2D data and provide reflected light from the terrain surfaces, images lack 3D information about the real-world objects. On the other hand, DSM could provide richer 3D geometric information of the objects than the imagery [17].

Feature is a primitive characteristic or attribute of the data. Thus, it is important to extract or derive unique features from the various data, then use those features for DL model training for semantic segmentation. Segmentation entails division or separation of the data into regions of similar characteristics or attributes. Therefore, the purpose of the segmentation is to form meaningful regions by grouping features that have common properties distinct from their neighboring regions [18]. In general, segmentation does not involve classifying each segment but rather subdivides regions without attempting to recognize the individual regions. On the other hand, semantic segmentation or classification involves identification and labeling of the regions or individual objects according to the category [15]. There is no theory or standard method available for the segmentation yet. Most of the existing methods are based on the hoc approach; therefore, DL is expected to be a promising innovative method to solve such a challenging task that requires human intelligence.

The experiments for this paper were carried out using the ISPRS benchmark dataset for terrain feature classification. The land cover of the study site is categorized into six classes; building, tree, low vegetation (grass), impervious surface (road), car, water, and clutter/background in the label data. The main intent of this paper is to classify terrain objects by training the DL model using multisource data; optical IR images, DSM, and DSM-derived data including slope, aspect, and multidirectional SRIs. The experiments were carried out as follows: (1) training with each type of data independently, and (2) training with combining all data. Each data collected from a specific sensor could not convey sufficient information about the real-world. In this aspect, multisource data would be complementary in the training process. The results were analyzed based on evaluation metrics and visual inspection. In conclusion, training by combining multisource data could provide a synergistic effect and multidirectional SRI plays an important role.

2. Materials and Proposed Methods

2.1. Description of Datasets. The ISPRS benchmark dataset [19, 20] was used for training, evaluation, and test of the SegNet model. Table 1 shows the description of the datasets, and Figure 1 presents the configuration of Vaihingen datasets. The datasets consist of IR true orthoimages, DSM, airborne LiDAR data, and label data. True orthoimages are

TABLE 1: Description of ISPRS benchmark dataset for the Vaihingen area.

Data type	GSD (m)	Data collection	Comment
Color IR (IR-R-G) true orthoimage	0.09	Intergraph DMC (altitude: 900 m)	33 patches with georeferenced TIFF format
DSM	0.09	Point clouds created by image matching	16 patches with georeferenced TIFF format
LiDAR data	0.50 (4 pts/m ²)	Leica ALS50 (altitude: 500 m)	10 strips with LAS format
Label data	0.09	Manually digitized ground truth	Land cover classification of 6 classes with TIFF format

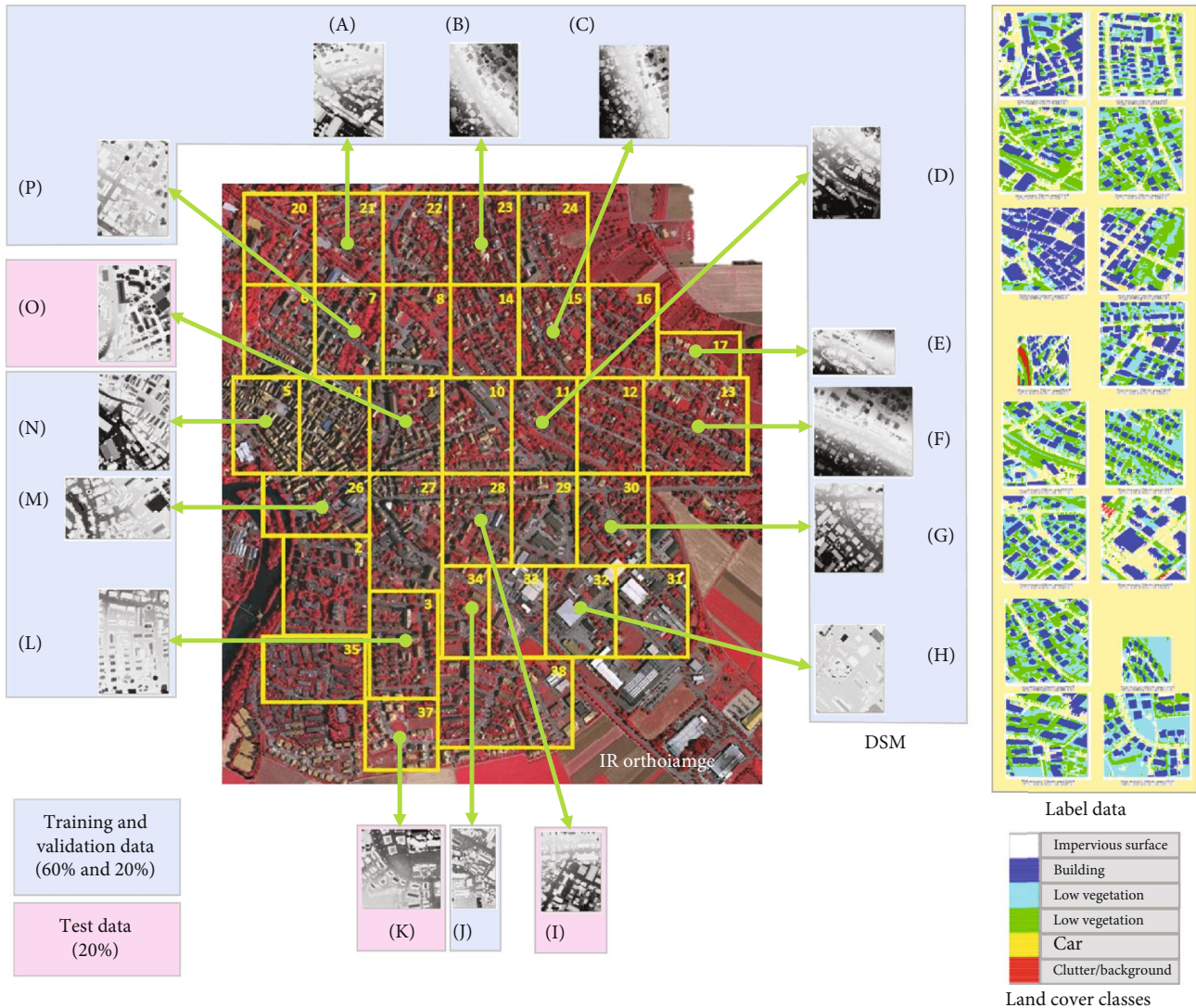


FIGURE 1: Vaihingen dataset is composed with IR images, DSM, and label data for training, validation, and test.

geometrically corrected, occlusion-free, and georeferenced images. 60%, 20%, and 20% of the datasets were used as training, evaluation, and test data, respectively. Figure 2 is another dataset from the Potsdam area that was used as “new unseen” data to apply to the trained DL model that was trained with the Vaihingen datasets. Four datasets that are 3_12, 4_10, 4_11, and 7_08 area were selected as new unseen data for the experiments. It is notable to mention that LiDAR data was not used in the experiment; instead, DSM was used to derive slope and aspect and to create multi-

directional SRIs. DSM could be generated using LiDAR data. However, some significant preprocessing is required such as noise removing and interpolation because of the irregularly distributed point clouds with low point density of the LiDAR data. The high-resolution DSM in the ISPRS dataset was created using INPHO MATCH-DSM software with sequential multiimage matching and finite element interpolation [21]. In this regard, we decided that utilizing DSM is more appropriate than LiDAR data to perform the proposed approach.



FIGURE 2: New unseen dataset from the Potsdam area.

2.2. *Slope and Aspect.* The slope and aspect (i.e., surface orientation) of the surface elements (i.e., DSM grid cells) can be computed from the coordinates of the 3×3 grid window of the DSM (see Figure 3).

Slope and aspect of each surface element is computed by Equations (1)–(10).

$$dX_1 = \frac{Z(X+1, Y+1) - Z(X-1, Y+1)}{2 \cdot \Delta X}, \quad (1)$$

$$dX_2 = \frac{Z(X+1, Y) - Z(X-1, Y)}{2 \cdot \Delta X}, \quad (2)$$

$$dX_3 = \frac{Z(X+1, Y-1) - Z(X-1, Y-1)}{2 \cdot \Delta X}. \quad (3)$$

The average slope in X -direction is

$$S_X = \frac{dX_1 + dX_2 + dX_3}{3}. \quad (4)$$

Slopes in Y -direction are computed by

$$dY_1 = \frac{Z(X-1, Y+1) - Z(X-1, Y-1)}{2 \cdot \Delta Y}, \quad (5)$$

$$dY_2 = \frac{Z(X, Y+1) - Z(X, Y-1)}{2 \cdot \Delta Y}, \quad (6)$$

$$dY_3 = \frac{Z(X+1, Y+1) - Z(X+1, Y-1)}{2 \cdot \Delta Y}. \quad (7)$$

The average slope in Y -direction is

$$S_Y = \frac{dY_1 + dY_2 + dY_3}{3}, \quad (8)$$

where ΔX and ΔY are ground sampling distance (GSD) of DSM in X - and Y -direction, respectively. Finally, slope (S) and aspect (A) at the center of the 3×3 DSM grid window are computed as follows [22]:

$$S = \tan^{-1} \left(\sqrt{S_X^2 + S_Y^2} \right), \quad (9)$$

$$A = \tan^{-1} \left(\frac{S_Y}{S_X} \right). \quad (10)$$

Figure 4 illustrates examples of DSM, slope, and aspect using Vaihingen test data.

2.3. *Multidirectional Shaded Relief Images.* Shading is an important visual cue (i.e., shape-from-shading) to recognize

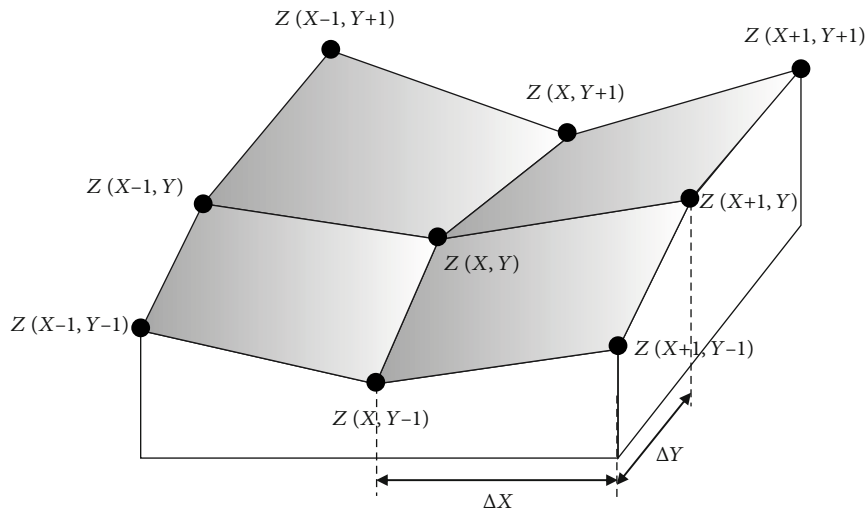


FIGURE 3: 3 × 3 grid window of DSM for computing slope and aspect.

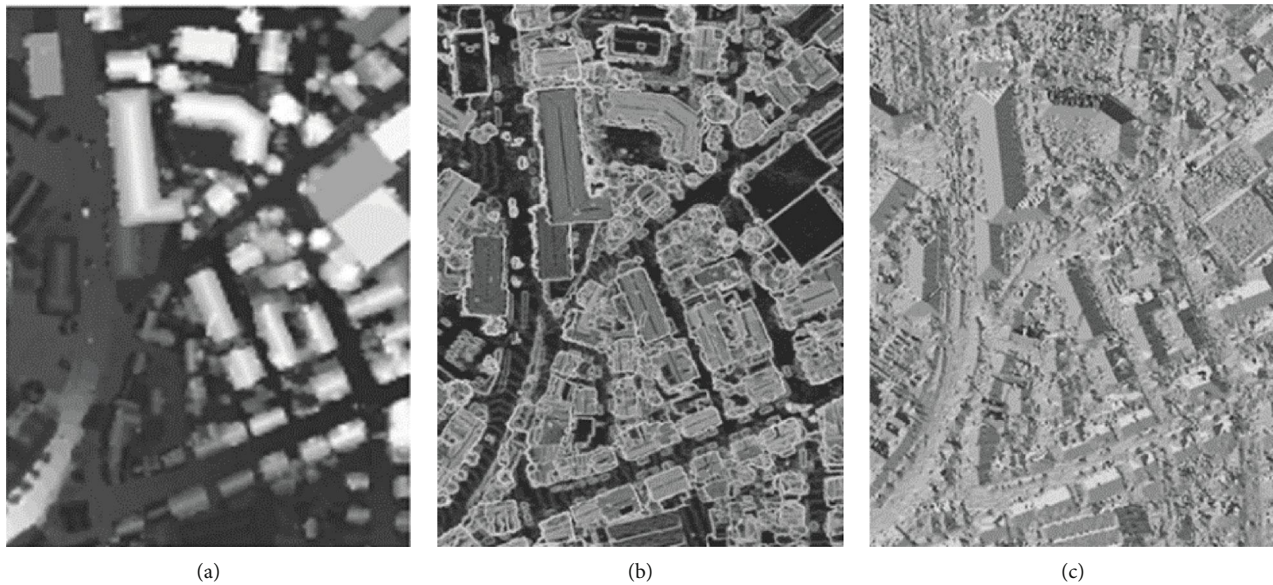


FIGURE 4: DSM and derived data. (a) DSM. (b) Slope. (c) Aspect.

the shape of the objects because each surface patch of the objects appears in different brightness depending on the orientation of the surfaces [23–25]. The SRIs can be created using surface orientation (i.e., slope and aspect) of the DSM and virtual light source. The amount of the reflected light from each surface element is recorded in the corresponding pixel of the SRIs and computed by Equation (11).

$$R = \cos(90^\circ - \beta) \cos(S) + \sin(90^\circ - \beta) \sin(S) \cos(\alpha - A), \quad (11)$$

where R is the magnitude of the reflected light, α and β are azimuth and elevation angle of the light source, respectively, and S and A are slope and aspect of the surface element, respectively. Each value of R is converted to an 8-bit image

with range from 0 to 255. The multidirectional SRIs are generated by changing the location of the light source [26]. Four SRIs were generated with light sources of NW, NE, SE, and SW direction (see Figure 5).

2.4. Preparation of Dataset for Training, Evaluation, and Label Data. Each region of the dataset, including training, evaluation, and label data, is partitioned with a 50% overlap along both X and Y directions, and the partitioned tile size is 150×250 as shown in Figure 6. Data partitioning, especially with overlapping, could provide significant benefits: (1) increasing amount of the training data because DL requires a large number of data, and (2) improving DL performance by training the same object (or area) in various situations.

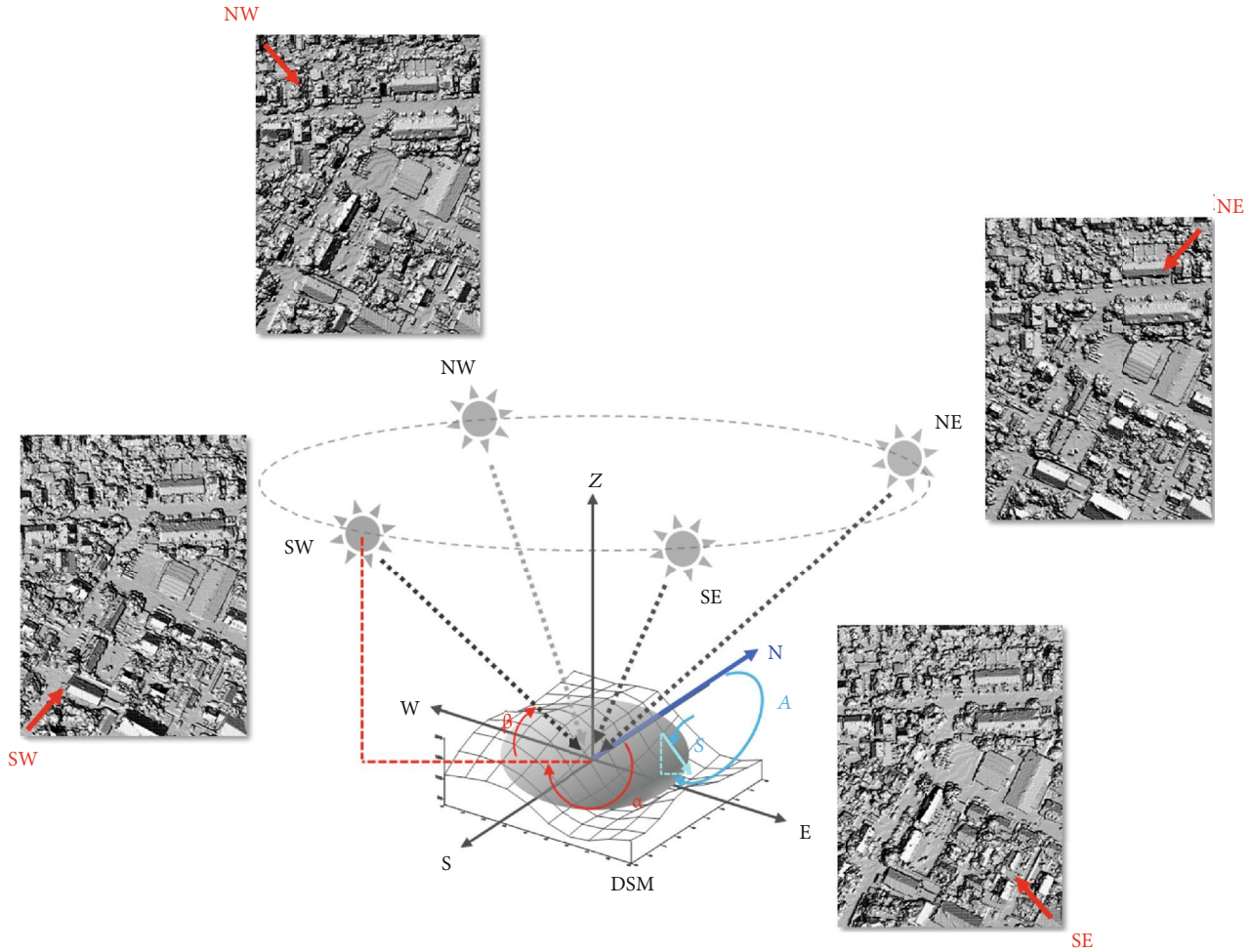


FIGURE 5: Multidirectional shaded reliefs images created from different locations of the light source.

2.5. SegNet Model. DL using the CNN has been proved to be successful and effective on semantic segmentation (or called semantic labeling) [15]. The SegNet model is a deep fully CNN for multiclass pixel-wise semantic segmentation. SegNet, by Computer Vision and Robotics Group at the University of Cambridge, was designed to be efficient in memory and computational time. It is also significantly smaller in the number of the trainable parameters than other network models. The SegNet is composed of encoder and decoder with symmetrical architecture, and the encoder is based on the convolutional layers from VGG16 that has 13 convolutional and 3 fully-connected layers. VGG16 has been developed by Visual Geometry Group of the University of Oxford and 1st runner-up of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) with an error rate of under 10% in 2014 [27].

On the other hand, the fully-connected layers in VGG16 are replaced by convolutional layers in SegNet. The SegNet uses a bottleneck architecture in which the feature maps are upsampled to match the original input resolution, and therefore it is possible to perform pixel-wise segmentation with one-to-one resolution. This is beneficial to label data with the same resolution of the original input data [6]. As with most of the CNN models, the encoder network of the SegNet

performs convolution with a filter bank to produce feature maps, batch normalization to increase the stability of the network, and accelerate training [28]. Then, rectified linear unit (ReLU) is applied as an activation function, and max-pooling with a 2×2 window and stride 2 is performed.

The number of convolutional kernels (i.e., filters) is from 64 to 512 with a size of 3×3 . The decoder network upsamples the feature maps using max-pooling indices from the corresponding encoder feature maps. The feature maps are convolved with a trainable decoder filter bank to produce dense feature maps. Then, batch normalization is applied to each feature map and then fed to the softmax classifier at the output layer of the decoder (see Figure 7). Softmax is used to map the nonnormalized output of the network to a probability distribution over predicted classes [29].

2.6. SegNet Model Training. Partitioned patches (or tiles) of the multisource dataset (i.e., IR orthoimage, DSM, slope, aspect, and multidirectional SRIs) generated by partitioning are fed into the SegNet model for training and evaluation. The numbers of the input patches of each data are 4870, 1730, and 1390 for training, evaluation, and test, respectively. Most of the current DL models require a trial-and-error method to achieve optimal solutions. This approach is not

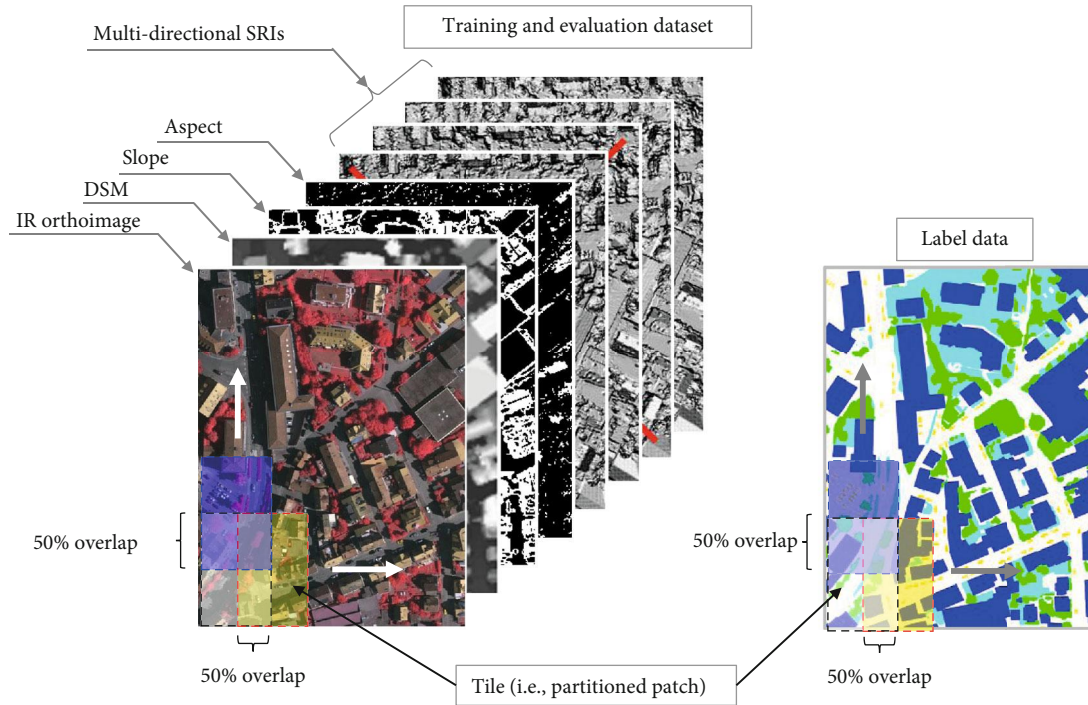


FIGURE 6: Preparation of training, evaluation, and label dataset by partitioning with 50% overlap.

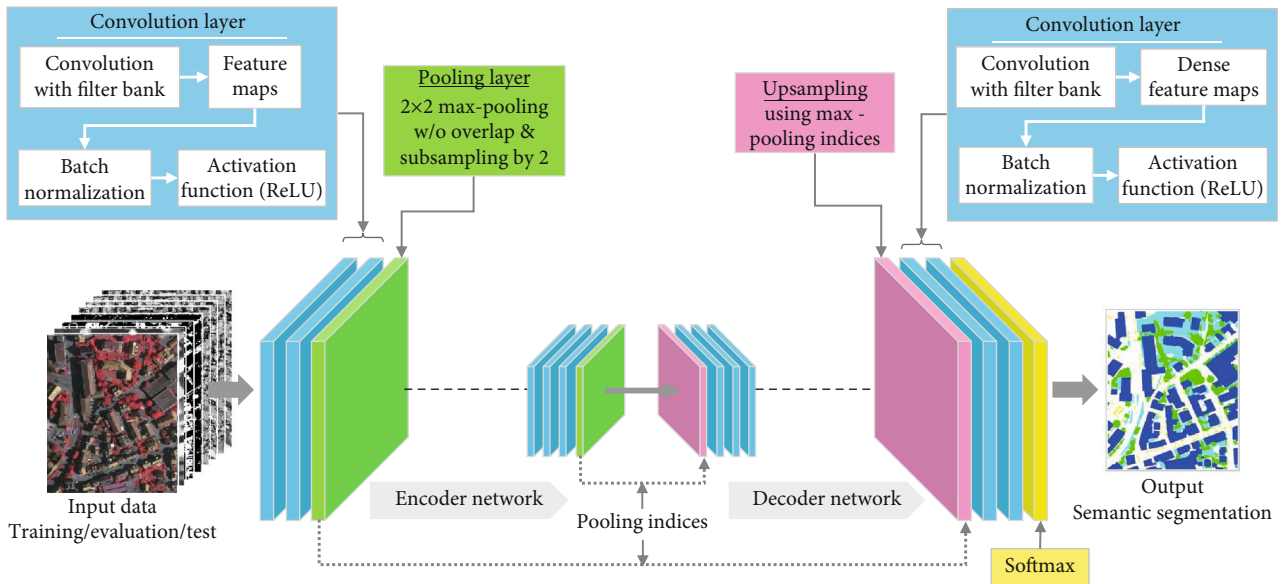


FIGURE 7: SegNet model architecture. Series of dots represent convolutional layer groups.

trivial and takes ample training time because diverse parameters are involved in the DL model. The combination of different parameter setting leads to tremendous numbers of cases. Therefore, transfer learning was adopted to determine the hyper-parameters including learning rate, mini-batch size, and number of the epochs. Another consideration is normalized DSM (nDSM). The nDSM is obtained by subtracting bare-ground height (i.e., digital terrain model (DTM)) from the DSM as Equation (12):

$$nDSM = DSM - DTM. \quad (12)$$

Many DL models utilize nDSM instead of DSM [6, 9, 30]. The main reason to use nDSM is that the nDSM only reflects relative heights of the objects regardless of the terrain elevation. In other words, differences among objects in elevation are taken into account in nDSM [31]. DTM has to be available to create nDSM (e.g., contours from topographic maps) or complicated filtering process to separate ground and

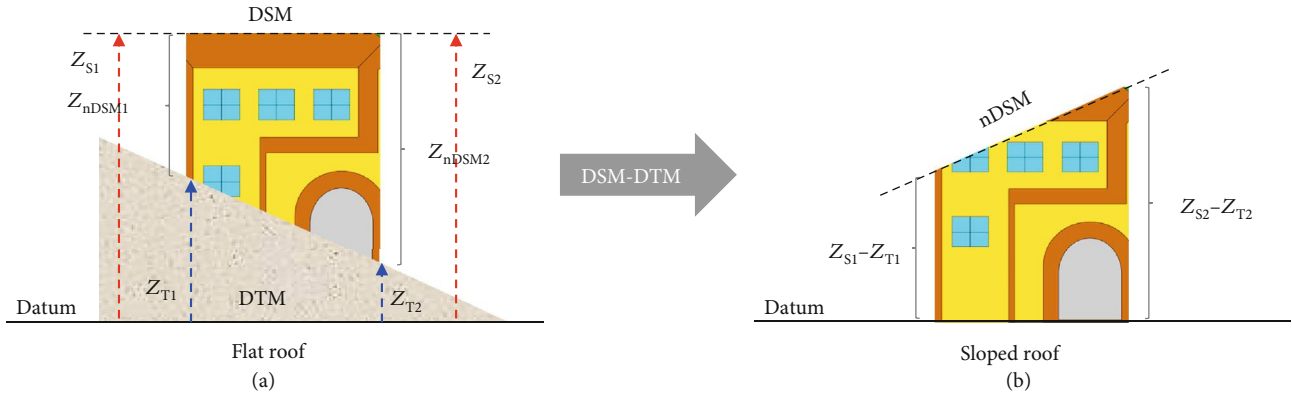


FIGURE 8: nDSM changes the original shape of object. (a) DSM. (b) nDSM. Z_T and Z_S denote terrain and object height from datum, respectively.

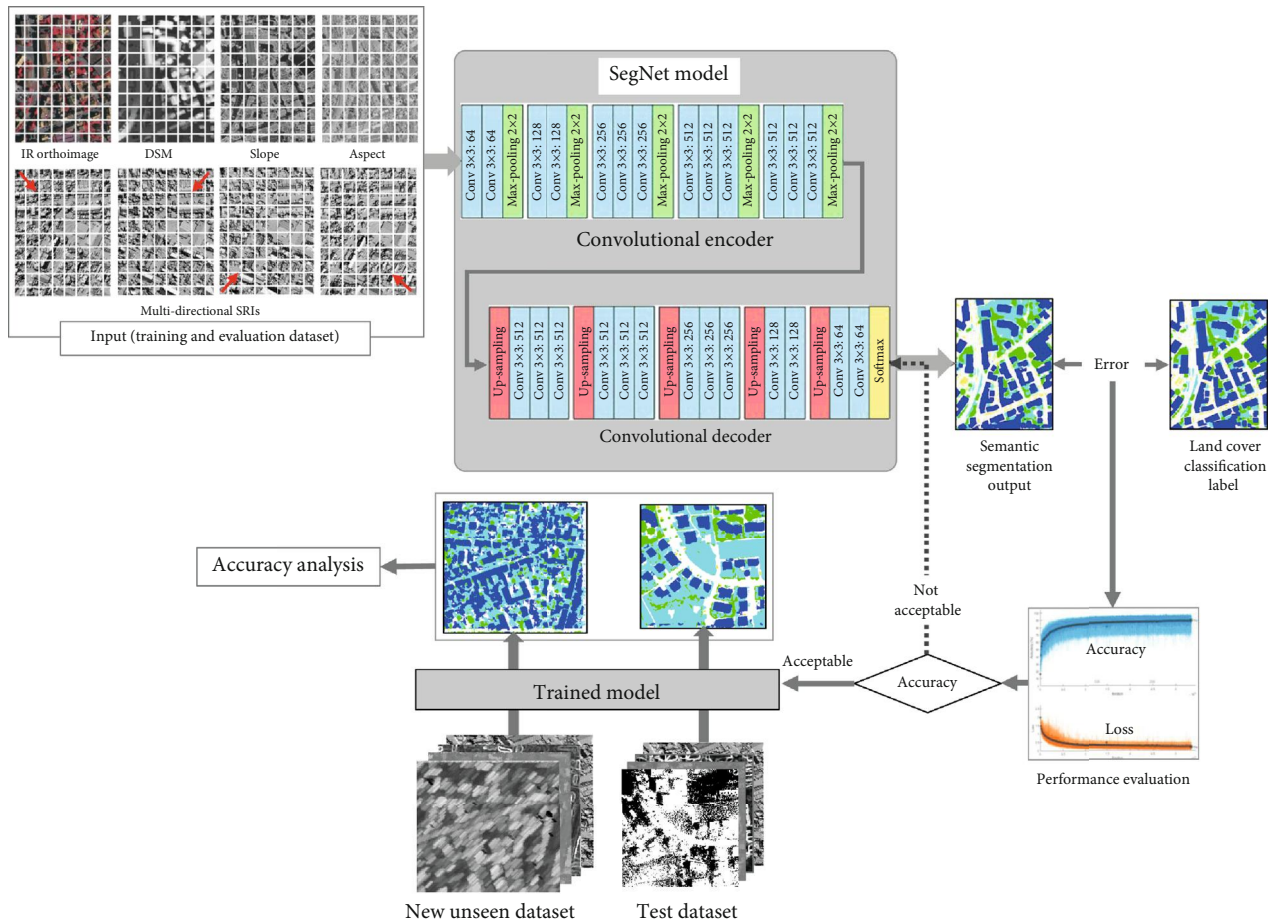


FIGURE 9: Workflow of proposed method for land cover classification.

nonground features (e.g., morphological filtering, progressive densification, surface-based method, and segment-based method) is required [32]. However, the robust DL has to require less preprocessing such as generating nDSM. There might be some controversial issues about using nDSM for object recognition or classification. The ultimate goal of DL is to resemble human intelligence. Human beings are able

to recognize objects regardless of their location in 3D space. Namely, it would be invariant with respect to geometric transformation including shift, rotation, and scale.

nDSM might be a controversial issue in DL. The concept of nDSM is that all objects have to be vertically relocated onto the reference datum before training DL models that utilize DSM. However, nDSM might distort the shape of the objects

TABLE 2: Hyper-parameter values for experiments.

Parameter	Value
Learning rate	0.001
Training cycle (iteration)	334,675
Maximum iteration	608,500
Iteration/epoch	1217/275
Validation frequency	1217
Mini-batch size	4
Shuffle	Every epoch

in some cases. For example, a building with flat roof on a sloped surface is to be a building with a sloped roof if nDSM is applied as shown in Figure 8.

Another issue of the nDSM is for DSM obtained from terrestrial LiDAR data or images of the street scene. If nDSM should be used, then objects have to be moved horizontally to the reference plane. In this case, the horizontal reference plane must be defined. Since DL attempts to resemble human recognition ability, it is much better and robust models that nDSM is not required because human beings recognize objects regardless of objects' location. nDSM should not affect the computation of slope and aspect and creating SRI. In this matter, we did not use nDSM as training but the original DSM. DSM and DSM-derived data (i.e., slope, aspect, and SRI) could provide 3D spatial geometric characteristics that are important information to identify and distinguish different types of objects while images provide only 2D visual information. In this regard, we propose utilizing various data for training the DL model to obtain reliable results. Figure 9 shows the workflow of our proposed method.

2.7. Evaluation Metrics for Performance Measures. Evaluating performance is one of the fundamental tasks in DL. Classification accuracy, in general, can be described as the number of the correct predictions from the total predictions performed. Classification accuracy alone is not sufficient to evaluate the performance of the DL model. Since the ultimate goal of DL is to expand the trained model to other datasets (i.e., new unseen data) that is not involved with training, it is an important issue to evaluate the robustness of a DL model for the new unseen data. The test data has similar characteristics to the training data since the test data, in most cases, belongs to the same region of the training data (i.e., Vaihingen dataset), while the new unseen data has somewhat different characteristics because the new unseen data is selected from a different place (i.e., Potsdam dataset).

Different evaluation criteria have been proposed to assess the quality of the semantic segmentation. Commonly used evaluation metrics for classification are accuracy, precision, recall, F1 score, and intersection-of-union (IoU) [33]. Usually, variations on pixel accuracy and IoU have been used frequently [34]. We applied overall accuracy (Equation (13)) and IoU (Equation (14)) to evaluate semantic segmentation results:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (13)$$

$$\begin{aligned} \text{IoU} &= \frac{\text{Area of overlap}}{\text{Area of union}} \\ &= \frac{(\text{Predicted area}) \cap (\text{Actual area})}{(\text{Predicted area}) \cup (\text{Actual area})} \\ &= \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \end{aligned} \quad (14)$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Accuracy might lead to misinterpretation when the class representation is small within the image, as the measure is biased in mainly reporting on how well in identifying negative cases. On the other hand, IoU is calculated for each class separately, then averaged over all classes to provide mean IoU score of semantic segmentation prediction. The criterion to be correct or not to be is 0.5. If IoU is larger than 0.5 (i.e., 50%), it is normally considered a good prediction. IoU has been used in numerous papers and popular object detection challenges such as ImageNet ILSVRC, Microsoft COCO, and PASCAL VOC.

3. Experimental Results and Discussion

Semantic segmentation results of three test datasets from Vaihingen data (i.e., areas I, K, and O in Figure 1) and four new unseen datasets from Potsdam data (i.e., areas 3_12, 4_10, 4_11, and 7_08 in Figure 2) are presented. The major hyper-parameters for experiments were set as shown in Table 2.

Figures 10 and 11 are IR image and label data of Vaihingen and Potsdam datasets, respectively. Figures 12 and 13 show the classification results of Vaihingen and Potsdam datasets, respectively. Results of the Potsdam data were obtained by using trained model with the Vaihingen dataset. It is obvious that the accuracy of the test data is higher than that of the new unseen data.

Classification accuracies of the training, evaluation, test, and new unseen data are listed in Tables 3 and 4 (Note: evaluation metrics are expressed in percent (%)). The result from combining all data (IoU = 84.90%) for the test data was improved compared with the results from individually trained data. However, multidirectional SRI provided better results (IoU = 66.27%) than combining all data (IoU = 52.45%) for new unseen data. We expect that multidirectional SRI has a relatively larger contribution to the model. Even though IoU from combining all data is lower than that from multidirectional SRI, in general, IoU with higher than 0.5 (i.e., 50%) is considered a successful classification. Since man-made features such as buildings have distinctive geometric characteristics compared to natural features, training DL model using DSM-derived data is particularly effective to buildings. Therefore, the proposed approach might be feasible to identify and extract buildings for the further application of 3D building modeling.

The evaluation metrics could not provide information on how well ground truth and predicted classes for individual

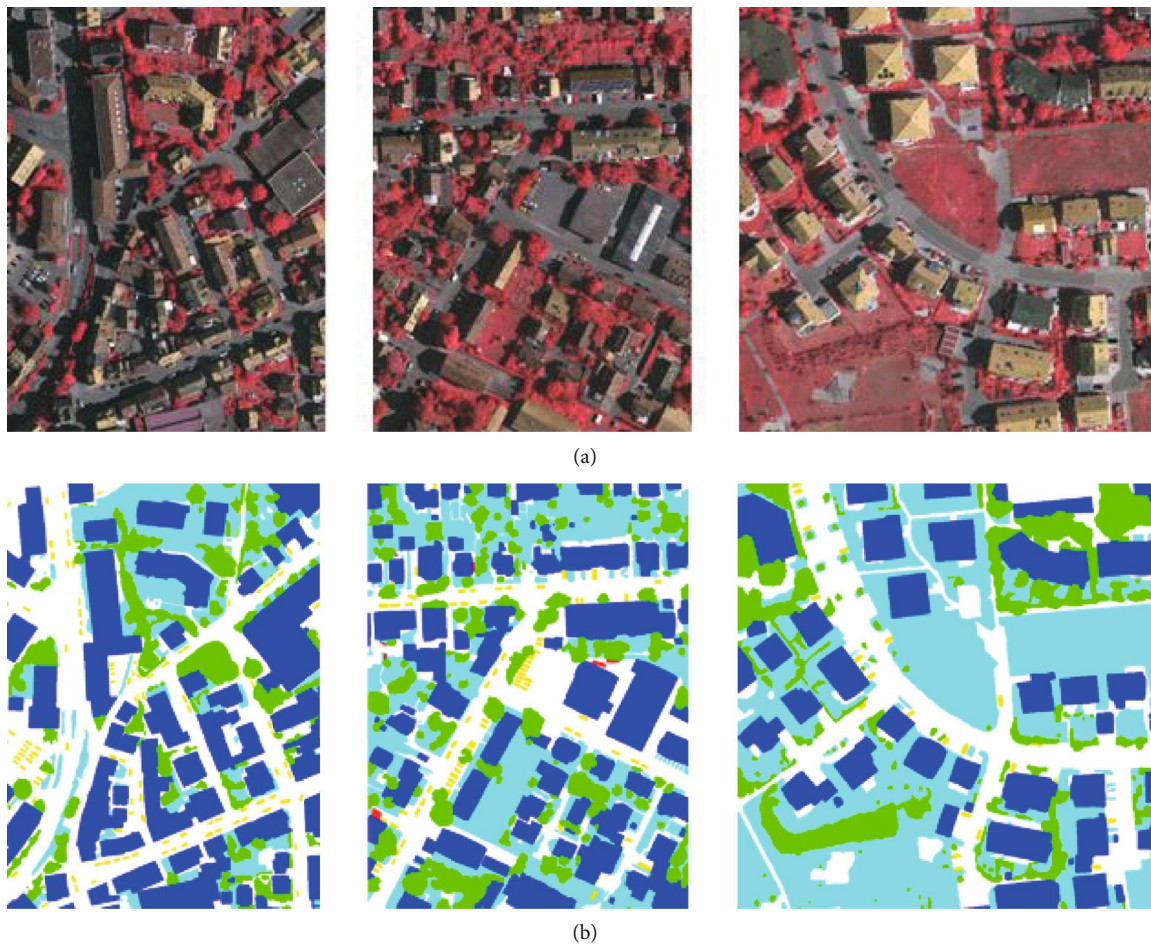


FIGURE 10: Vaihingen test area. (a) IR images. (b) Label data (ground truth).

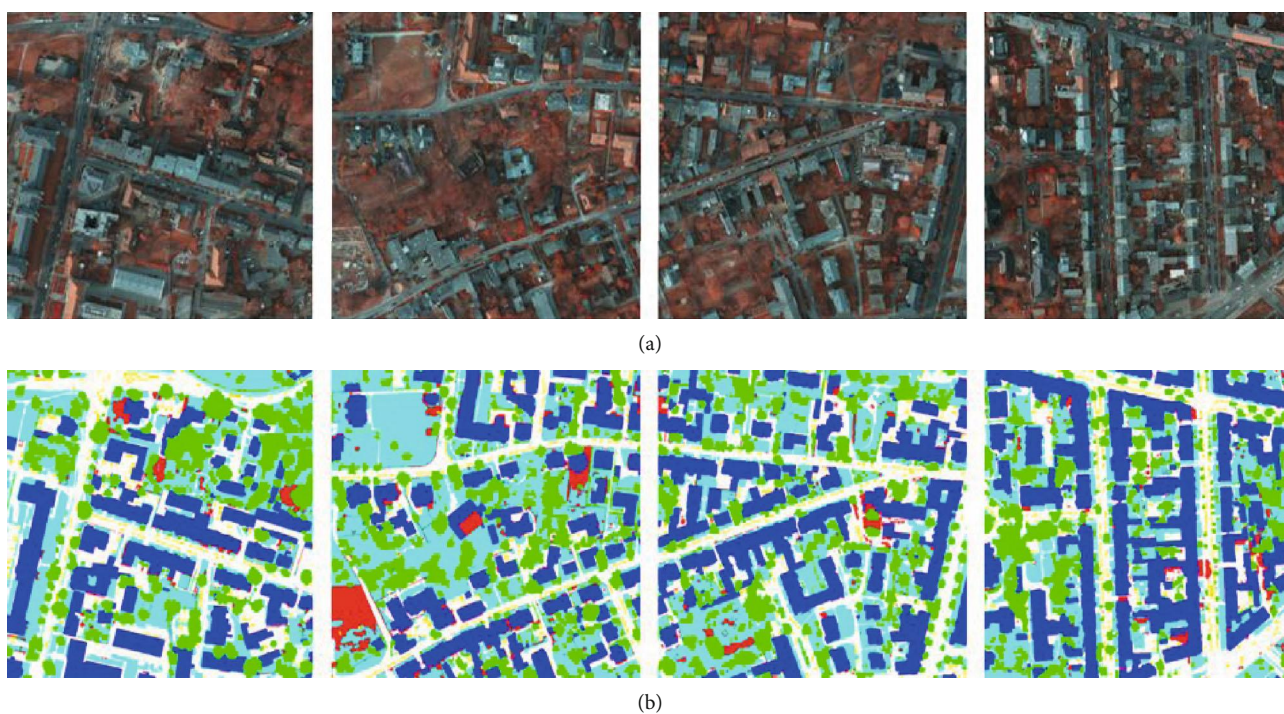


FIGURE 11: Potsdam new unseen data. (a) IR images. (b) Label data (ground truth).

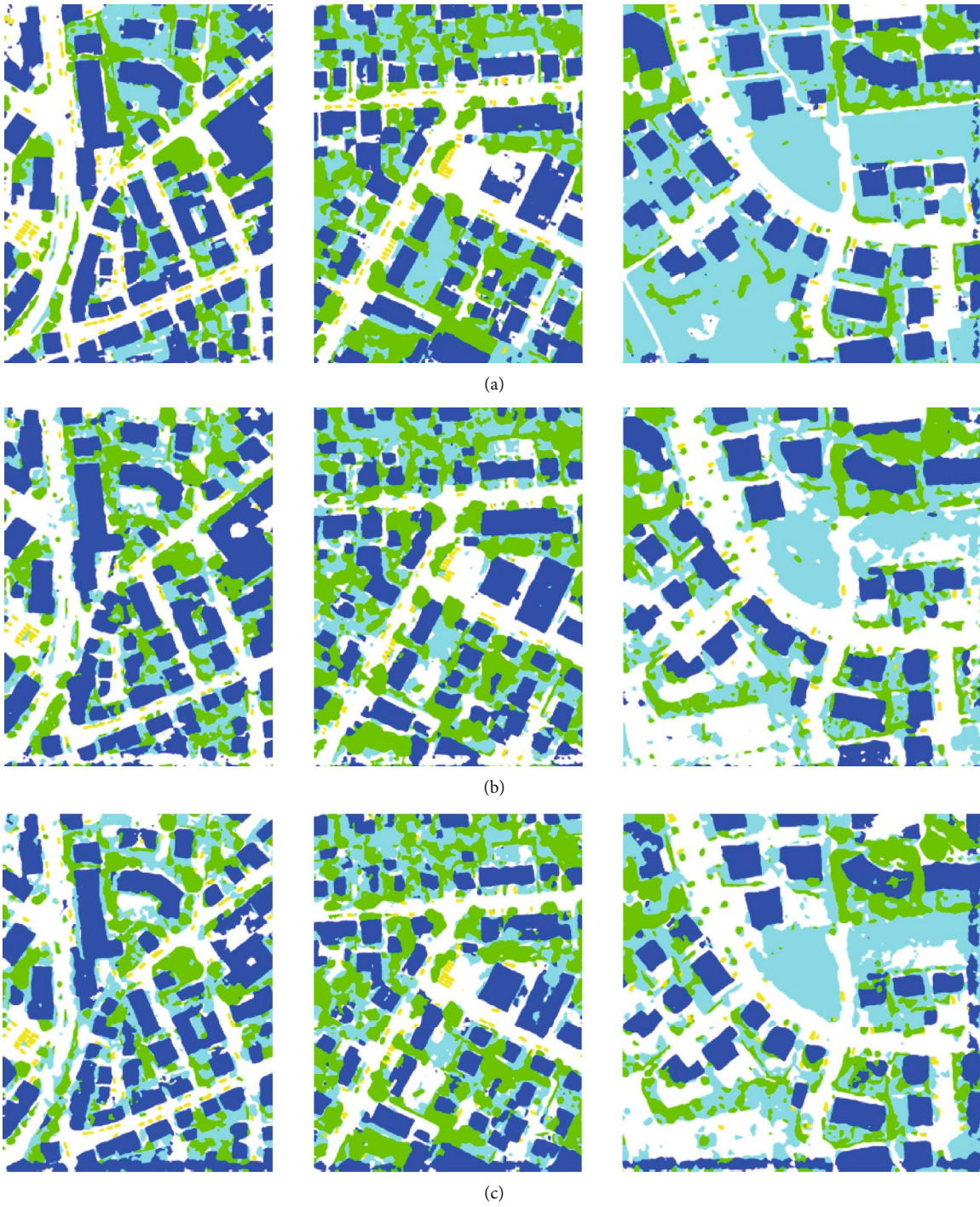


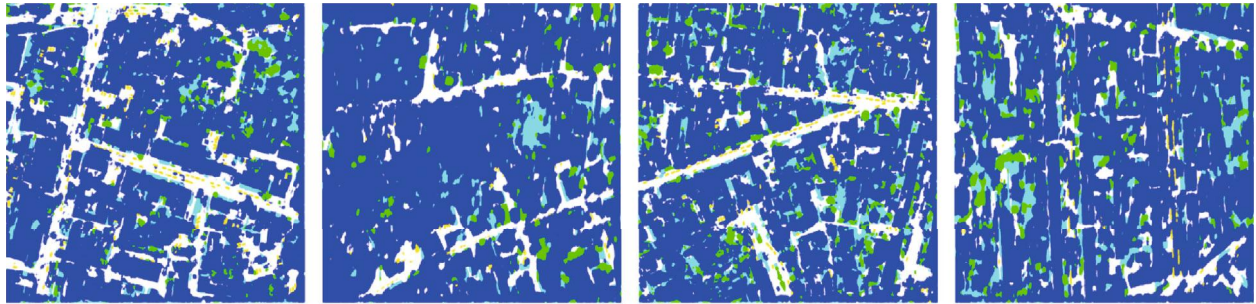
FIGURE 12: Continued.



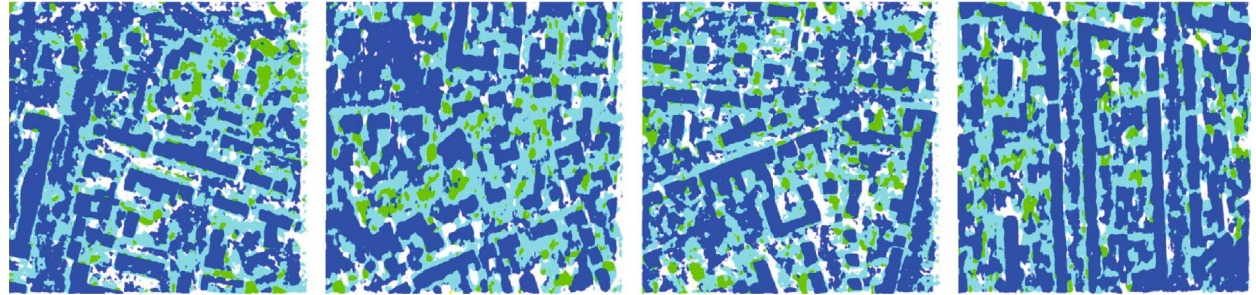
FIGURE 12: Classification results of Vaihingen test data. (a) IR images. (b) DSM. (c) Slope. (d) Aspect. (e) Multidirectional SRI. (f) Combining all data.

objects matched. In this matter, differences between each result and label data were depicted for visual evaluation. White and black indicate correctly classified and misclassified pixels, respectively (see Figures 14 and 15). Interpolation of the DSM might cause delocalization and zagging effect, especially on the objects' boundaries. Thus, most of the misclassified pixels are found along the boundaries of the objects.

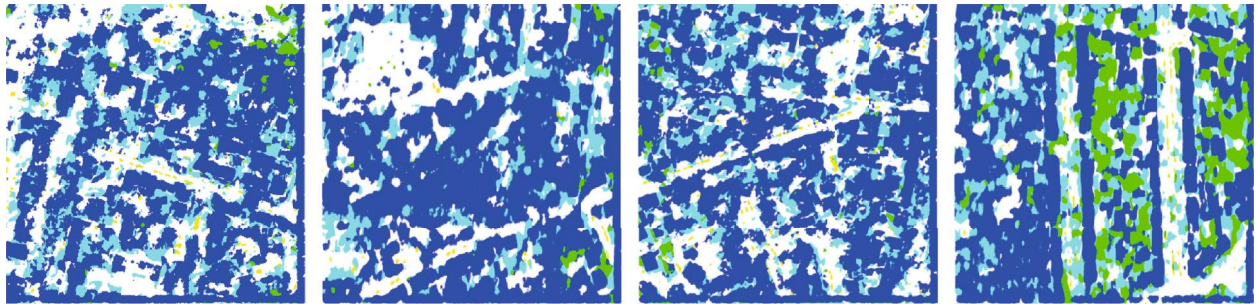
For the Vaihingen test data, some of the low vegetation areas were classified as impervious surfaces (i.e., roads) with DSM and DSM-derived data (i.e., slope, aspect, and SRI). However, training the DL model using all data could improve classification results (see Figure 14(f)). As for the Potsdam new unseen data, more misclassified regions were found except for the case of buildings. Similar to the Vaihingen test



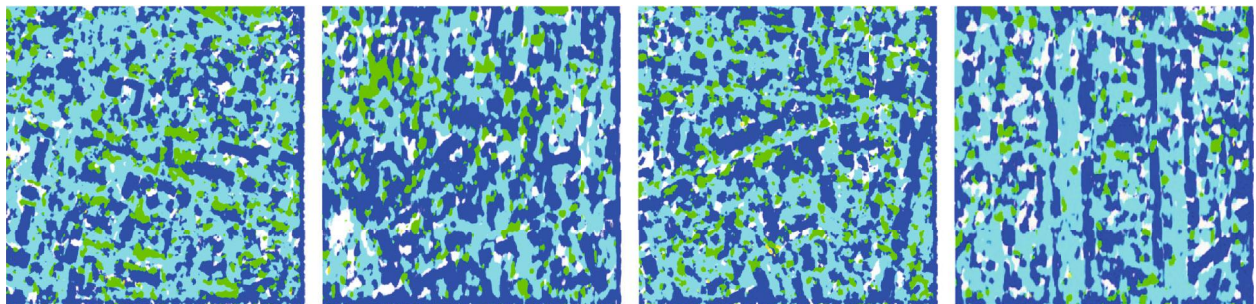
(a)



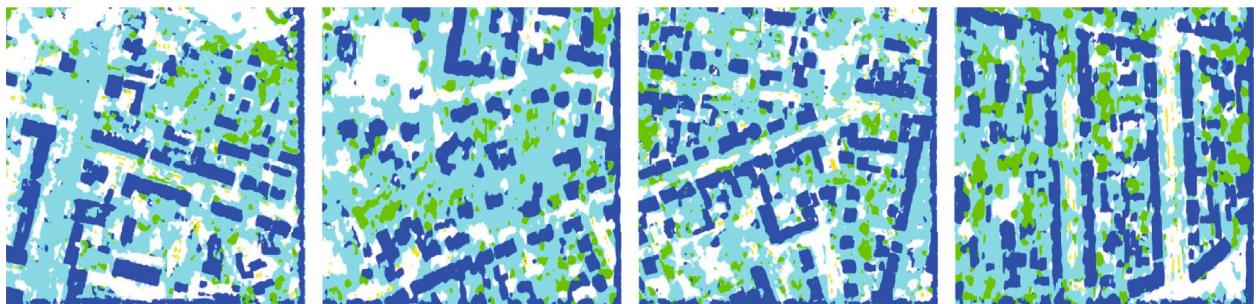
(b)



(c)



(d)



(e)

FIGURE 13: Continued.

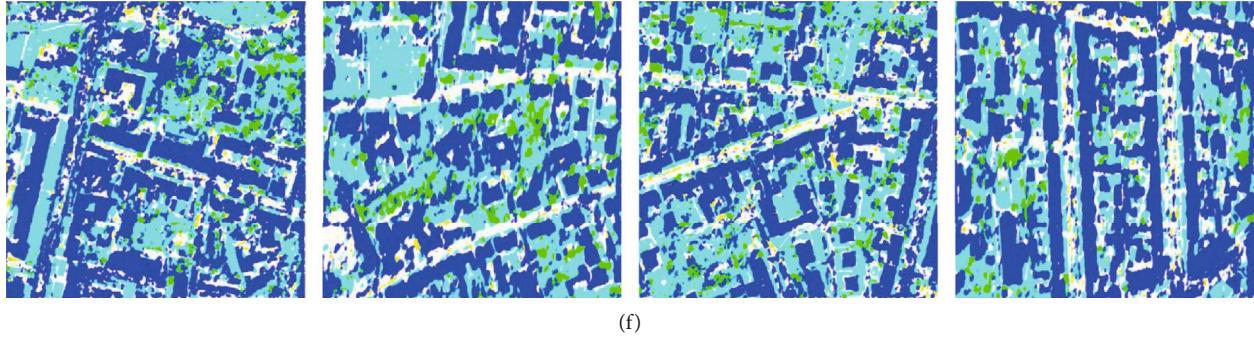


FIGURE 13: Classification results of Potsdam new unseen data. (a) IR images. (b) DSM. (c) Slope. (d) Aspect. (e) Multidirectional SRI. (f) Combining all data.

TABLE 3: Classification evaluation of Vaihingen dataset for training, evaluation, and test.

Dataset	Training accuracy (%)	Evaluation accuracy (%)	Test data (%)			
			Overall Accuracy	IoU	Building Accuracy	IoU
IR image	97.74	91.49	83.71	57.98	89.34	84.08
DSM	91.97	85.25	69.90	41.32	83.88	77.66
Slope	94.42	88.41	67.93	40.31	78.59	71.28
Aspect	95.87	89.07	64.55	35.90	79.01	70.58
SRI	93.50	89.37	70.00	42.59	83.74	75.90
All data	91.36	90.61	83.56	55.75	89.45	84.90

TABLE 4: Classification evaluation of Potsdam dataset for new unseen data.

Dataset	New unseen data (%)			
	Overall		Building	
	Accuracy	IoU	Accuracy	IoU
IR image	45.56	23.34	94.24	35.95
DSM	43.86	17.66	82.71	44.89
Slope	41.09	18.11	77.68	37.68
Aspect	32.97	11.68	61.80	34.24
SRI	52.53	28.28	77.92	66.27
All data	58.86	30.75	89.37	52.45

data, a training model using a combination of all data provided better results (see Figure 15(f)). Test data might be insufficient to evaluate the performance of the trained DL models particularly to investigate the possibility of universal use. For this reason, we applied not only test data but also new unseen data to the trained model; as a result, some meaningful results were confirmed. Utilizing multisource data and derived information that are appropriate to represent characteristics of the objects could improve semantic segmentation. DSM and DSM-derived information could play an important role in recognizing buildings.

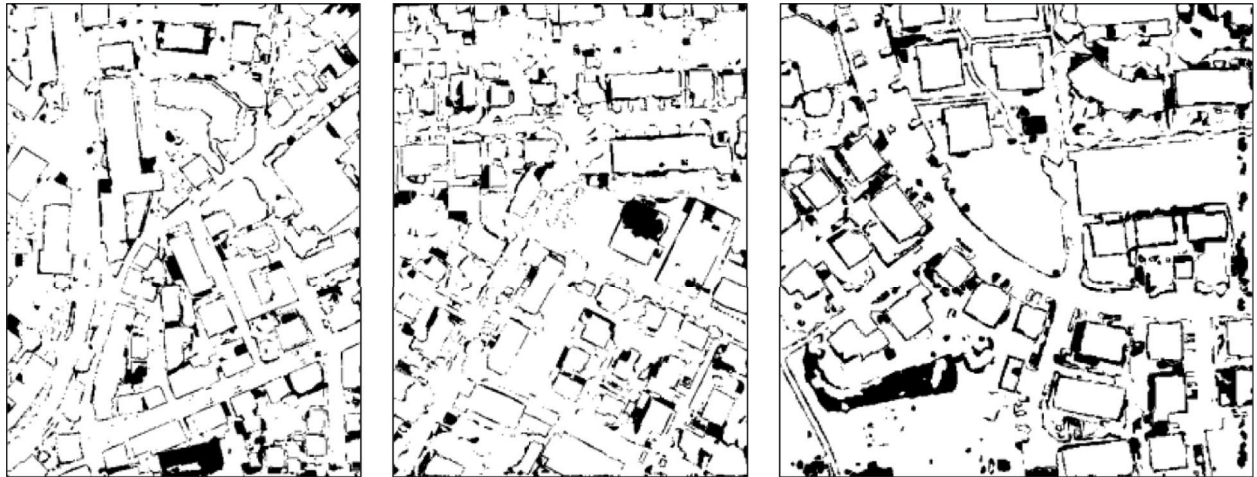
Classification results from training all data (i.e., IR image, DSM, slope, aspect, and SRI) are better than results from training individual data. The experiments show that DSM and its derived information could be appropriate to recognize man-made objects including buildings because DSM

and DSM-derived information have explicit geometric characteristics. In consequence, the classification accuracy of the building is much higher than that of other objects. Figures 16–18 show buildings of the label data and corresponding classification results from using all data of Vaihingen and Potsdam areas, respectively. It is noticeable that buildings in the test data were well identified, while buildings in the new unseen data were less accurately identified. A considerable number of cars of the Potsdam data were identified as buildings; however, most of the major buildings are correctly classified.

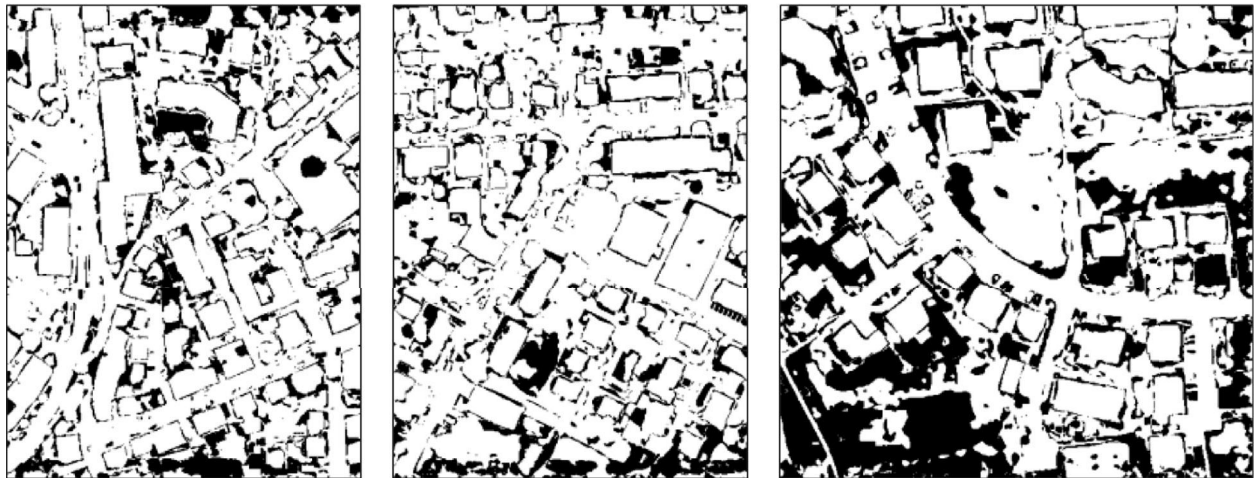
Some conventional image processing technique (e.g., morphological filtering with erosion and dilation [35]) might improve the results. Figure 18 illustrates an example of applying morphological filtering to the building class from Potsdam data. Therefore, it is recommendable to consider postprocessing if necessary.

4. Conclusions

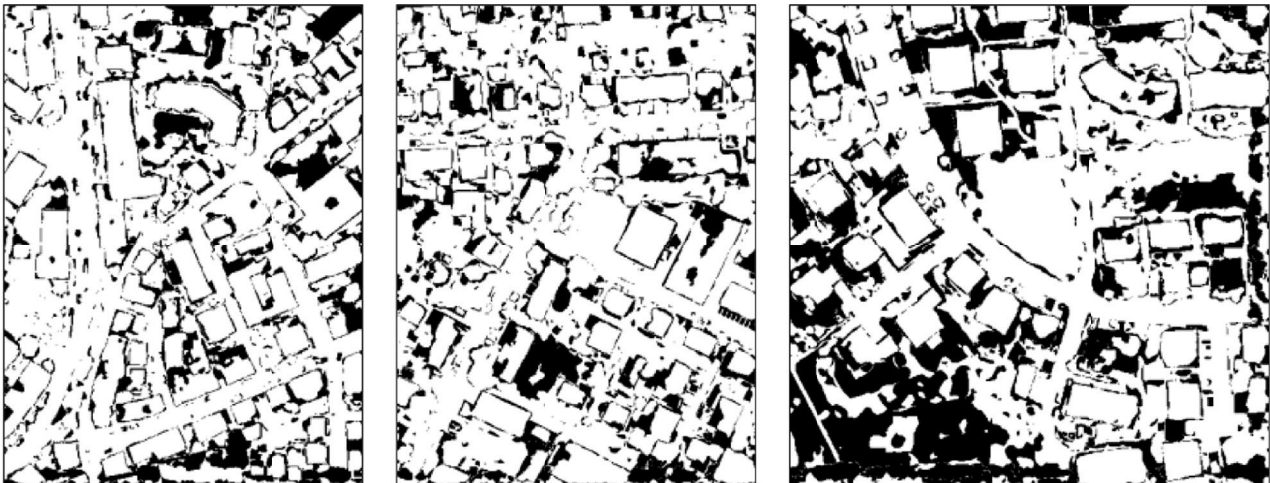
Human beings have the ability to integrate various visual cues (i.e., shape, size, color, tone, texture, depth, stereopsis, and motion) based on knowledge, experience, and innate cognition for visual perception. Semantic segmentation or classification involves object detection, recognition, and identification by utilizing various information. In consequence, it is not sufficient to successfully carry out the semantic segmentation of real-world objects by one type of sensory data. Every object has its own unique physical and geometrical characteristics. Such intrinsic characteristics could be obtained partially from imagery, 3D data, and derived



(a)



(b)



(c)

FIGURE 14: Continued.

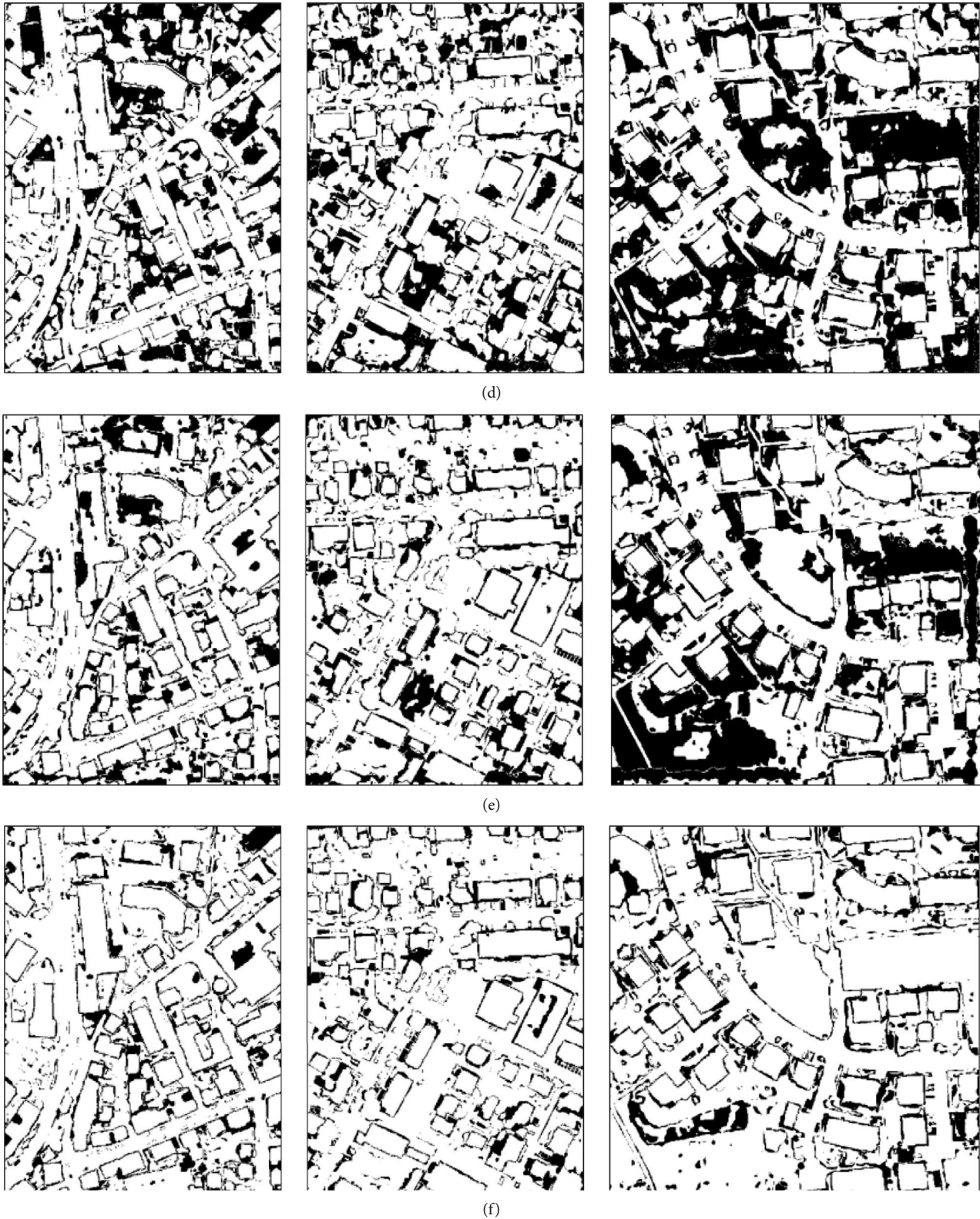


FIGURE 14: Differences between classification results and label data of Vaihingen test data. (a) IR image. (b) DSM. (c) Slope. (d) Aspect. (e) Multidirectional SRI. (f) Combining all data.

information from various sources. Therefore, one of the key tasks is to utilize multisource data for complementary effects to extract characteristics of the objects from the data.

In this paper, we proposed a DL model that utilizes multisource data including optical IR image, DSM, and DSM-derived slope, aspect, and multidirectional SRIs for semantic

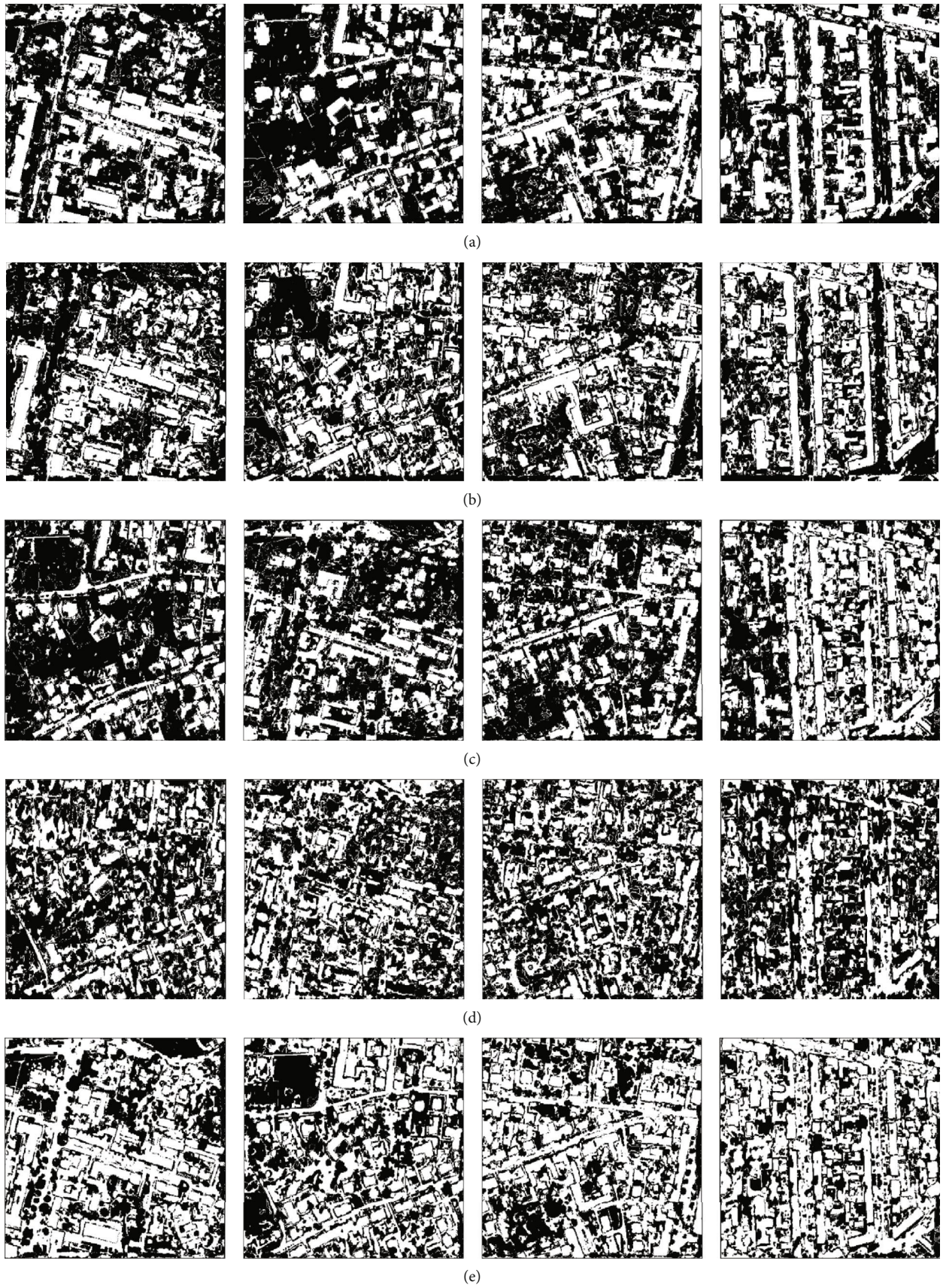
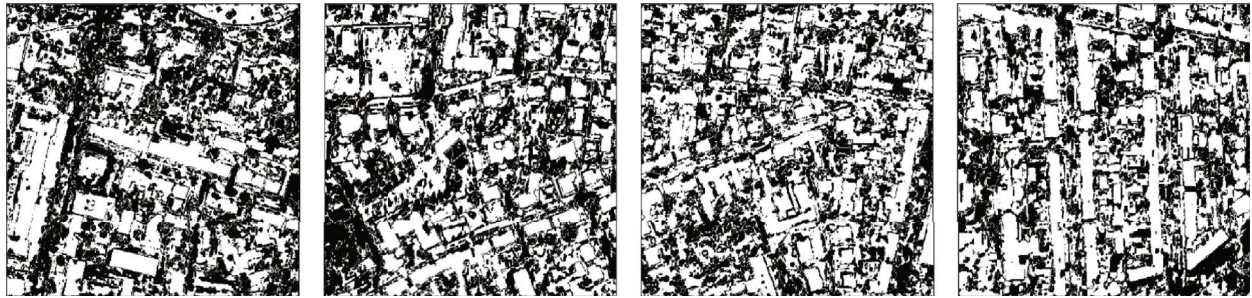
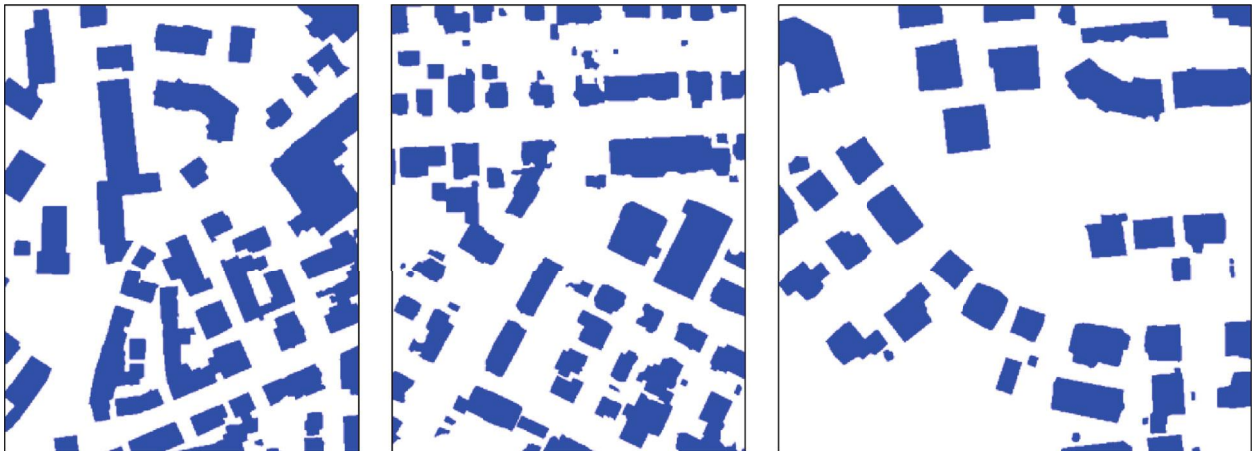


FIGURE 15: Continued.

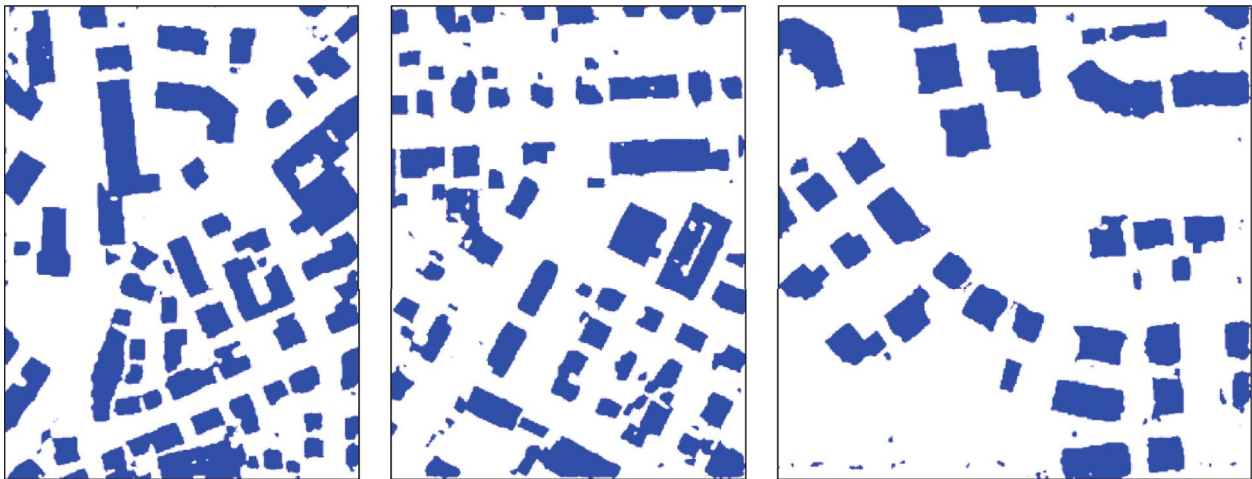


(f)

FIGURE 15: Differences between classification results and label data of Potsdam new unseen data. (a) IR image. (b) DSM. (c) Slope. (d) Aspect. (e) Multidirectional SRI. (f) Combining all data.



(a)



(b)

FIGURE 16: Building class of Vaihingen test data. (a) Label data. (b) Results from combining all data.

segmentation to classify land cover features. The dataset used for training the DL model are based on the DSM that has distinct 3D geometric characteristics. Therefore, specific objects (i.e., buildings) are identified well dominantly. Nevertheless, overall results from the new unseen data were not quite satisfied; buildings were adequately identified compared to the other objects due to the DSM

with derived information. Each object has its own unique characteristics with various aspects. The key task of DL is to reveal representative features from the objects during training. Therefore, training by utilizing various types of data that are suitable to specific objects could improve the performance of DL. In particular, DSM-derived data would be helpful to identify buildings.

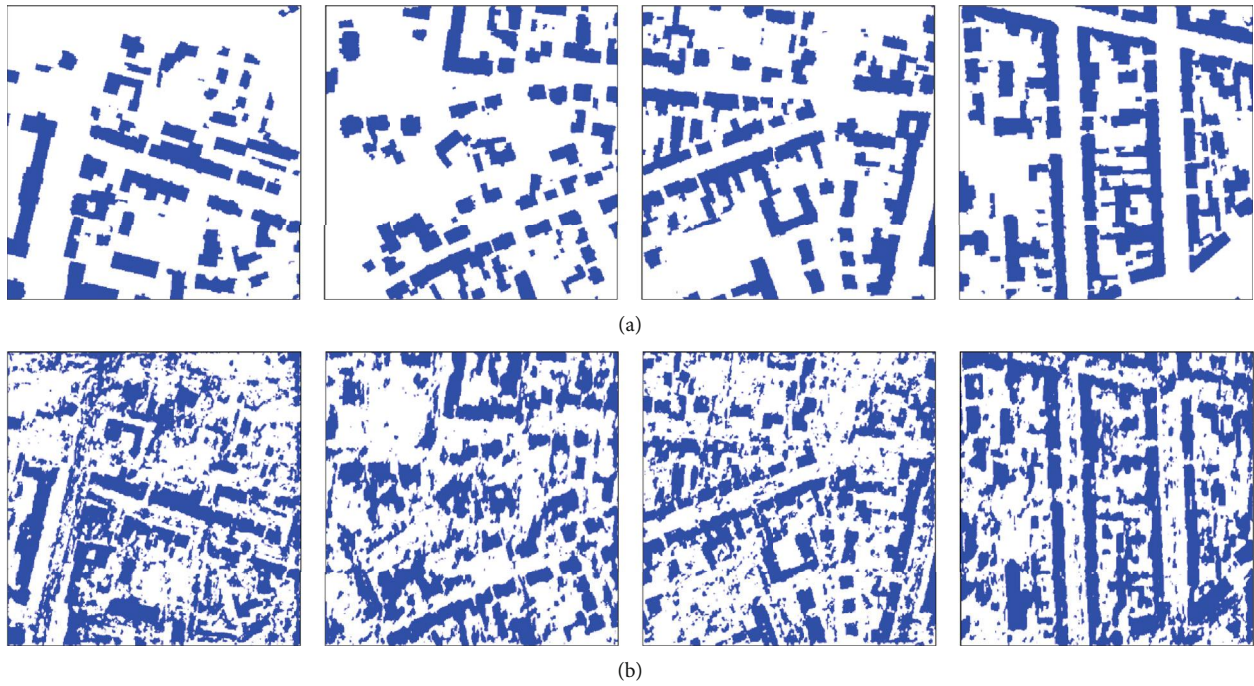


FIGURE 17: Building class of Potsdam new unseen data. (a) Label data. (b) Results from combining all data.

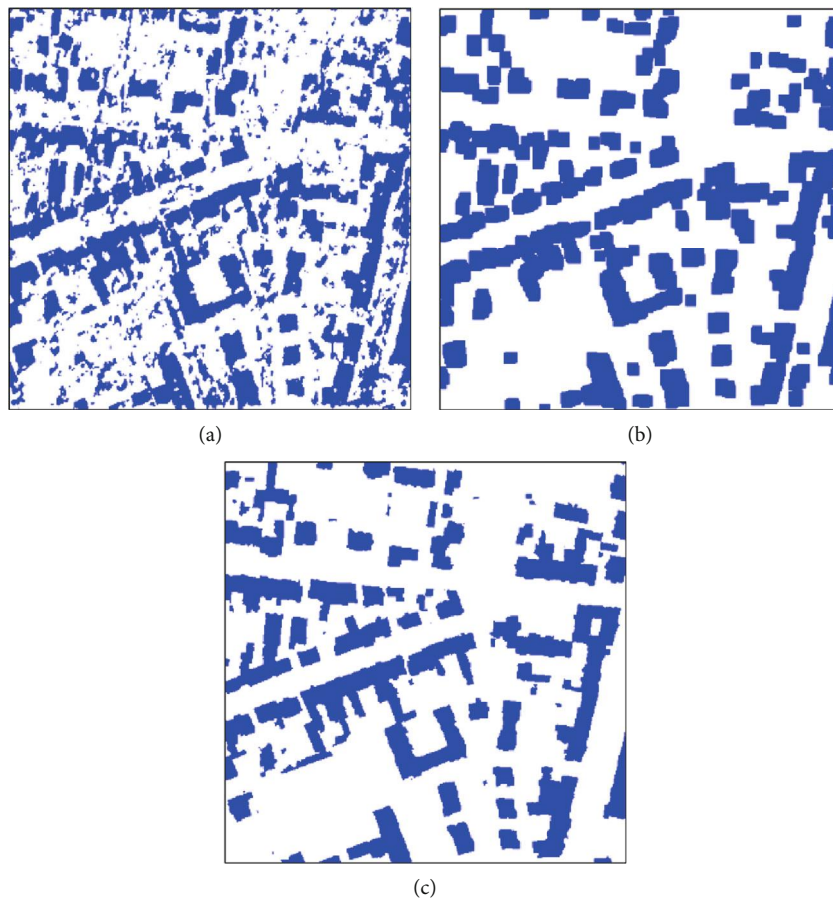


FIGURE 18: Comparison of (a) classification result, (b) morphological filtering, and (c) label data of Potsdam new unseen data.

In general, the scalability of the trained DL models is evaluated using test data. The problem of using the test data is that, in most cases, the test data belongs to the same area with similar properties as the training data. Therefore, we evaluated the model using a dataset from different areas, i.e., test data from the Vaihingen area, and new unseen data from the Potsdam area. We found some meaningful results from the test data and new unseen data. Combining all data yielded the highest IoU of the building for the test data (i.e., 84.90%). On the other hand, multidirectional SRI provided the highest IoU of the building (i.e., 66.27%) while combining all data yielded IoU of 52.45%, but it was higher than individually trained data. We expect that multidirectional SRI plays an important role for training DL model.

The ultimate goal of DL is to expand trained models for universal use. However, the major problem of DL is data dependency. Geospatial data has a variety of regional properties; hence, it is a valuable task to provide pretrained models suitable for the specific regions (e.g., urban, residential, mountainous, and agricultural areas). In addition, flexible transfer learning that could avoid training DL models from scratch would make DL a more powerful tool in various fields. In addition, the integration strategy of various combinations of the multisource dataset might maximize the synergistic effect, for example, applying different priorities or weights for each dataset.

Data Availability

The datasets generated during the current study are available from the corresponding author on reasonable request. Vaihingen datasets are available in <http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html>.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1D1A1B07048732). The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) [21]: <http://www.ifp.uni-stuttgart.de/dgpf/DKEPAllg.html>.

References

- [1] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
- [2] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: a survey," 2020, <https://arxiv.org/abs/2001.05566>.
- [3] D. Turcsány, *Deep Learning Models of Biological Visual Information Processing*, Doctor of Philosophy theses, The University of Nottingham, 2016.
- [4] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: a meta-analysis and review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, pp. 166–177, 2019.
- [5] E. Maltezos, A. Doulamis, N. Doulamis, and C. Ioannidis, "Building extraction from LiDAR data applying deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 1, pp. 155–159, 2019.
- [6] N. Audebert, B. le Saux, and S. Lefèvre, "Beyond RGB: very high resolution urban remote sensing with multimodal deep networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 140, pp. 20–32, 2018.
- [7] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.
- [8] F. H. Nahhas, H. Z. M. Shafri, M. I. Sameen, B. Pradhan, and S. Mansor, "Deep learning approach for building detection using LiDAR - orthophoto fusion," *Journal of Sensors*, vol. 2018, Article ID 7212307, 12 pages, 2018.
- [9] K. Zhou, D. Ming, X. Lv, J. Fang, and M. Wang, "CNN-based land cover classification combining stratified segmentation and fusion of point cloud and very high-spatial resolution remote sensing image data," *Remote Sensing*, vol. 11, no. 17, p. 2065, 2019.
- [10] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [11] F. Alidoost and H. Arefi, "Knowledge based 3D building model recognition using convolutional neural networks from LIDAR and aerial imageries," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B3, pp. 833–840, 2016.
- [12] L. Pibre, M. Chaumont, G. Subsol, D. Ienco, and M. Derras, "How to deal with multi-source data for tree detection based on deep learning," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1150–1154, Montreal, QC, Canada, November 2017.
- [13] P. Bodani, K. Shreshtha, and S. Sharma, "OrthoSeg: a deep multimodal CONVOLUTIONAL neural network architecture for semantic segmentation of orthoimagery," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-5, pp. 621–628, 2018.
- [14] O. Russakovsky, J. Deng, H. Su et al., "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, Boston, MA, USA, June 2015.
- [16] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (ICCV)*, pp. 1520–1528, Santiago, Chile, December 2015.
- [17] D. Lee, E. Cho, and D. C. Lee, "Semantic classification of DSM using convolutional neural network based deep learning," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 37, no. 6, pp. 435–444, 2019.
- [18] D. C. Lee, *An Adaptive Texture Segmentation Approach for Applications in Digital Photogrammetry*, Ph.D. Dissertation, The Ohio State University, 1997.

- [19] F. Rottensteiner, G. Sohn, J. Jung et al., "The ISPRS benchmark on urban object classification and 3D building reconstruction," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-3, pp. 293–298, 2012.
- [20] M. Cramer, "The DGPF-test on digital airborne camera evaluation - overview and test design," *Photogrammetrie - Fernerkundung - Geoinformation*, vol. 2010, no. 2, pp. 73–82, 2010.
- [21] C. Lemaire, "Aspects of the DSM production with high resolution images," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37, Part B4, pp. 1143–1146, 2008.
- [22] P. A. Burrough, *Principles of geographical information systems for land resources assessment*, Clarendon Press, 1988.
- [23] B. K. P. Horn, "Hill shading and the reflectance map," *Proceedings of the IEEE*, vol. 69, no. 1, pp. 14–47, 1981.
- [24] V. Bruce and P. R. Green, *Visual perception: physiology, psychology and ecology*, Lawrence Erlbaum Associates, 1990.
- [25] R. Szeliski, *Computer vision: algorithms and applications*, Springer-Verlag, 2011.
- [26] D. C. Lee, D. H. Lee, and D. G. Lee, "Determination of building model key points using multidirectional shaded relief images generated from airborne LiDAR data," *Journal of Sensors*, vol. 2019, Article ID 2985014, 19 pages, 2019.
- [27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, <https://arxiv.org/abs/1409.1556>.
- [28] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, pp. 448–456, Lille, France, 2015.
- [29] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," 2016, <https://arxiv.org/abs/1511.00561v3>.
- [30] M. Gerke, T. Speldekamp, C. Fries, and C. Gevaert, "Automatic semantic labelling of urban areas using a rule-based approach and realized with MeVisLab," 2015.
- [31] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.
- [32] C. M. Gevaert, C. Persello, F. Nex, and G. Vosselman, "A deep learning approach to DTM extraction from imagery using rule-based training labels," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 142, pp. 106–123, 2018.
- [33] B. Benjdira, A. Ammar, A. Koubaa, and K. Ouni, "Data-efficient domain adaptation for semantic segmentation of aerial imagery using generative adversarial networks," *Applied Sciences*, vol. 10, no. 3, article 1092, 2020.
- [34] A. Garcia-Garcia, S. Orts-Escolano, S. O. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," 2017, <https://arxiv.org/abs/1704.06857v1>.
- [35] A. Hornberg, *Handbook of machine vision*, Wiley-Verlag, 2006.

Research Article

Pre- and Postcollapse Ground Deformation Revealed by SAR Interferometry: A Case Study of Foshan (China) Ground Collapse

Wu Zhu ¹, Yan Zhang,¹ Zhanke Liu ², and Qian Zhu¹

¹College of Geology Engineering and Geomatics, Chang'an University, Xi'an, China

²The First Geodetic Surveying Brigade of Ministry of Natural Resources, Xi'an, China

Correspondence should be addressed to Wu Zhu; zhuwu@chd.edu.cn and Zhanke Liu; lzk_111@163.com

Received 26 May 2020; Revised 8 August 2020; Accepted 19 August 2020; Published 10 September 2020

Academic Editor: Sang-Hoon Hong

Copyright © 2020 Wu Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

On the evening of 7 February 2018, a deadly collapse of a metro tunnel under construction in the Southern China city of Foshan caused 11 deaths, 8 injuries, and 1 missing person. For disaster prevention and mitigation, the spatiotemporal ground deformations before and after the collapse event were derived from 55 Sentinel-1A synthetic aperture radar (SAR) images spanning from March 2017 to January 2019. The results showed that prominent ground subsidence in the shape of a funnel with a maximum rate of 42 mm/year was observed in the vicinity of the collapse area before the accident. After the accident, the area and magnitude of subsidence decreased compared with precollapse subsidence. This decrease is related to the progress of tunnel excavation and groundwater changes. In the temporal domain, continuous subsidence was observed over a year before and after the accident, and accelerated subsidence appeared one month before the collapse accident. Soft soil consolidation and tunnel-induced soil losses were the main reasons for the subsidence over the study area. The leakage of groundwater accounted for the collapse event. The leaked groundwater eroded the soil, resulting in the formation of an arched hole. The connection between the arched hole and the tunnel reduced the bearing capacity of the soil layer above the arched hole, triggering the collapse event. The findings provide scientific evidence for future collapse monitoring and early warning due to tunnel excavation.

1. Introduction

As an advanced space observation technique, synthetic aperture radar (SAR) interferometry has demonstrated advantages for monitoring the ground deformation for collapse sinkholes compared with some ground-based instruments and techniques, such as high-precision leveling [1], robotic total stations [2], global positioning systems [3], and terrestrial laser scanner [4]. The sinkholes along the Dead Sea coast in Israel and Jordan have been successfully identified and monitored by SAR interferometry since 2002, providing rich references and experiences for studying the deformation of collapse sinkholes [5–11]. The continuous subsidence of sinkholes in Wink, Texas, USA, was observed from time-series SAR interferometry, suggesting that continuous monitoring of subsidence in the vicinity of sinkholes could help prevent and mitigate catastrophic outcomes [12–14]. A dramatic increase in vertical deformation in the last few years preceding the sinkhole formation at a shopping mall in Heer-

len, the Netherlands, was recorded in SAR data between 1992 and 2011, showing the feasibility of using satellite radar interferometry to detect a migrating cavity [15]. Similar techniques were applied to the collapse sinkhole in the Ebro Valley, Spain [16]; South Africa [17]; Poland [18]; and Arizona, USA [19].

On the evening of 7 February 2018, a deadly ground collapse of a metro tunnel under construction occurred in Foshan, Guangdong Province, China. The accident caused 11 deaths, 8 injuries, 1 missing person, and a direct economic loss of more than 53 million yuan (USA \$7.8 million) [20]. The metro construction is an east–west line including 12 underground stations and 5 elevated stations. The project has 5 shield tunnel intervals, and some of the intervals are constructed in strata that feature rich underground water and complex geological conditions. The collapse area is part of the shield tunnel interval between Huchong Station and Lvdaohu Station, which is 1804 m long and 17.2–34.4 m deep [20]. After the collapse event, a sinkhole 81 m long, 65 m



FIGURE 1: Foshan collapse view: aerial (a) and terrestrial (b).

wide, and 8 m deep was formed, as shown in Figure 1. Since the collapse area was under metro construction, the event was attributed to the underground tunnel excavation. The results of a field investigation indicated that the accident was due to the groundwater leakage, which let the water pour into the tunnel and triggered the collapse [21]. Based on this investigation, the leakage of groundwater was the crucial factor causing the collapse event. Accordingly, several researchers analyzed the reasons for the groundwater leakage, such as complex geological strata [21], surface loading [22], and the leakage at the tail of the shield machine [23]. For the ground deformation, existing research focused on the whole city and there was no information available for deformation monitoring and analysis regarding to the collapse area [22, 24, 25].

To better understand the Foshan collapse event, multi-temporal SAR interferometry was used in this study to obtain the spatiotemporally evolution of the ground deformation in the vicinity of the collapse sinkhole. For this, 55 C-band Sentinel-1A images were acquired in an ascending orbit. The annual deformation velocity was estimated to characterize the spatial pattern of deformation over the collapse area. Subsequently, the time-series deformation was interpreted in detail. Finally, the cause of the accident was discussed based on the observed deformation.

2. Study Area and Data Collection

2.1. Geological Setting of the Study Area. The study area is located at Foshan, which lies in the northern part of the Pearl River, in Southern China. Crossed by the Tanzhou, Dongping, and Jili Rivers, the collapse area is a large alluvial delta, where it is mostly flat with few hills and has an altitude from 1.3 to 4.6 m above sea level [21]. Belonging to the subtropical maritime monsoon climate, the study area is characterized by short spring and autumn, long summer, warm winter, abundant sunshine, and plentiful rainfall [26].

Quaternary strata, with unconsolidated sediments from 0 to 60 m thick, are widespread in the study area. Figure 2(a) shows that the thickness of quaternary strata decreases from

west to east. Figure 2(b), displaying the vertical stratigraphic structure around the constructed tunnel, shows that the stratum from top to bottom is artificial filling soil, silty clay, mucky soil, powdered fine sand, medium sand, gravel, sandstone, silty mudstone, and siltstone [21]. The artificial filling soil with a thickness of 0-5 m is distributed along the road and mainly from the weathered residual soil, weathered rock, and gravel. Below the artificial filling soil, it is distributed by the silty clay characterized by high water content and high compressibility. The experiment indicates that the water content is about 10%-48% and compression modulus is about 3-13 MPa for silty clay over the study area [26]. The mucky soil, distributed below the silty clay, is mainly from the sea alluvial silt and mucky soil, which has properties of high water content, high compressibility, high void ratio, and low strength. For the study area, the mucky soil has the water content of 24.3%-68.8%, compression modulus of 1.6-2.4 MPa, and void ratio of 0.8-2.0 [26]. Powdered fine sand and medium sand belong to the sandy soil, which is characterized by high permeability and distributed below the mucky soil. Gravel, sandstone, silty mudstone, and siltstone are classified as sandstone soil, which is distributed at the bottom of quaternary and has the properties of high permeability, macropores, and high shear strength.

Groundwater of the study area is mainly sourced from quaternary pore water and bedrock fissure water, where the pore water is further divided into unconfined and confined water. The unconfined water mainly lies in the soft soil layer, while the confined water is mainly found in the sandy and sandstone soil layers [21]. The changes of groundwater are affected by the precipitation and anthropogenic activities [23]. The annual average precipitation is about 1,690 mm, of which 50% concentrates in summer, 40% concentrates in spring and autumn, and 10% concentrates in winter [21]. Therefore, changes of groundwater due to the precipitation present the long-term periodical variations: high groundwater level in summer, middle groundwater level in spring and autumn, and low groundwater level in winter. Changes of groundwater due to the anthropogenic activities show the short-term fluctuations with domestic and industrial

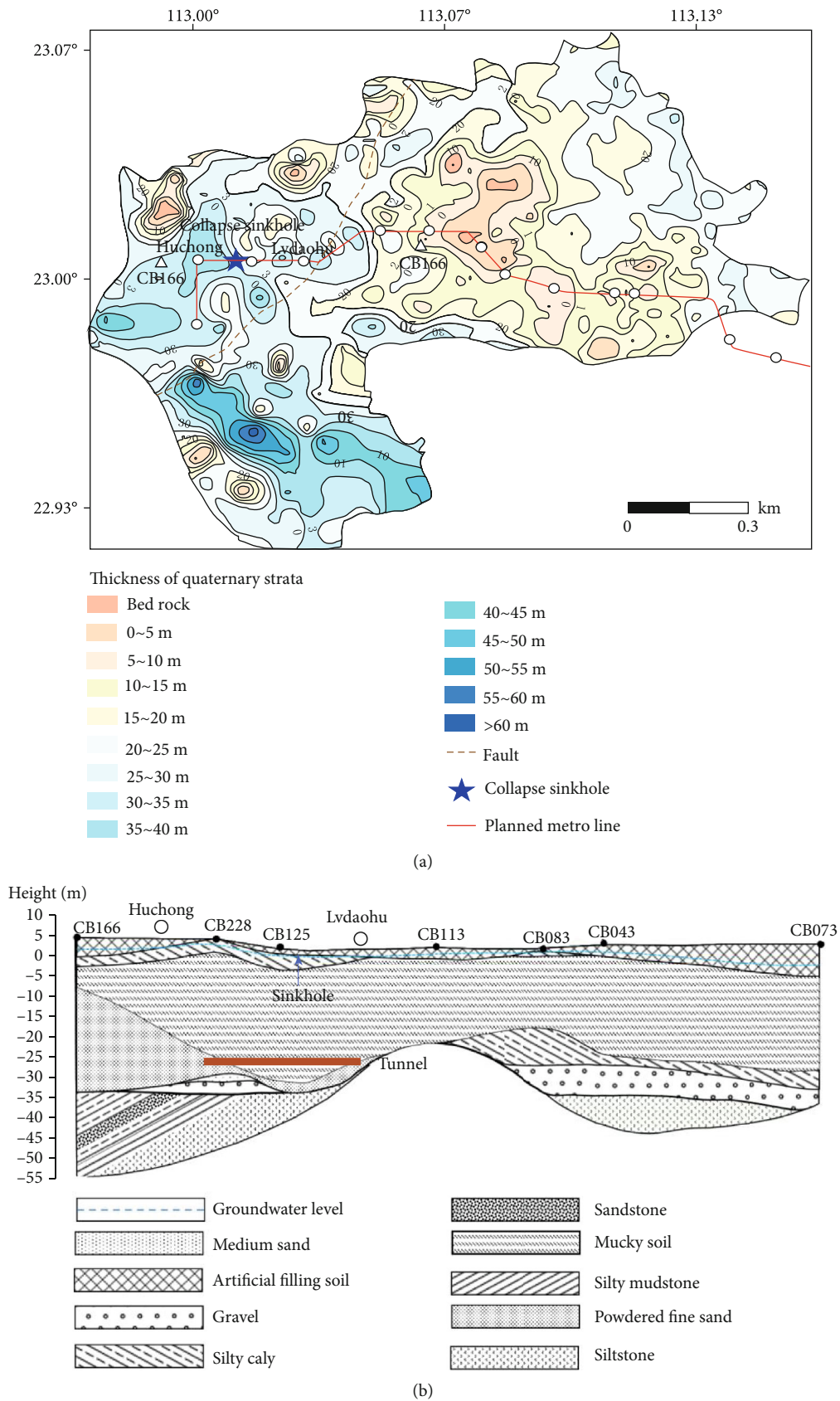


FIGURE 2: (a) Thickness of the quaternary strata and (b) stratigraphic profile over the study area. The reference height is the sea level in (b).

TABLE 1: SAR parameters in this study.

Satellite	Orbit direction	Azimuth angle (°)	Incidence angle (°)	Number of SAR images	Data period
Sentinel-1A	Ascending	-10	40	55	12/03/2017–25/01/2019

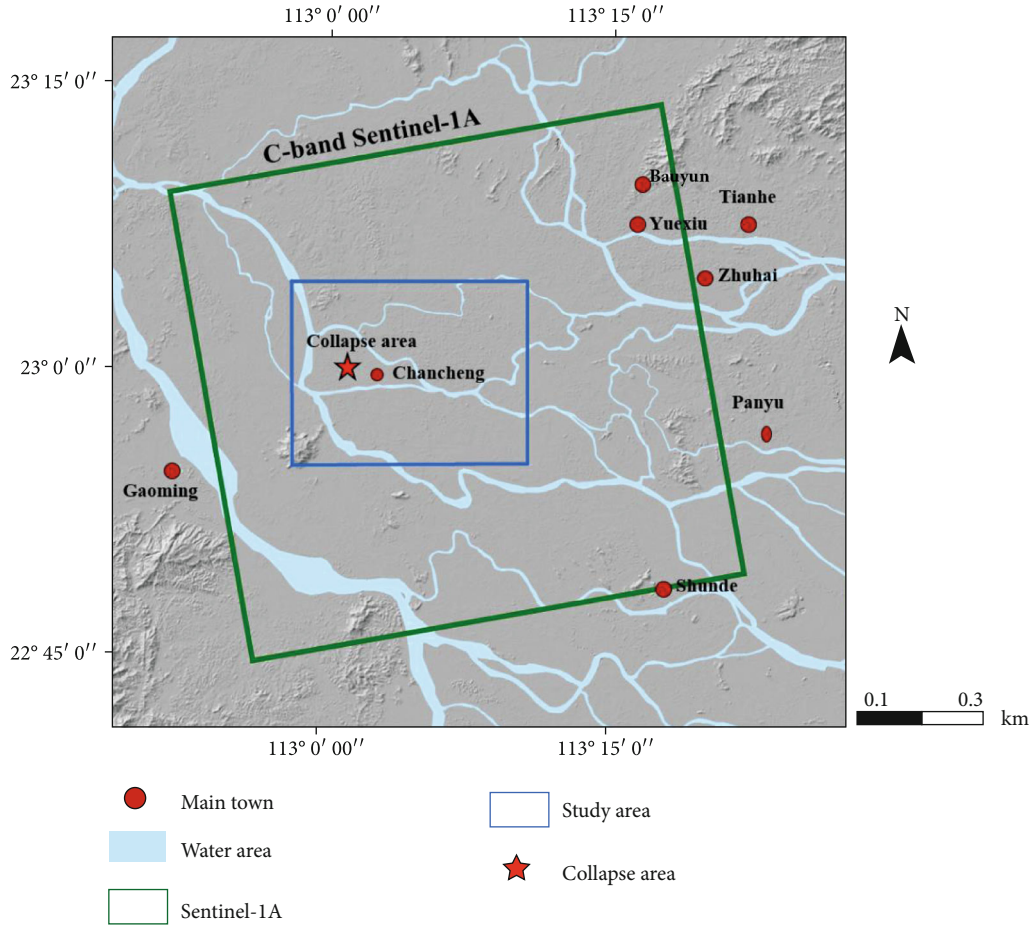


FIGURE 3: Synthetic aperture radar (SAR) data coverage used in this study; corresponding data are superimposed on the digital elevation model (DEM).

water consumptions, as well as underground construction. The domestic and industrial water consumptions are not dominated over the study area due to the rich surface water [26]. The operations of pumping and discharging groundwater during the underground construction, such as excavations of a metro tunnel, may cause the significant groundwater changes.

2.2. Data Collection. A total of 55 C-band Sentinel-1A images were collected to derive the ground deformation over the Foshan collapse area. Table 1 and Figure 3 provide the parameters and coverage of these SAR images, respectively. The Sentinel-1A images in the ascending orbit were acquired between March 2017 and January 2019, which allowed us to obtain the ground deformation over one year before and after the formation of the collapse sinkhole. A TerraSAR-X add-on for Digital Elevation Measurements (TanDEM-X) with a spatial resolution of 90 m was acquired as an external digital

elevation model (DEM) to remove the topographic phase from the differential interferograms.

3. Methodology

Using the collected 55 C-band Sentinel-1A images, the annual deformation velocity and time-series deformation were calculated using the multitemporal SAR interferometry technique. During this procedure, we followed the methods of Chen et al. [27]. Interferometric pairs were generated by setting small temporal and spatial baselines. On the basis of the experiment, a spatial baseline below 100 m and a temporal baseline less than 30 days were designed to generate the interferograms. A total of 96 interferograms were produced, among which 42 interferograms were used for precollapse deformation estimation and 54 interferograms were used for postcollapse deformation estimation, as shown in Figure 4. The topographic and orbital phases were simulated

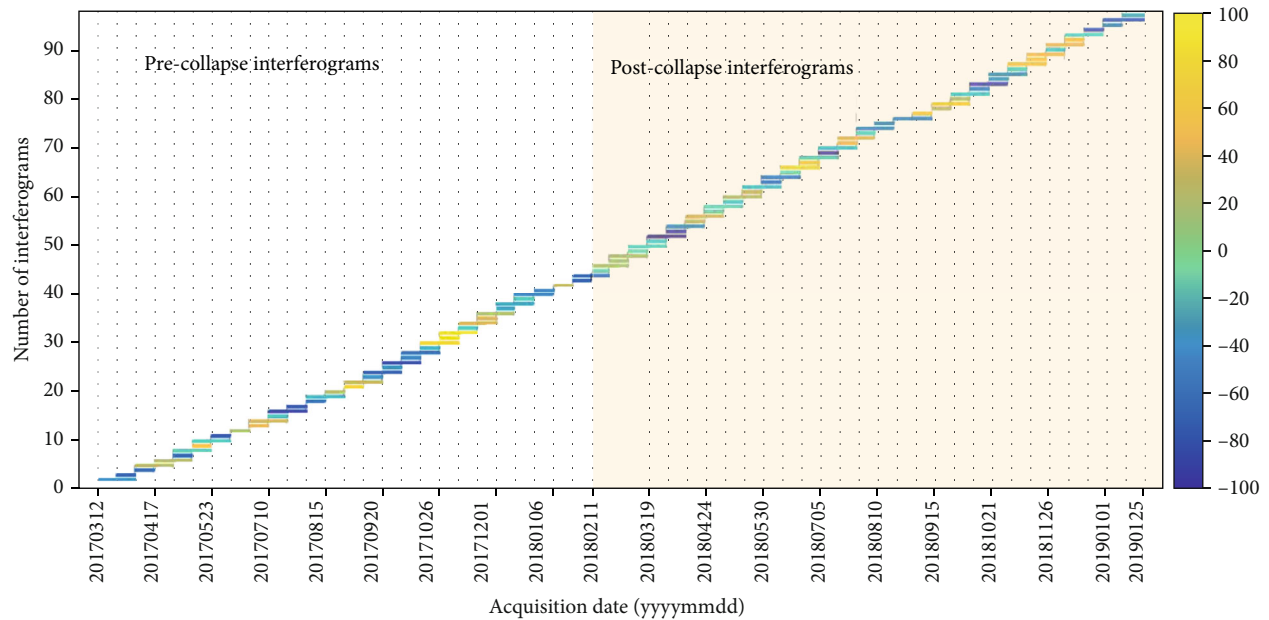


FIGURE 4: The distribution of the temporal and spatial baselines before and after the collapse event.

and subsequently removed from the differential interferometric phase using the collected TanDEM-X DEM and SAR orbital data [27]. Then, adaptive spectral filtering with a window size of 32 was executed to suppress the interferometric noise [28]. After that, phase unwrapping using the minimum cost flow (MCF) method was conducted to retrieve the absolute phase [29]. Finally, the atmospheric delay errors were reduced through the cascade of a low-pass filtering implemented in the two-dimensional spatial domain, followed by a temporal high-pass filtering [30]. After these processes, the unwrapped interferograms were used to estimate the annual deformation velocity and time-series deformation before and after the formation of the collapse sinkhole.

The annual deformation velocity, which allowed the characterization of the spatial pattern of ground deformation over the study area, was estimated by the method in [31]:

$$\text{Ph}_{\text{rate}} = \frac{\sum_{j=1}^N \Delta t_j \varphi_j}{\sum_{j=1}^N \Delta t_j^2}, \quad (1)$$

where Ph_{rate} is the annual deformation velocity, N is the number of interferograms involved in the estimation, Δt_j is the time interval between the master and slave pair, and φ_j is the unwrapped interferograms. The standard deviation of Ph_{rate} was estimated by the method in [32]:

$$\text{Var}(\text{Ph}_{\text{rate}}) = \sum_{j=1}^N \left(\varphi_j - \frac{4\pi}{\lambda} \cdot \text{Ph}_{\text{rate}} \cdot \Delta t_j \right)^2 / \Delta t_j^2, \quad (2)$$

where λ is the radar wavelength and $\text{Var}(\text{Ph}_{\text{rate}})$ is the standard deviation of Ph_{rate} . $\text{Var}(\text{Ph}_{\text{rate}})$ was used to validate our observed annual deformation velocity due to lack of

additional observations, such as ground-based leveling measurements.

The time-series deformation demonstrates the temporal variations in ground deformation and can be estimated by the method in [33]:

$$A \cdot \cos \theta \cdot V_u = \varphi, \quad (3)$$

where matrix A consists of time intervals between consecutive SAR acquisition, θ is the incidence angle, V_u represents the unknown vertical velocities that are to be determined, and φ represents the unwrapped interferometric phase. The unknown parameters V_u for each pixel are solved by applying the singular value decomposition (SVD) [34]. Once V_u was estimated, the time-series deformation $d = [d_1 d_2 \dots d_{M-1} d_M]$ was reconstructed from the computed deformation rates by numerical integration, where M is the number of SAR images involved in calculation.

4. Results

4.1. Precollapse Ground Deformation. The annual deformation velocity was estimated by Equation (1) using 26 Sentinel-1A images acquired between 12 March 2017 and 30 January 2018, which revealed the spatial pattern of ground deformation over the year prior to the collapse accident, as shown in Figure 5(a). All measurement values are relative to a common reference point that was considered to be stable over the entire time period, and positive and negative values represent uplift and subsidence, respectively. The prominently subsiding region was detected in the vicinity of the collapse sinkholes, as indicated by the black outline polygon in Figure 5(a). This subsiding region was further divided into three subregions. The first subregion contained the collapse

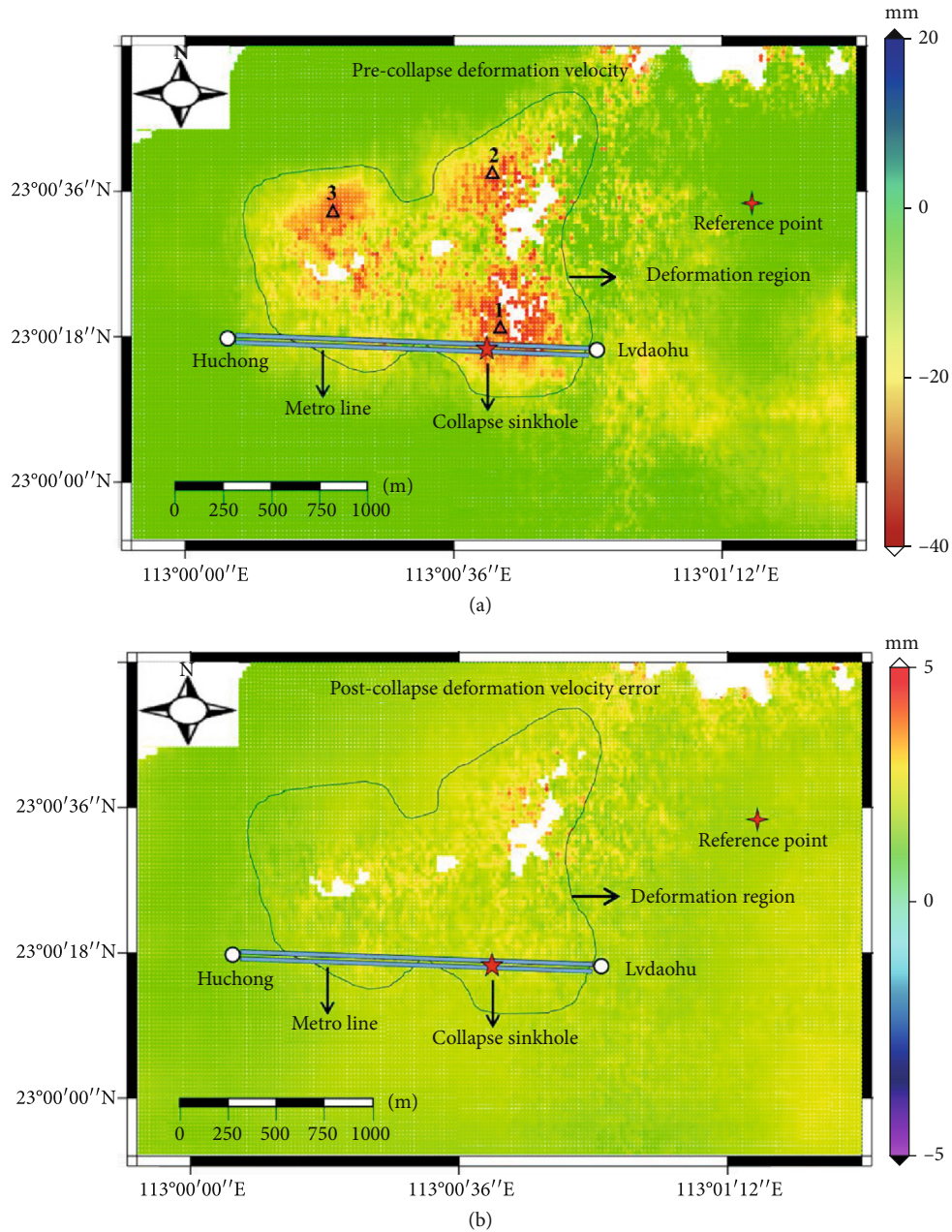


FIGURE 5: (a) Annual deformation velocity map and (b) its error distribution before the formation of the collapse sinkhole. The red cross indicates the reference point, the red star denotes the location of the collapse sinkhole, the white circles show the metro station locations, the blue double line shows the metro line, and the red triangles show three subregions.

region and showed the maximum subsiding rate of 42 mm/year. The second subregion, with a maximum subsiding rate of 38 mm/year, was located at about 600 m north of the collapse sinkhole. The third subregion was located at about 900 m northwest of the collapse sinkhole, where the maximum subsiding rate was up to 34 mm/year. To evaluate the observed result, the error distribution of deformation velocity was estimated using Equation (2), as shown in Figure 5(b). It was observed that the errors were below 3 mm for most regions and up to 5 mm for the severe subsiding area, suggesting the reliability of our observed subsidence.

Based on this observation, it was summarized that prominent ground subsidence with a maximum rate of 42 mm/year appeared in the vicinity of the sinkhole before the collapse.

To investigate the temporal variation of precollapse deformation, the time-series deformation was estimated using multitemporal SAR interferometry processing. Figure 6 shows the time-series deformation from 12 March 2017 to 30 January 2018, which was derived from 26 Sentinel-1A images acquired before the formation of the collapse sinkhole. Continuous subsidence was observed in the vicinity of the collapse, where the cumulative maximum

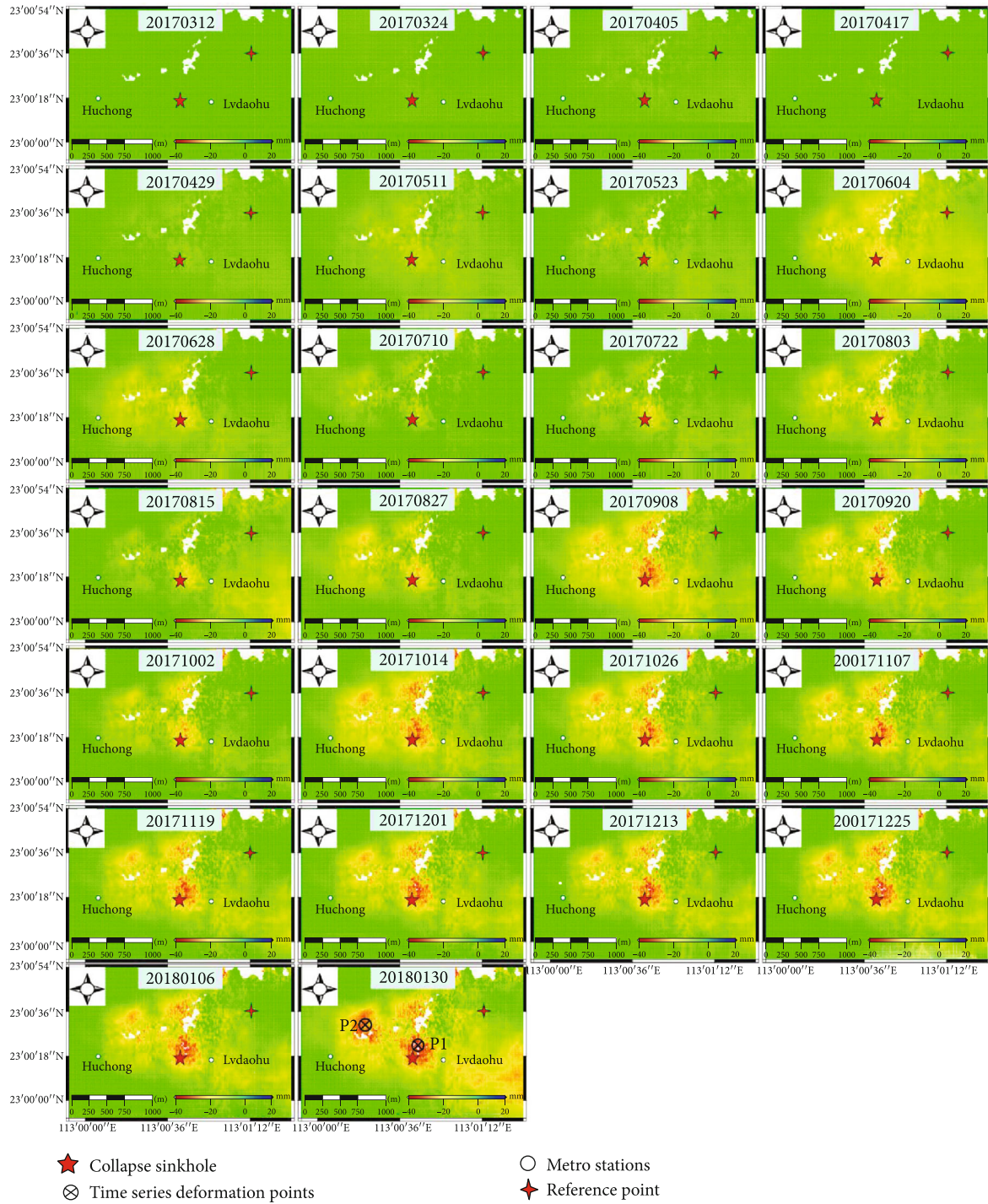


FIGURE 6: Time-series deformation before the formation of the collapse sinkhole observed from Sentinel-1A images. The time-series deformation at points P1 and P2 was extracted for further analysis.

subsidence reached 40 mm. Figure 7 displays the time-series deformation at points P1 (blue rectangle) and P2 (red circle), both of which are marked in the last image of Figure 6. Point P1, which is about 100 m away from the center of collapse, showed a cumulative subsidence of 35 mm between 12 March 2017 and 30 January 2018. Further observation indicated a fluctuation in time-series deformation: alternative subsi-

dence and uplift from 17 April to 5 August 2017 and subsequent continuous subsidence until 30 January 2018. The metro construction started from Huchong Station on 10 May 2017. Connecting this information with the observed deformation at point P1, it was found that the clear fluctuation of deformation appeared during this stage, suggesting the relationship between subsidence and metro construction.

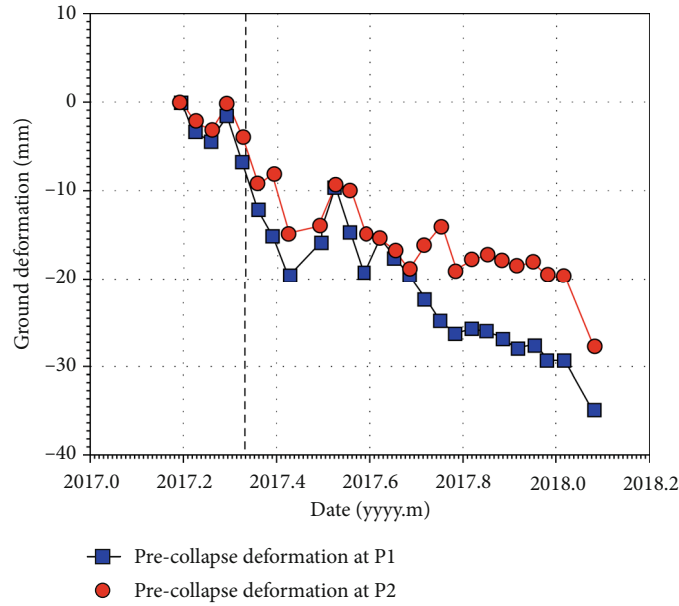


FIGURE 7: Time-series deformation at points P1 (a) and P2 (b) marked in Figures 6.

Point P2, which is about 900 m away from the center of the collapse sinkhole, showed similar variations in time-series deformation as point P1. The cumulative subsidence was about 28 mm for point 2, which is less than at point P1. As for the deformation in the month prior to the collapse event on 7 February 2018, accelerated subsidence was observed at both points: the deformation decreased by 7 mm and 8 mm between 6 January and 30 January 2018 for P1 and P2, respectively. In summary, continuous subsidence occurred in the vicinity of the collapse before the accident, and accelerated subsidence was observed when metro construction started and one month before the formation of the sinkhole.

4.2. Postcollapse Ground Deformation. Figure 8 shows the annual deformation velocity and its error distribution after the formation of the collapse sinkhole, which was derived from 29 Sentinel-1A images acquired between 11 February 2018 and 25 January 2019. Similar to the precollapse deformation velocity map in Figure 5(a), the subsiding region occurred in the vicinity of the collapse, as shown by the black outline in Figure 8(a). However, the area and magnitude of subsidence are smaller compared with Figure 5(a). The statistics showed that the maximum subsiding rates were about 30, 31, and 32 mm/year for the first, second, and third subsiding subregions, respectively. Meanwhile, it was observed that the subsidence started to extend to the east, where subsidence area and magnitude increased. The lower subsiding rates were due to the completion of the metro construction over the study area, and the construction continuing to the east extended the subsidence. This is similar to the precollapse deformation where the deformation velocity errors were below 3 mm for most regions and up to 5 mm for the severely subsiding area, as shown in Figure 8(b). Based on this observation, we found decreases in the area and magnitude of subsidence after the collapse, which may be related with the groundwater level changes.

Figure 9 shows the time-series deformation from 11 February 2018 to 25 January 2019, which was derived from 29 Sentinel-1A images after the collapse of the sinkhole. Like the precollapse time-series deformation in Figure 6, continuous subsidence was observed in the vicinity of the collapse sinkholes, where the cumulative maximum subsidence reached 30 mm. However, the area and magnitude of subsidence are smaller compared with Figure 6. Figure 10 displays the time-series deformation at P1 (blue rectangle) and P2 (red circle), both of which are marked in the last image of Figure 9. Point P1 experienced a cumulative subsidence of 20 mm between 11 February 2018 and 25 January 2019, which is a decrease of 15 mm compared with the precollapse subsidence. Although the wave-shaped deformation variation was retained, the amplitude of variation was lower when compared with Figure 7(a). Point 2 experienced a cumulative subsidence of 28 mm, which was the same as the precollapse subsidence. In summary, continuous but decelerated subsidence was observed after the incident, and accelerated subsidence was observed over 48 days after the formation of the sinkhole.

5. Discussion

5.1. Subsidence due to Soft Soil Consolidation. Located on a large alluvial delta, unconsolidated sediments are widespread in the study area (Figure 2(a)). The stratum from top to bottom is artificial filling soil, silty clay, mucky soil, powdered fine sand, medium sand, gravel, sandstone, silty mudstone, and siltstone (Figure 2(b)). Considering the presented characteristics of high water content and high compressibility, the upper silty clay and mucky soil are classified as soft soil, which has many properties that are unfavorable for their use in projects [26]. The lower soil layers, including powdered fine sand, medium sand, gravel, sandstone, silty mudstone, and siltstone, are classified as sandy and sandstone

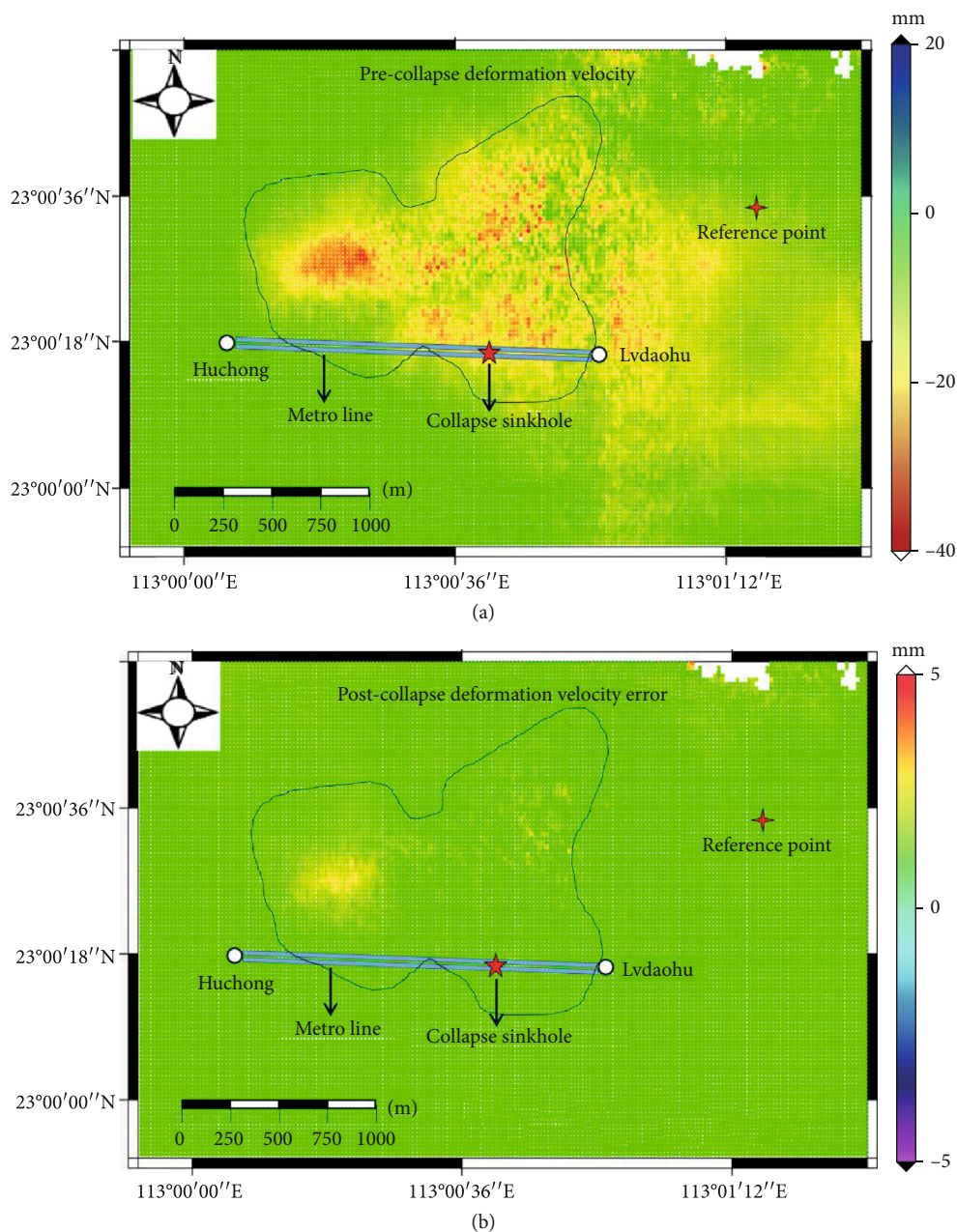


FIGURE 8: (a) Annual deformation velocity map and (b) its error distribution after the formation of the collapsed sinkhole. The red cross shows the reference point, the red star denotes the location of the collapse sinkhole, the white circles show the location of the metro station, and the blue double line indicates the metro line.

soils, which are characterized by high permeability, low compressibility, and high shear strength. In this context, pore water in soft soils is easily discharged along the sandy and sandstone soils in the presence of natural and human loadings, resulting in the decrease of pore fluid volume, compression of soft soils, and subsequent ground subsidence. The relevant materials show that the magnitude of subsidence is related to the thickness of soft soils [27, 34]. Thus, the relationship between ground deformation and soft soil thickness was investigated and analyzed over the study area. Figure 11(a) displays the thickness of soft soils over the study

area, where the thickness between Huchong Station and Lvdaohu Station is from 20 to 30 m. Figure 11(b) shows the relationship between ground deformation and soft soil thickness. The green circle in Figure 11(b) represents the ground deformation between Huchong Station and Lvdaohu Station, which is extracted from Figure 5(a). The red and blue lines are, respectively, the fitting deformation and soft soil thickness between Huchong Station and Lvdaohu Station. The comparison results show the approximate consistency between the ground deformation and the soft soil thickness in the spatial distribution; the greater the subsidence, the

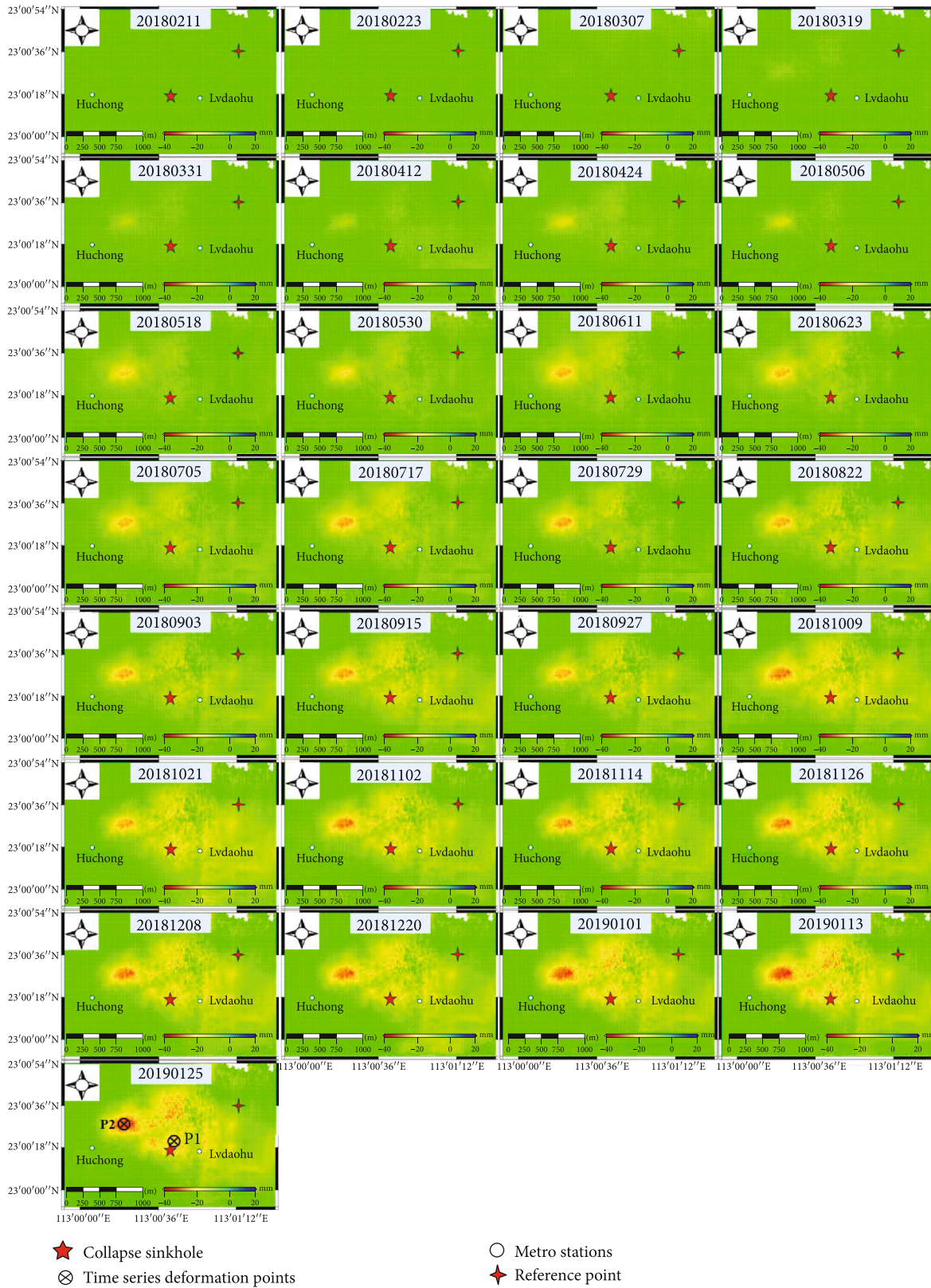


FIGURE 9: Time-series deformation after the formation of the collapse sinkhole observed from Sentinel-1A images. The time-series deformation at points P1 and P2 was extracted for further analysis.

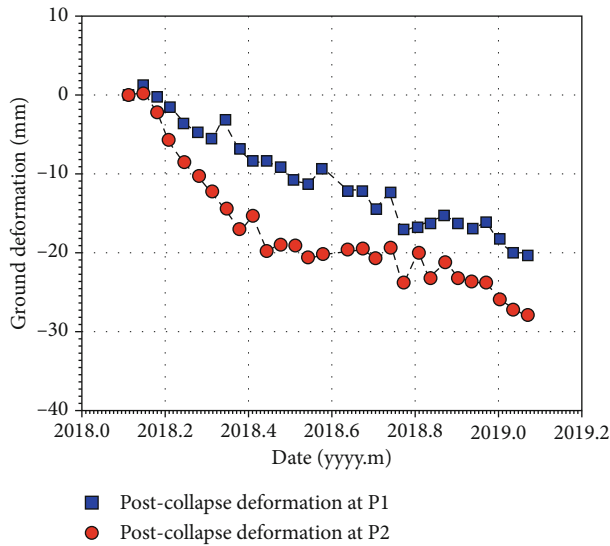


FIGURE 10: Time-series deformation at points P1 (a) and P2 (b).

thicker the soft soils, particularly in the vicinity of the collapse sinkhole, where the maximum subsidence and soft soil thickness are observed.

The literature indicates that the subsidence is due to the soft soil consolidation mainly from the surface loadings [22] and groundwater exploitation [25]. With respect to the collapse area, surface loadings come from the vehicle loadings and some buildings and infrastructure. Groundwater exploitation to meet the needs of domestic and industrial water consumption was rare over the study area due to the rich surface water [26]. However, this process occurs during the tunnel excavation, where the operations of pumping and discharging groundwater are needed to reduce the external water pressure [35].

5.2. Subsidence due to Soil Losses. The studies indicate that the soil losses during the tunnel excavation may cause the subsidence, which is located at the center of the tunnel [36, 37]. This subsidence occurs when a difference exists in the volumes of the excavated soil and tunnel. In this case, soil losses contributed to the subsidence according to the adopted shield tunnel excavating method and the local geological conditions [37]. Figure 12 shows the annual deformation velocity along the planned metro line. It is observed that prominent subsidence appears along the completed metro line, which is caused by the tunnel-induced soft soil consolidation and soil losses. However, the deformation is not significant for the region where the tunnel excavation has not started. This observation suggests the effect of tunnel excavation on ground subsidence.

In order to further show the ground subsidence induced by soil loss, Peck model, an empirical model to calculate the deformation, was used in this study [36]. A schematic diagram of the Peck model is shown in Figure 13, where point O is the center of the tunnel, R is the tunnel radius, H is the depth of the tunnel centerline, S_{\max} is the maximum subsidence, $S(x)$ is the deformation due to soil loss, and i is the width coefficient of the subsiding trough, which can be

empirically estimated as a function of the depth of the tunnel centerline H , and the horizontal direction of the tunnel cross section is the x axis. The model is expressed as [36]

$$S(x) = S_{\max} \exp\left(\frac{-x^2}{2i^2}\right), \quad (4)$$

$$S_{\max} = -\frac{V_s}{i\sqrt{2\pi}} = -\frac{V_s}{2.5i}, \quad (5)$$

$$V_s = V_1 \times \pi R^2, \quad (6)$$

where V_s and V_1 are the soil loss and soil loss rate, respectively, both of which are difficult to accurately estimate since they are related to many factors, such as stratigraphic condition, hydrogeology, geotechnical parameters, rock/soil porosity, water content, granulometry, construction technique, and excavating method. In this study, R and H in Equations (4), (5), and (6) were directly derived from [21] and parameters $V_1 = 3\%$ and $i = 0.76 * H$ were derived from [37]. Based on these parameters, the ground subsidence due to soil loss was estimated, as shown in Figure 14. It was found that the subsidence due to soil losses was in the shape of a funnel, with the subsiding center was located at the center of the tunnel. Compared with the deformation map in Figure 6, the three detected subsiding subregions were approximately shaped as funnels, suggesting consistency in the deformation shape. For the subsiding center, the first subsiding subregion showed a center along the metro line. However, the other two subregions did not overlap the metro line. This phenomenon was due to the reduction in the subsiding area after the tunnel excavation, which was identified in the time-series deformation map. In terms of magnitude of subsidence, the modeled maximum subsidence due to soil losses in this case was about 25 mm. The observed subsidence due to soil losses at point P1 in Figure 7 was about 28 mm when considering 80% of total subsidence from soil loss [37]. The modeled and observed subsidence at point P1 is comparable in magnitude. Based on the discussion, it was summarized that the observed subsidence in this case was due to soft soil consolidation and soil losses. Subsidence due to soft soil consolidation occurred before and after tunnel excavation, whereas subsidence due to soil losses occurred during tunnel excavation.

5.3. Collapse due to Groundwater Leakage. The result from an investigation report showed that the Foshan collapse occurred due to leakage of groundwater above the tunnel [21]. The groundwater poured into the tunnel, causing the tunnel structure damage and triggering the collapse. In view of filed investigation, the damage of underground pipelines is considered to be related with the leakage of groundwater. Here, three aspects are shown to support this point. The first is the aging of underground pipelines over the study area. Like most other cities, there are more than 30-year-old underground pipelines in Foshan region [38]. This aging of underground pipelines is easily damaged in the case of ever-increasing urbanization and the continuous construction, development, and expansion of urban areas, such as

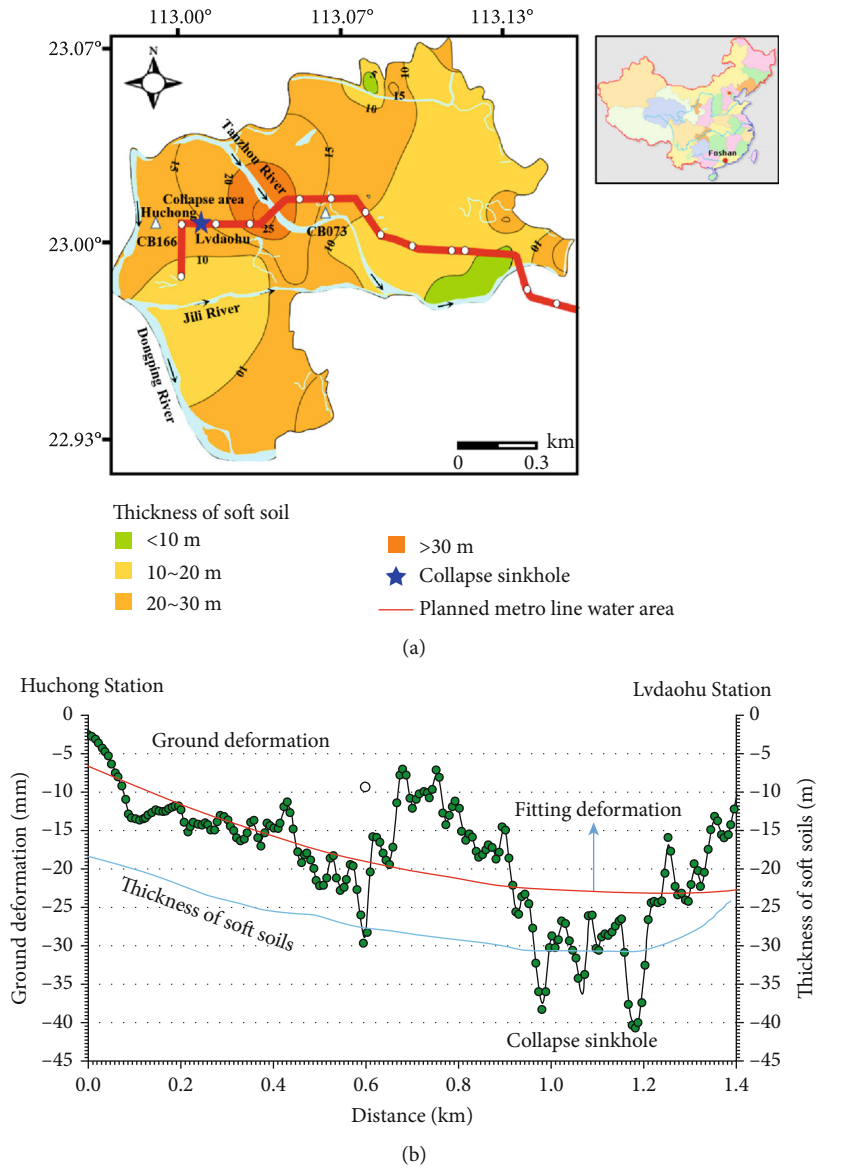


FIGURE 11: (a) Thickness of soft soils and (b) relationship between ground deformation and thickness of soft soils.

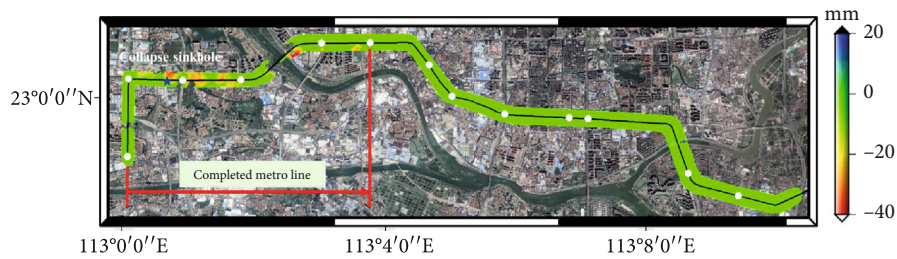


FIGURE 12: Ground deformation along the planned metro line.

the abrupt collapse due to a 44-year-old sewer pipeline in Fraser, USA [39]. The second is the observed inhomogeneous ground subsidence over the study area. The filed investigation suggests that some infrastructures have been affected by ground subsidence in the vicinity of the collapse area, such

as ground crack, wall crack, and pipeline rupture, as shown in Figures 15(a)–15(d). Although there is no direct evidence to show this effect on the underground pipelines, plenty of research has proved that the inhomogeneous ground subsidence can cause the underground pipeline leakage [40–42].

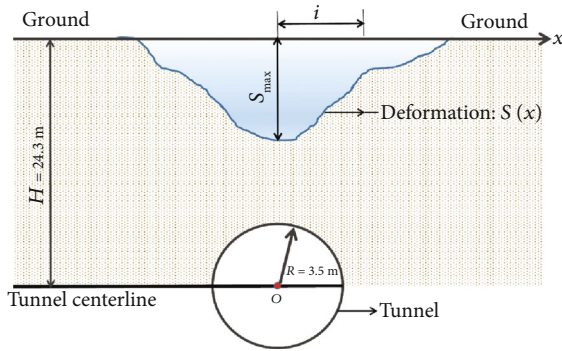


FIGURE 13: Schematic diagram of ground subsidence due to the tunnel excavation.

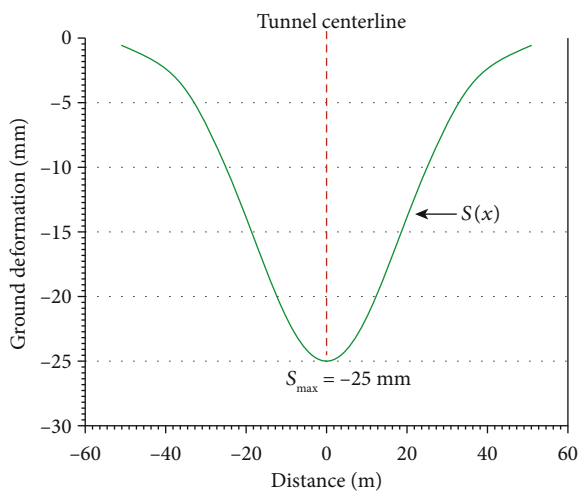


FIGURE 14: The modeled ground subsidence due to tunnel excavation.

The leakage in underground pipelines mainly refers to the leakages, bursts, or blockages in sewer, drain, and/or water pipelines [43, 44]. The third is the similar collapse event in the vicinity of the collapse area. On November 5, 2017, a small ground collapse event was observed near the station of Huchong, which caused the formation of a collapse sinkhole with an area of 60 m^2 [2]. It is confirmed that the collapse is caused by the underground pipeline leakage, as shown in Figures 15(e) and 15(f). Based on these three aspects, it is the damage of underground pipelines causing the leakage of groundwater. Therefore, the detection of underground pipeline leakage should be conducted to prevent the collapse events, particularly for subsiding areas.

Related studies investigated the mechanisms and necessary conditions for collapse sinkhole formation due to the leakage of groundwater [6, 16, 45–47]. Based on these studies, the widely accepted process of collapse sinkhole formation is shown in Figure 16. Firstly, the leakage of groundwater erodes the soil, forming an arched hole. This arched hole is generally related to the size and location of the later formed collapse sinkhole and is therefore prerequisite to the formation of a collapse sinkhole [48, 49]. Anthropogenic influences and the presence of urban infrastructure

can significantly expedite the process of arched hole formation [50, 51]. With the excavation of the tunnel, the arched hole and the tunnel are connected. This connection reduces the bearing capacity of the soil layer above the arched hole; hence, a ground collapse sinkhole forms. Thus, the collapse event occurred in this region due to groundwater leakage and arched hole formation.

6. Conclusions

On the evening of 7 February 2018, a deadly collapse of a metro tunnel under construction in the Southern China city of Foshan caused 11 deaths, 8 injuries, and 1 missing person. To better understand this event, the ground deformations before and after the collapse was derived using a multitemporal InSAR technique. The driving factors for the ground deformation and the collapse were discussed based on the investigation report, local geological conditions, and observed deformation. The findings provide scientific evidence for future collapse monitoring and early warning due to tunnel excavation. Based on this study, the following conclusions were made:

- (1) Precursory subsidence was observed in the vicinity of collapse sinkhole. Using 55 C-band Sentinel-1A images, the annual deformation velocity and time-series deformation maps before and after the event were retrieved through multitemporal SAR interferometry processing. The results showed that prominent ground subsidence with a maximum rate of 42 mm/year occurred in the vicinity of the collapse before the formation of the sinkhole. Compared with precollapse deformation, the area and magnitude of subsidence decreased during postcollapse deformation. In the temporal domain, accelerated subsidence was observed during metro construction and one month before the formation of the collapsed sinkhole. These observations indicated the presence of precursory subsidence for collapse sinkhole, which allowed us to forecast the collapses and thus reduce the risk for people and property
- (2) Soft soil consolidation and soil losses were the main factors affecting the subsidence over the study area. The quaternary stratum from top to bottom was soft, sandy, and sandstone soils. In this context, the ground subsidence easily occurred due to soft soil consolidation and tunnel excavation. The time-series deformation maps showed the continuous subsidence before and after the metro construction. This subsidence was mainly from the soft soil consolidation due to surface loadings and groundwater exploitation. The modeled and observed subsidence is comparable in magnitude, suggesting the subsidence due to soil losses during tunnel excavation
- (3) The leakage of groundwater accounted for the collapse event. The result from an investigation report showed that the collapse occurred due to the leakage of groundwater above the tunnel. Considering the



(a)



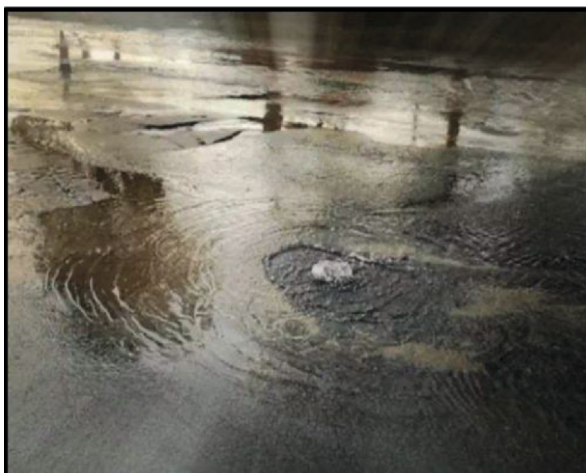
(b)



(c)



(d)



(e)



(f)

FIGURE 15: The effects of ground deformation in the vicinity of the collapse area, such as surface subsidence (a), ground crack (b), wall crack (c, d), and leakage of groundwater (e, f).

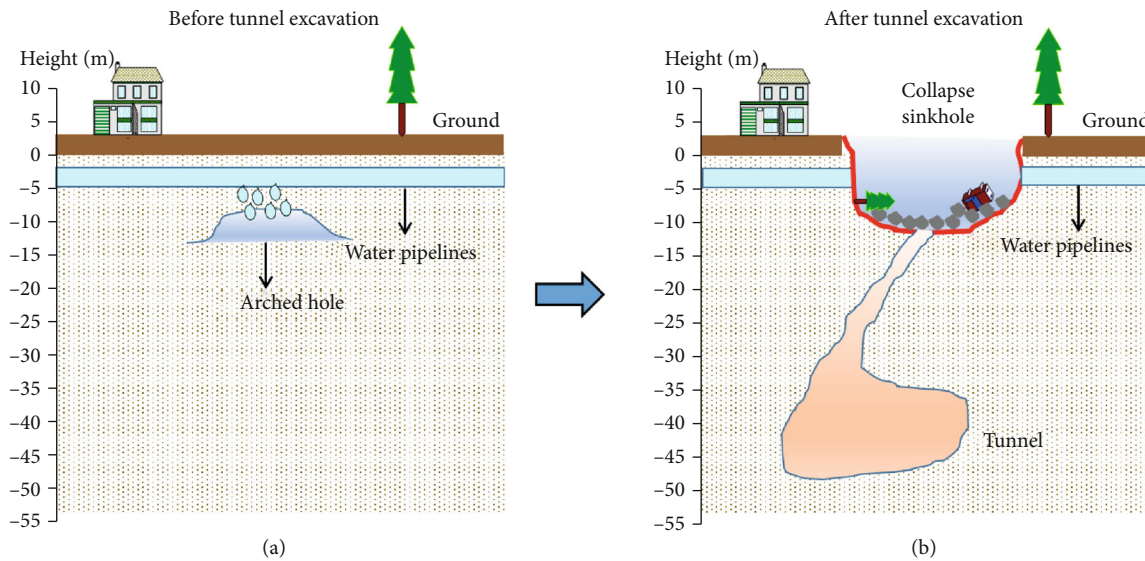


FIGURE 16: Schematic diagram of the formation of the collapse sinkhole (a) before and (b) after tunnel excavation.

mechanisms and necessary conditions for collapse sinkhole formation, the process of collapse was retrieved. The leaked groundwater eroded the soil, forming an arched hole. The connection between the arched hole and the tunnel reduced the bearing capacity of the soil layer above the arched hole, thereby triggering the collapse event

These findings contribute to the understanding of the spatial-temporal evolution of ground deformation before and after the collapse event. However, clearly describing the reasons for collapse is challenging due to a lack of related data, such as continuous ground-based deformation monitoring data. Therefore, sufficient in situ measurements will be collected in the future to analyze the reasons for the collapse.

Data Availability

Sentinel-1A/B images were from the European Space Agency (ESA) project, and these data were downloaded from the Sentinel-1 Scientific Data Hub. ALOS-2 images were provided by the Japan Aerospace Exploration Agency (JAXA).

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

Several figures were prepared by using Generic Mapping Tools software. This research was funded by Chang'an University (Xi'an, China) through the National Key R&D Program of China (2019YFC1509802 and 2018YFC1505101), Natural Science Foundation of China projects (NSFC) (Nos. 41941019, 41790445, 41731066, 41674001, and 41604001), Natural Science Basic Research Plan in Shaanxi Province of

China (2018JQ4031 and 2018JQ4027), and the Fundamental Research Funds for the Central Universities, CHD (Nos. 300102269207, 300102269303, and 300102268204).

References

- [1] F. Gutiérrez, A. Benito-Calvo, D. Carbonel et al., "Review on sinkhole monitoring and performance of remediation measures by high-precision leveling and terrestrial laser scanner in the salt karst of the Ebro Valley, Spain," *Engineering Geology*, vol. 248, pp. 283–308, 2019.
- [2] W. Lienhart, "Geotechnical monitoring using total stations and laser scanners: critical aspects and solutions," *Journal of Civil Structural Health Monitoring*, vol. 7, no. 3, pp. 315–324, 2017.
- [3] J. D. Kent and L. Dunaway, "Real-time GPS network monitors Bayou Corne sinkhole event," *Eos, Transactions American Geophysical Union*, vol. 94, no. 43, pp. 385–386, 2013.
- [4] C. L. Aiken, J. S. Oldow, M. Alfarhan, L. S. White, and T. Ahmed, "Monitoring growth of the Daisetta, Texas sinkhole with terrestrial laser scanning, close range digital photography and GPS," *Geological Society of America Abstracts with Programs*, vol. 41, no. 2, p. 34, 2009.
- [5] G. Baer, U. Schattner, D. Wachs, D. Sandwell, S. Wdowski, and S. Frydman, "The lowest place on earth is subsiding—an InSAR (interferometric synthetic aperture radar) perspective," *Geological Society of America Bulletin*, vol. 114, no. 1, pp. 12–23, 2002.
- [6] M. Abelson, G. Baer, V. Shtivelman et al., "Collapse-sinkholes and radar interferometry reveal neotectonics concealed within the Dead Sea basin," *Geophysical Research Letters*, vol. 30, p. 1545, 2003.
- [7] D. Closson, N. A. Karaki, H. Hansen, D. Derauw, C. Barbier, and A. Ozer, "Space-borne radar interferometric mapping of precursory deformations of a dyke collapse, Dead Sea area," *International Journal of Remote Sensing*, vol. 24, pp. 843–849, 2010.

- [8] R. N. Nof, G. Baer, A. Ziv, E. Raz, S. Atzori, and S. Salvi, "Sinkhole precursors along the Dead Sea, Israel, revealed by SAR interferometry," *Geology*, vol. 41, no. 9, pp. 1019–1022, 2013.
- [9] S. Atzori, A. Antonioli, S. Salvi, and G. Baer, "InSAR-based modeling and analysis of sinkholes along the Dead Sea coastline," *Geophysical Research Letters*, vol. 42, no. 20, pp. 8383–8390, 2015.
- [10] G. Baer, Y. Magen, R. N. Nof, E. Raz, V. Lyakhovskiy, and E. Shalev, "InSAR measurements and viscoelastic modeling of sinkhole precursory subsidence: implications for sinkhole formation, early warning, and sediment properties," *Journal of Geophysical Research - Earth Surface*, vol. 123, no. 4, pp. 678–693, 2018.
- [11] R. N. Nof, M. Abelson, E. Raz et al., "SAR interferometry for sinkhole early warning and susceptibility assessment along the Dead Sea, Israel," *Remote Sensing*, vol. 11, no. 1, p. 89, 2019.
- [12] J. W. Kim, Z. Lu, and K. Degrandpre, "Ongoing deformation of sinkholes in Wink, Texas, observed by time-series Sentinel-1A SAR interferometry (preliminary results)," *Remote Sensing*, vol. 8, no. 4, p. 313, 2016.
- [13] J. W. Kim, Z. Lu, and J. Kaufmann, "Evolution of sinkholes over Wink, Texas, observed by high-resolution optical and SAR imagery," *Remote Sensing of Environment*, vol. 222, pp. 119–132, 2019.
- [14] Y. Shi, Y. Tang, Z. Lu, J. W. Kim, and J. Peng, "Subsidence of sinkholes in Wink, Texas from 2007 to 2011 detected by time-series InSAR analysis," *Geomatics, Natural Hazards and Risk*, vol. 10, no. 1, pp. 1125–1138, 2019.
- [15] L. Chang and R. F. Hanssen, "Detection of cavity migration and sinkhole risk using radar interferometric time series," *Remote Sensing of Environment*, vol. 147, pp. 56–64, 2014.
- [16] F. Gutiérrez, J. P. Galve, P. Lucha, C. Castañeda, J. Bonachea, and J. Guerrero, "Integrating geomorphological mapping, trenching, InSAR and GPR for the identification and characterization of sinkholes: a review and application in the mantled evaporite karst of the Ebro Valley (NE Spain)," *Geomorphology*, vol. 134, no. 1–2, pp. 144–156, 2011.
- [17] A. Theron, J. Engelbrecht, J. Kemp, W. Kleynhans, and T. Turnbull, "Detection of sinkhole precursors through SAR interferometry: first results from South Africa," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 5398–5401, Beijing, China, July 2016.
- [18] A. A. Malinowska, W. T. Witkowski, R. Hejmanowski, L. Chang, F. J. van Leijen, and R. F. Hanssen, "Sinkhole occurrence monitoring over shallow abandoned coal mines with satellite-based persistent scatterer interferometry," *Engineering Geology*, vol. 262, p. 105336, 2019.
- [19] B. Conway and J. Cook, "Monitoring evaporite karst activity and land subsidence in the Holbrook Basin, Arizona using interferometric synthetic aperture radar (InSAR)," *Proceedings of the Sinkhole Conference*, pp. 187–194, Carlsbad, New Mexico, 2013.
- [20] J. Wang and R. Li, "Subway Construction Collapse Blamed on Negligence, Misjudgment," *Society and Culture*, 2018, <https://www.caixinglobal.com/2018-07-31/subway-construction-collapse-blamed-on-negligence-misjudgment-101310265.html>.
- [21] Investigation Committee, *Investigation report on "2.7" water leakage and ground collapse major accident in Project I of Foshan Metro Line 2 in Guangdong*, The people's Government of Guangdong Province of China, Guangdong, China, 2018.
- [22] A. H. M. Ng, H. Wang, Y. Dai et al., "InSAR reveals land deformation at Guangzhou and Foshan, China between 2011 and 2017 with COSMO-SkyMed data," *Remote Sensing*, vol. 10, no. 6, p. 813, 2018.
- [23] C. Yu, A. Zhou, J. Chen, A. Arulrajah, and S. Horpibulsuk, "Analysis of a tunnel failure caused by leakage of the shield tail seal system," *Underground Space*, vol. 5, no. 2, pp. 105–114, 2019.
- [24] H. Wang, Y.-P. Yu, and L. L. Jiang, "Monitoring land subsidence in Guangzhou and Foshan using InSAR," *Science Survey Mapping*, vol. 39, pp. 66–71, 2014.
- [25] Q. Liu, G. Yue, X. Ding, K. Yang, G. Feng, and Z. Xiong, "Temporal and spatial characteristics analysis of deformation along Foshan Subway using time series InSAR," *Geomatics and Information Science of Wuhan University*, vol. 44, pp. 1099–1106, 2019.
- [26] Geological bureau of Foshan, *The analysis report of ground subsidence in Foshan*, Geological bureau of Foshan, Guangdong, 2018.
- [27] G. Chen, Y. Zhang, R. Q. Zeng et al., "Detection of land subsidence associated with land creation and rapid urbanization in the Chinese loess plateau using time series InSAR: a case study of Lanzhou New District," *Remote Sensing*, vol. 10, no. 2, p. 270, 2018.
- [28] I. Baran, M. P. Stewart, B. M. Kampes, Z. Perski, and P. Lilly, "A modification to the Goldstein radar interferogram filter," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 2114–2118, 2003.
- [29] M. Costantini, "A novel phase unwrapping method based on network programming," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 813–821, 1998.
- [30] A. Ferretti, "Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 5, pp. 2202–2212, 2000.
- [31] D. T. Sandwell and E. J. Price, "Phase gradient approach to stacking interferograms," *Journal of Geophysical Research - Solid Earth*, vol. 103, no. B12, pp. 30183–30204, 1998.
- [32] D. T. Sandwell and L. Sichoix, "Topographic phase recovery from stacked ERS interferometry and a low-resolution digital elevation model," *Journal of Geophysical Research - Solid Earth*, vol. 105, no. B12, pp. 28211–28222, 2000.
- [33] P. Berardino, G. Fornaro, R. Lanari, and E. Sansosti, "A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2375–2383, 2002.
- [34] Beijing Vastitude Technology company and Geological bureau of Foshan, *The link between ground subsidence and geological parameters over Foshan*, Geological bureau of Foshan, Guangdong, 2018.
- [35] T. Xu, *Study on the deformation of pipelines used to earth pressure balance shield excavation in weak geology*. Master's Thesis, Southeast University, Nanjing, China, 2017.
- [36] R. B. Peck, *Deep excavations and tunneling in soft ground*, pp. 225–290, 1969.
- [37] X. Han, N. Li, and J. R. Standing, "An adaptability study of Gaussian equation applied to predicting ground settlements induced by tunnel," *Rock and Soil Mechanics*, vol. 1, pp. 23–28, 2007.
- [38] X. Ding, L. Wang, C. Yang, and P. Tian, "The essential analyses of measures for informatization construction of underground

- pipelines in Foshan,” *Geomatics of Beijing*, vol. 4, pp. 69–71, 2011.
- [39] S. Roger, *Huge sinkholes are now appearing in the wrong places. AP News*, 2017, <https://apnews.com/41a17da407d241769e54085a8118ac7c>.
- [40] D. Bouziou, *Earthquake-induced ground deformation effects on buried pipelines. Ph.D. Thesis*, Cornell University, USA, 2015.
- [41] S. Jeon, “Permanent ground deformation induced by Christchurch earthquake and estimation of underground water pipeline performance in New Zealand,” *Journal of the Korea Academia-Industrial cooperation Society*, vol. 16, no. 6, pp. 4201–4207, 2015.
- [42] T. D. Rourke, J. K. Jung, and C. Argyrou, “Underground pipeline response to earthquake-induced ground deformation,” *Soil Dynamics and Earthquake Engineering*, vol. 91, pp. 272–283, 2016.
- [43] D. Larry and J. Seale, *Creation, analysis, and evaluation of remote sensing sinkhole databases for Florida. Master’s Thesis*, University of South Florida, Tampa, FL, USA, 2005.
- [44] W. Zhou and B. F. Beck, “Management and mitigation of sinkholes on karst lands: an overview of practical applications,” *Environmental Geology*, vol. 55, no. 4, pp. 837–851, 2008.
- [45] A. Theron and J. Engelbrecht, “The role of earth observation, with a focus on SAR interferometry, for sinkhole hazard assessment,” *Remote Sensing*, vol. 10, no. 10, p. 1506, 2018.
- [46] M. Parise and P. Lollino, “A preliminary analysis of failure mechanisms in karst and man-made underground caves in Southern Italy,” *Geomorphology*, vol. 134, no. 1-2, pp. 132–143, 2011.
- [47] S. Richardson, *Sinkhole and subsidence record in the Chuniespoort Group dolomite, Gauteng, South Africa*, University of Pretoria, Pretoria, South Africa, 2013.
- [48] G. A. Momubaghan, *Geotechnical investigations for the Gautrain mass transit rapid link over dolomite bedrock in the Centurion area. Master’s Thesis*, University of Pretoria, Pretoria, South Africa, 2012.
- [49] A. C. Oosthuizen and S. Richardson, “Sinkholes and subsidence in South Africa,” *Council for Geoscience*, vol. 27, pp. 1–31, 2011.
- [50] H. Ali and J. H. Choi, “A review of underground pipeline leakage and sinkhole monitoring methods based on wireless sensor networking,” *Sustainability*, vol. 11, no. 15, p. 4007, 2019.
- [51] H. Ali and J. H. Choi, “Risk prediction of sinkhole occurrence for different subsurface soil profiles due to leakage from underground sewer and water pipelines,” *Sustainability*, vol. 12, p. 310, 2020.

Research Article

Search Space Reduction for Determination of Earthquake Source Parameters Using PCA and k -Means Clustering

Seongjae Lee and Taehyoun Kim 

Department of Mechanical and Information Engineering and Department of Smart Cities, University of Seoul, Seoul 02504, Republic of Korea

Correspondence should be addressed to Taehyoun Kim; thkim@uos.ac.kr

Received 21 July 2020; Revised 14 August 2020; Accepted 23 August 2020; Published 7 September 2020

Academic Editor: Sang-Hoon Hong

Copyright © 2020 Seongjae Lee and Taehyoun Kim. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The characteristics of an earthquake can be derived by estimating the source geometries of the earthquake using parameter inversion that minimizes the L2 norm of residuals between the measured and the synthetic displacement calculated from a dislocation model. Estimating source geometries in a dislocation model has been regarded as solving a nonlinear inverse problem. To avoid local minima and describe uncertainties, the Monte-Carlo restarts are often used to solve the problem, assuming the initial parameter search space provided by seismological studies. Since search space size significantly affects the accuracy and execution time of this procedure, faulty initial search space from seismological studies may adversely affect the accuracy of the results and the computation time. Besides, many source parameters describing physical faults lead to bad data visualization. In this paper, we propose a new machine learning-based search space reduction algorithm to overcome these challenges. This paper assumes a rectangular dislocation model, i.e., the Okada model, to calculate the surface deformation mathematically. As for the geodetic measurement of three-dimensional (3D) surface deformation, we used the stacking interferometric synthetic aperture radar (InSAR) and the multiple-aperture SAR interferometry (MAI). We define a wide initial search space and perform the Monte-Carlo restarts to collect the data points with root-mean-square error (RMSE) between measured and modeled displacement. Then, the principal component analysis (PCA) and the k -means clustering are used to project data points with low RMSE in the 2D latent space preserving the variance of original data as much as possible and extract k clusters of data with similar locations and RMSE to each other. Finally, we reduce the parameter search space using the cluster with the lowest mean RMSE. The evaluation results illustrate that our approach achieves 55.1~98.1% reductions in search space size and 60~80.5% reductions in 95% confidence interval size for all source parameters compared with the conventional method. It was also observed that the reduced search space significantly saves the computational burden of solving the nonlinear least square problem.

1. Introduction

In the past decades, interferometric synthetic aperture radar (InSAR) has been a powerful method to acquire geophysical features such as surface deformation or topography by comparing the phases of at least two complex-valued SAR images obtained from different location or time. Since SAR provides high-resolution images, InSAR can measure the surface deformation with centimetric or even millimetric accuracy. This accurate deformation map enables the observation of ocean and ground surface changes, the measurement of ice

drift and glacier elevations, and the analysis of seismic deformation or volcanic activities [1].

Surface deformation acquired from InSAR measurements provides essential information for studying earthquakes and volcanic activities. To derive the characteristics of an earthquake, we can estimate the source geometries of an earthquake by performing parameter inversion that minimizes the L2 norm of residuals between measured and modeled displacement calculated from a dislocation model [2]. In many cases, the root-mean-square error (RMSE) values of residuals are used for the evaluation of the inversion result.

A dislocation model provides the mathematically calculated coseismic displacement of the earth's surface. The most popular dislocation model is a rectangular dislocation model (Okada model) [3] that assumes isotropic homogeneous half-space and finite rectangular source. Other dislocation models that provide closed analytical expression include the prolate spheroid model (Yang model) [4] and the spherical source model (Mogi model) [5].

In general, since the dislocation model cannot express the surface displacement linearly, estimating source geometries is a nonlinear inverse problem that can be solved by a nonlinear least square method. The nonlinear least square method iteratively searches local minima based on the partial derivatives of the objective function that indicates the misfit between observed and synthetic data. Hence, the solution may vary according to the starting point of parameters, and thus, we cannot assure that the solution is a global minimum [6]. To avoid local minima and describe uncertainties, the Monte-Carlo restarts are often used to solve the problem [7].

In solving a nonlinear least square problem with the Monte-Carlo restarts, we should designate a search space by setting boundary conditions of each parameter to pick random starting points. Since a nonlinear least square method iteratively explores the parameter search space, the search space size significantly affects the accuracy and execution time of the algorithm. Previous studies [8, 9] used the initial range of search space relying on the focal mechanism from seismologists. However, the accuracy of seismological studies can influence the result of an inverse problem. Besides, a dislocation model with many source parameters describing fault leads to bad data visualization and thus analyzing the distribution of the results becomes quite tricky.

This paper presents a new search space reduction algorithm based on machine learning techniques to come up with the challenges mentioned above. Our proposed method proceeds as follows:

- (1) Initially, assume a wide range of search space and perform the Monte-Carlo restarts of five thousand iterations with randomly initialized starting points
- (2) Then, choose the samples that preserve maximum variance and have a low RMSE and project them to the two-dimensional (2D) space using PCA
- (3) Decide hyperparameter k based on the distribution of data on the 2D space and cluster samples using k -means clustering
- (4) Choose the cluster with the lowest mean RMSE and reduce the search space using the cluster
- (5) Perform additional Monte-Carlo restarts of five thousand iterations in the reduced search space and determine the final source parameters

Our search space reduction algorithm was compared with the conventional approach that uses geological studies to define the search space in the same study area, the 2017 Pohang earthquake. The evaluation was performed in terms of the size of search space, RMSE calculated from determined

source parameters, a 95% confidence interval of each source parameter, and the average iteration number of a nonlinear least square.

This paper is organized as follows. Section 2 introduces the related work and background of our approach. In Section 3, our machine learning-based search space reduction algorithm is described in detail. Section 4 presents the results and discussion of our approach in the study area, Pohang, Korea. Finally, Section 5 concludes the paper.

2. Related Work

2.1. Interferometric SAR and Earthquake Source Parameter Determination. InSAR is the method that extracts geophysical features such as surface topography and deformation by comparing the phase offsets of at least two complex-valued SAR images [1]. Comparing SAR images measured at different time instants, InSAR can retrieve the surface displacement at specific time instants, e.g., the surface deformation induced from earthquakes or volcanic activities, landslides [10], thaw-derived slope failure [11], and glacier elevations and changes [12]. However, conventional InSAR has a limitation that it can measure only the line-of-sight (LOS) component of displacement [13]. To overcome the limitation and measure 3D surface displacement accurately, the offset tracking method [14] and stacking InSAR and multiple aperture interferometry (MAI) [15] were proposed.

The source geometries of an earthquake can also be estimated by performing parameter inversion that finds the best approximate source parameters having the minimum misfit between the measured displacement and the synthetic 3D surface displacement of a theoretical dislocation model. Typical dislocation models are as follows. Mogi's model [5] creates a 2D displacement map in radial and tangential direction with four input parameters assuming ideal semi-infinite elastic crust and spherical source. The Yang model [4] generates a 3D displacement map in east, north, and vertical directions with eight source parameters assuming finite prolate spheroid source in an elastic half-space. Okada's model [3] builds a 3D displacement map similar to the Yang model, but it assumes a finite rectangular source and uses ten source parameters. All these dislocation models are nonlinear because the relation between the source parameters and surface displacement cannot be expressed as linear equations.

To estimate the source parameters using a nonlinear dislocation model, we can use the nonlinear least square method that minimizes the Euclidean norm of residuals between the measured and the modeled displacement. Let us denote the dislocation model as G , source parameters as \mathbf{m} , and observed displacement as \mathbf{d} , respectively. Then, the objective function of the nonlinear least square method is formulated as Equation (1) [6].

$$f(\mathbf{m}) = \|G(\mathbf{m}) - \mathbf{d}\|_2. \quad (1)$$

Source parameters \mathbf{m} can have boundary constraints. As for the parameter indicating an angle, the range of parameters is physically limited. For example, the parameter Strike that means the horizontal angle of the fault should be on

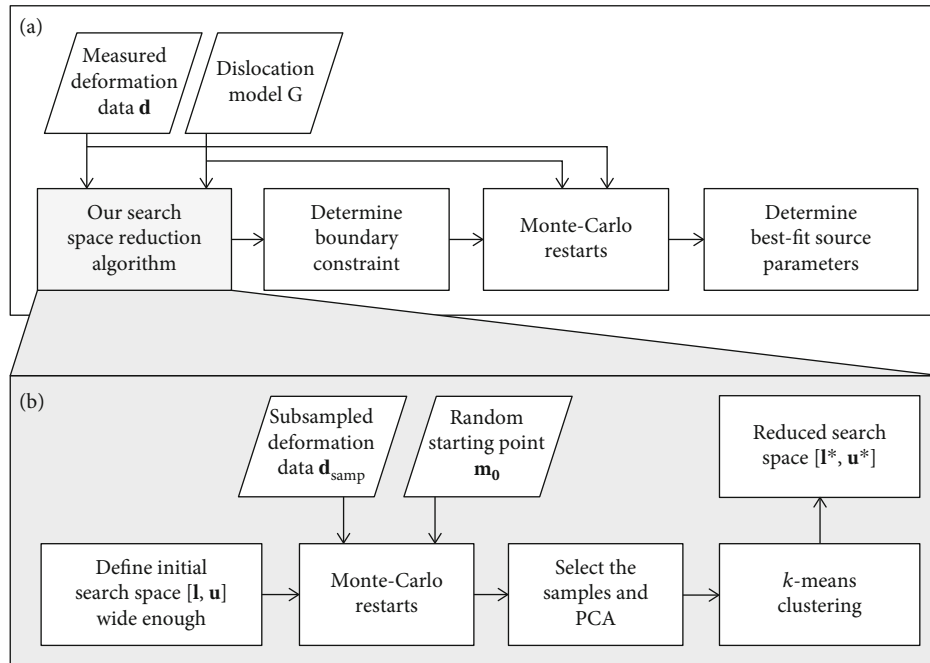


FIGURE 1: Schematic workflow of our approach. (a) is a big workflow that estimates earthquake source parameters. (b) shows our machine learning approach, included in (a), to reduce search space.

TABLE 1: Source parameters of the Okada model.

Parameter	Unit	Description
E	km	Distance from the reference point to the east
N	km	Distance from the reference point to the north
Depth	km	Depth of source
Strike	Degree	Angle of fault relative north
Dip	Degree	Angle between the fault and a horizontal plane
Length	km	Length of fault
Width	km	Width of fault
Rake	Degree	Angle of slip relative to the Width direction
Slip	mm	Dislocation in rake direction
Open	km	Dislocation in tensile component

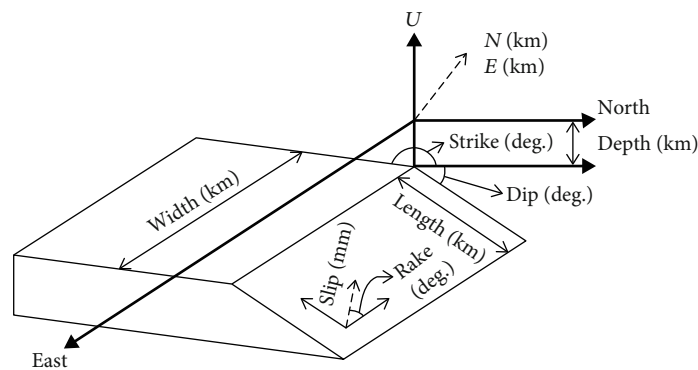


FIGURE 2: Fault geometry of the Okada model.

TABLE 2: Initial search space of source parameters.

	E (km)	N (km)	Depth (km)	Strike (deg.)	Dip (deg.)	Length (km)	Width (km)	Rake (deg.)	Slip (mm)
LB	-40	-40	1	0	0	0.01	0.01	0	-10000
UB	40	40	60	360	90	120	120	180	10000

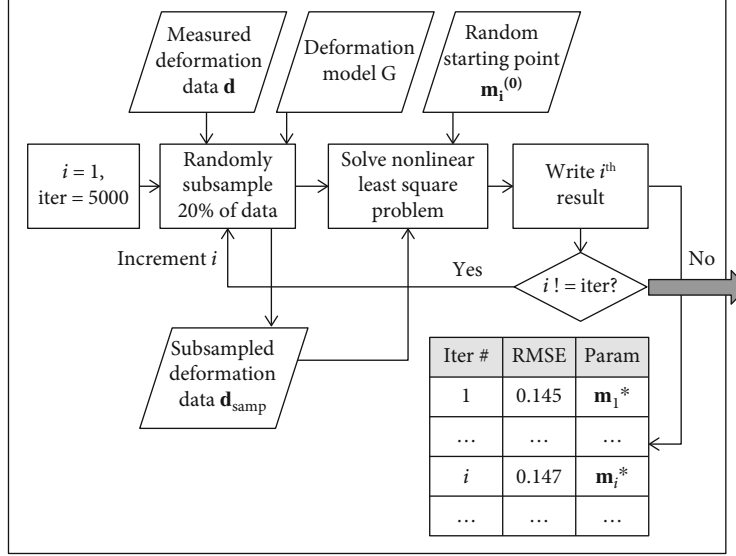


FIGURE 3: Data acquisition in the initial search space using the Monte-Carlo restarts.

TABLE 3: Options of MATLAB lsqnonlin() function.

Option	Value
Algorithm	Trust-region-reflective
MaxIter	500
TolX	0.01
TolFun	0.01

the range $[0^\circ, +360^\circ]$. Also, seismological studies such as the focal mechanism can be used to establish additional boundary conditions. Equation (2) shows the objective function of a nonlinear least square with lower bound \mathbf{l} and upper bound \mathbf{u} .

$$f(\mathbf{m}) = \|\mathbf{G}(\mathbf{m}) - \mathbf{d}\|_2, \quad \mathbf{l} \leq \mathbf{m} \leq \mathbf{u}. \quad (2)$$

The nonlinear least square method iteratively explores the search space from a specific starting point to local minima, so the method cannot derive global minima. To avoid local minima and describe uncertainties of parameters, we can use the Monte-Carlo restarts that generate many initial starting points and then solve the nonlinear least square problem for each starting point. Obtaining results from multiple starting points helps to find the global minimum and show the probabilistic descriptions such as standard deviation and histogram. Many studies employed the Monte-Carlo restarts to obtain the best-fit source parameters of earthquakes or volcanic activities [8, 9, 16–18]. Besides,

Bayesian estimation of source parameters can be used to determine the source parameters [19, 20].

Our approach used stacking InSAR and multiaperture InSAR (MAI) [15] and assumed a wide initial search space. Then, we performed Monte-Carlo restarts for randomly initialized starting points to get data that includes the evidence of the initial search space.

2.2. Machine Learning Applications for Remote Sensing. Machine learning has been used widely in remote sensing [21]. It can be classified into supervised learning and unsupervised learning by whether a given set of training data includes the desired solution called *labels*, or not. In supervised learning, the label y corresponding to the data x exists, and the problem that the data x is provided without the label is called unsupervised learning. Typical supervised learning algorithms include the support vector machine (SVM), random forest (RF), artificial neural network (ANN), and k -nearest neighbor (KNN). In remote sensing, these algorithms have been used for estimating PM2.5 concentrations [22], modeling aboveground biomass of maize [23], and land cover classification [24].

Unsupervised learning techniques have also been used broadly in remote sensing. Principal component analysis (PCA), one of the representatives among them, reduces the dimension of the given dataset. PCA finds a lower dimensional hyperplane that preserves the maximum variance to maintain the maximum amount of original information [25]. In remote sensing, PCA has been used to process hyperspectral images acquired from airborne or spaceborne

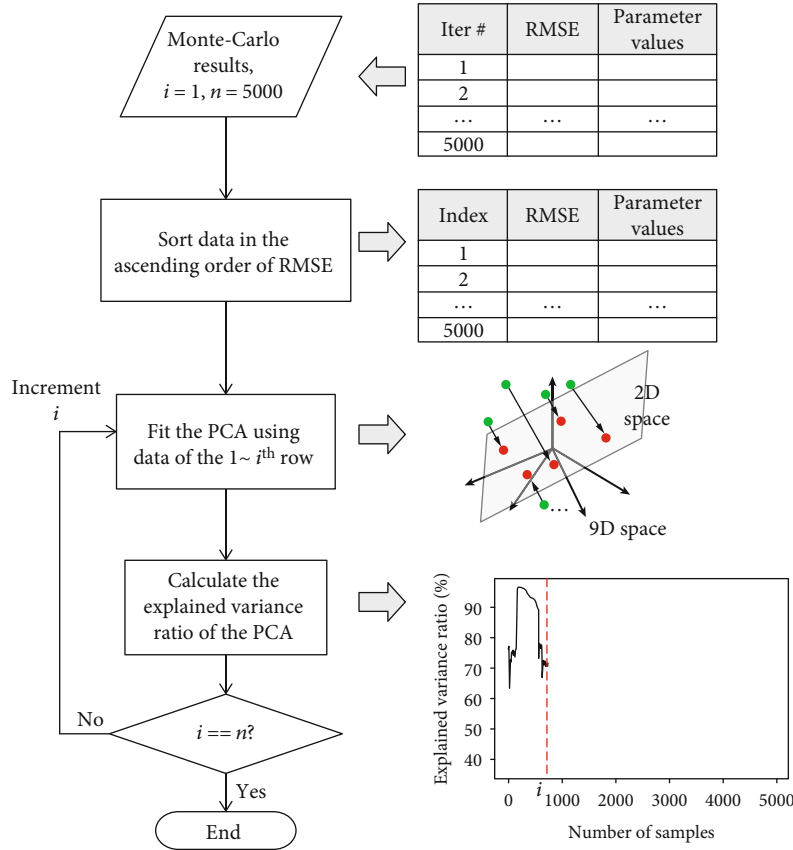


FIGURE 4: PCA procedure for dimensionality reduction of Monte-Carlo results.

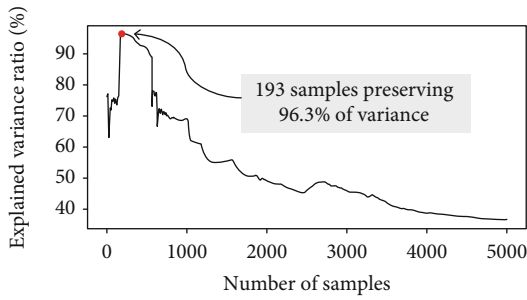


FIGURE 5: Explained variance ratio according to the number of samples.

TABLE 4: Principal components of each parameter.

Parameters	PC 1	PC 2
E	-0.16955	-0.94018
N	-0.34902	0.14051
Depth	0.35540	0.01463
Strike	0.35263	0.05568
Dip	0.34854	-0.01957
Length	-0.35430	0.03954
Width	-0.33501	0.27958
Rake	0.35549	0.04171
Slip	0.33640	-0.10564

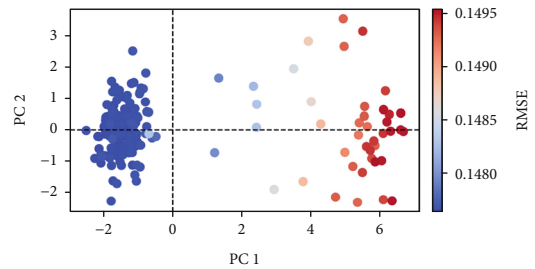


FIGURE 6: Samples projected into 2D space.

vehicles [26–28] and applied in various fields dealing with parameter inversion [29, 30].

k -means clustering is another popular unsupervised learning algorithm that groups data into the k clusters. The parameter k is a hyperparameter that should be defined manually, and data are grouped by minimizing the sum of the squared error of distances between the centroid of the cluster and the data points in the cluster [31]. In this paper, we used PCA to project the data in the initial search space to 2D latent space and performed k -means clustering to cluster the data.

3. Proposed Approach

This paper is aimed at reducing the parameter search space for the determination of earthquake source parameters with

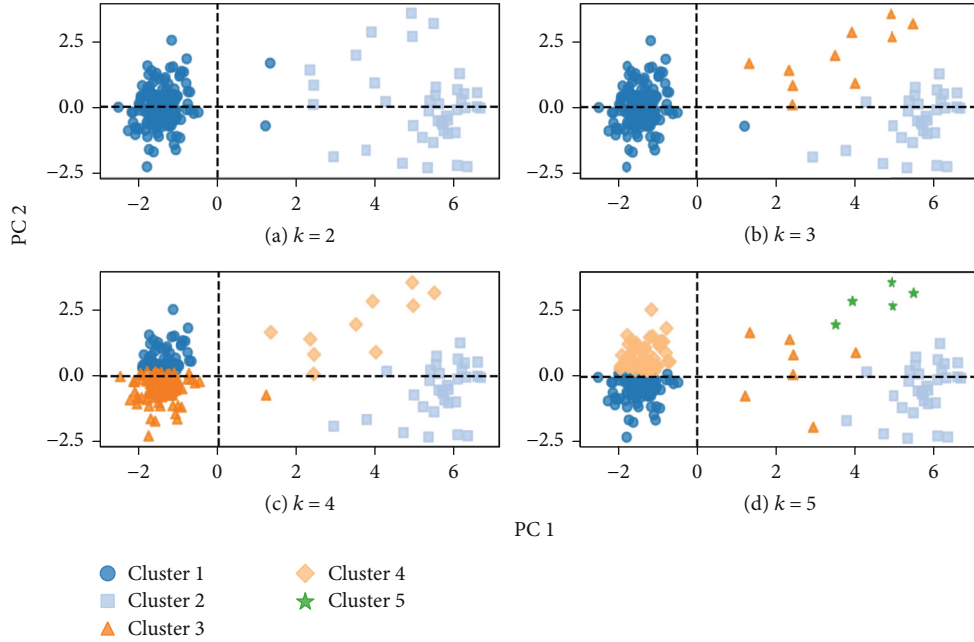


FIGURE 7: Results of k -means clustering for varying k .

TABLE 5: Statistics of each cluster.

Cluster no.	Mean of RMSE (cm)	Standard deviation of RMSE (cm)	Number of samples
1	0.14767	6.08458×10^{-5}	151
2	0.14934	2.13527×10^{-4}	32
3	0.14867	4.96763×10^{-4}	10

better computation time efficiency and data visualization. Figure 1 outlines our proposed search space reduction algorithm. Figure 1(a) illustrates the workflow that determines the source parameters using the Monte-Carlo restarts, and Figure 1(b) presents the proposed method that reduces the parameter search space included in Figure 1(a).

As the first step of our approach, we set a wide initial search space. Second, we perform the Monte-Carlo restarts of five thousand iterations to associate the RMSE and source parameters in the initial search space. In performing Monte-Carlo restarts, we randomly subsample measured deformation to reduce computation time. Third, we reduce the parameter search space using the PCA and the k -means clustering. In doing so, we first apply PCA to choose the samples that preserve the maximum variance of the original information with a low RMSE and perform the k -means clustering with varying k ranging from 2 to 5. Finally, we reduce the search space based on the data in the cluster having the lowest mean RMSE.

3.1. Datasets and Dislocation Model. Our approach is applied to the 2017 Pohang earthquake [32]. To acquire an accurate 3D surface deformation, we applied the stacking InSAR and

the MAI methods [15] to four SAR images obtained from the CSK and ALOS2. The size of the measured displacement is 18.84×15.54 km, and the pixel size is 30 m^2 .

Our study used the Okada model [3] as the dislocation model G. The Okada model assumes a finite rectangular source and isotropic homogeneous half-space. Table 1 describes the physical source parameters defining a rectangular source, and Figure 2 shows the fault geometry of the Okada model, respectively. The parameter E indicates the horizontal, i.e., E-W direction, distance of the upper-left point of the satellite image and fault, and N indicates the vertical one. The upper-left point of the measured deformation map is located at $36^{\circ}10'57''\text{N}$, $129^{\circ}17'03''\text{E}$. We fix the value of parameter Open, i.e., tensile component, to 0 and optimize the other nine parameters to determine the source parameter values.

3.2. Monte-Carlo Restarts in the Initial Search Space. As the first step, our algorithm conducts the Monte-Carlo restarts for a wide initial search space described in Table 2. Since the parameters Strike, Dip, and Rake represent angles, the lower and upper bounds for them are defined as the whole range physically available. Other parameters illustrating the length unit of the fault geometry are defined as large enough to cover the maximum movement of the fault possible in the study area.

Then, the Monte-Carlo restarts of five thousand iterations are performed to extract the [data point, RMSE] pairs that will be used in the next machine learning step. Figure 3 shows the procedure of the Monte-Carlo step. We randomly subsample 20% of measured deformation data \mathbf{d}_{samp} and define loose termination tolerance of the nonlinear least method to save the computational burden. In solving the nonlinear least square problem, we aim to minimize the objective function $\|G(\mathbf{m})_{\text{samp}} - \mathbf{d}_{\text{samp}}\|_2$ to retrieve the best-

TABLE 6: Search space determined from seismological approaches.

Parameters	E (km)	N (km)	Depth (km)	Strike (deg.)	Dip (deg.)	Length (km)	Width (km)	Rake (deg.)	Slip (mm)
LB	4.8	-12.0	4.1	220.0	33.0	4.0	4.0	120.0	10
UB	10.0	-6.0	5.7	235.0	55.0	6.5	6.7	150.0	300

TABLE 7: Reduced search space of three clusters.

Parameters	Cluster 1		Cluster 2		Cluster 3	
	LB	UB	LB	UB	LB	UB
E (km)	6.26911	6.41034	6.27024	6.37718	6.22817	6.3297
N (km)	-7.9826	-7.87204	-8.2007	-8.06074	-8.10795	-7.98743
Depth (km)	4.83734	5.19459	5.28377	5.54763	5.18088	5.49678
Strike (deg.)	224.9835	231.7157	232.4768	236.8167	231.5858	236.5544
Dip (deg.)	42.61413	44.68897	44.99918	46.47305	44.35434	46.34576
Length (km)	4.19294	4.92414	1.13009	3.04056	1.7581	4.19284
Width (km)	5.7689	6.45088	4.67679	5.51209	5.34835	6.08954
Rake (deg.)	127.4311	131.1753	132.5249	135.0445	131.1863	135.0024
Slip (mm)	170	230	330	950	220	520

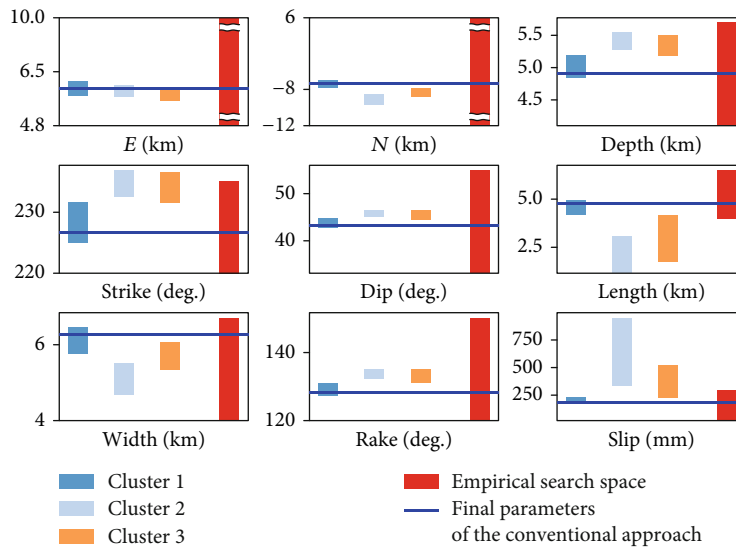


FIGURE 8: Comparison of search spaces between conventional approach and our approach.

TABLE 8: Search space size derived from conventional seismological studies and our proposed algorithm.

Parameters	Conventional approach	Our approach	Reduction ratio (%)
E (km)	5.2	0.1412	97.3
N (km)	6.0	0.1106	98.2
Depth (km)	1.6	0.3573	77.7
Strike (deg.)	15.0	6.7322	55.1
Dip (deg.)	22.0	2.0748	90.6
Length (km)	2.5	0.7312	70.8
Width (km)	2.7	0.6820	74.7
Rake (deg.)	30.0	3.7442	87.5
Slip (mm)	290	60	79.3

fit parameters with minimum residuals. As a result, optimization result \mathbf{m}_i^* and corresponding RMSE are calculated in each iteration. The nonlinear least square problem is solved by using the `lsqnonlin()` function of the MATLAB R2019a with the trust-region-reflective algorithm [33]. Table 3 specifies the option parameters used in the function. Option parameters not specified are set to the default values.

3.3. Search Space Reduction Using Machine Learning Approach. This study uses the PCA and the k -means clustering to reduce the search space. At first, we perform the PCA to reduce the dimensionality of the Monte-Carlo results, as described in Figure 4. Before the PCA fitting, the results obtained from the Monte-Carlo step are sorted in the ascending order of the RMSE value to select well-optimized

TABLE 9: Determined parameters and RMSE of conventional approach and our approach with a 95% confidence interval.

Parameters	Conventional approach	Our ML approach
E (km)	6.346 ± 0.098	6.341 ± 0.034
N (km)	-7.904 ± 0.107	-7.922 ± 0.029
Depth (km)	4.921 ± 0.172	4.901 ± 0.052
Strike (deg.)	226.563 ± 3.276	226.364 ± 1.059
Dip (deg.)	43.162 ± 0.985	43.04 ± 0.394
Length (km)	4.825 ± 0.465	4.786 ± 0.123
Width (km)	6.288 ± 0.724	6.21 ± 0.141
Rake (deg.)	128.424 ± 2.35	128.373 ± 0.65
Slip (mm)	175.243 ± 36.221	177.618 ± 7.417
RMSE (cm)	0.147637	0.147634

samples. Then, we perform the PCA fitting iteration for the sorted data in the index range of $[1, i]$ where $1 \leq i \leq 5000$ and calculate the explained variance ratio for the PCA results. The explained variance ratio is the ratio of the variance of projected data on the reduced dimensions to the total variance of the original information. Also, the explained variance ratio is regarded as a metric to evaluate the usefulness of the principal components extracted by the PCA, and thus a higher explained variance ratio means the information loss is kept smaller. For this reason, our algorithm chooses the samples showing the highest explained variance ratio for further processing.

PCA enables visualization of high-dimensional data in the low-dimensional space while preserving information of the original data as much as possible. In general, a 2D scatter matrix is used to visualize high-dimensional data, but each element of the scatter matrix represents only partial information. Furthermore, since the number of two-dimensional

scatter plots of p -dimensional parameters is $\binom{p}{2} = p(p-1)$,

visualizing high-dimensional data using a scatter matrix is cumbersome [34]. For example, visualizing the results of Monte-Carlo restarts by a scatter matrix requires 36 plots for the Okada model with nine parameters. To overcome this problem, this paper uses the PCA visualizing the nine-dimensional data on the 2D hyperplane preserving the original information so that we can analyze the [data, RMSE] pairs with just one 2D scatter plot.

For the data projected to the 2D space, we cluster the data using the k -means clustering. k -means clustering identifies k number of centroids and allocates data points to the nearest cluster, where k is a hyperparameter that a user should define manually. Our algorithm applies k -means clustering to the data on the 2D plane for varying k to find the most appropriate k that well divides projected samples with similar RMSE and position to each other. Once deciding the parameter k , we reduce the search space for each source parameter using the data cluster showing the lowest mean RMSE for the given k .

If the PCA preserves most of the variance of the Monte-Carlo results and all source parameters are on the same scale,

the distances between original data points in the latent 2D space can be directly assessed [35]. It means that the k -means clustering using the projected data can group the data in the adjacent location of the original nine-dimensional space because this algorithm minimizes the squared error between the centroid of the cluster and data points in the cluster. In summary, by using the PCA and the k -means clustering, we can extract the similarity of the nine-dimensional data on the 2D space with better data visualization.

Before the PCA fitting, each parameter is scaled to have a standard normal distribution $N(0, 1)$. PCA finds the principal components maximizing the variance. If parameters are unscaled, the principal component loading vector will be biased to a parameter showing a large variance. For example, the PCA fitting for the unscaled Monte-Carlo results may lead the principal component to have a significant loading for the Strike parameter that has the largest parameter search space among the source parameters of the Okada model. Furthermore, the distances between data points in the 2D space cannot be assessed. In this paper, we implemented our machine learning-based algorithm using the Scikit-learn package [36] in Python.

4. Experimental Results and Discussion

To evaluate the effectiveness of our machine learning-based search space reduction algorithm, this paper applied the proposed machine learning-based algorithm to the 2017 Pohang earthquake case [32]. The evaluation was performed in terms of the size of search space, RMSE values derived from determined source parameters, a 95% confidence interval of each source parameter, and the number of iterations for each non-linear least square of the Monte-Carlo restarts. The results were compared with those of the conventional approach, which refers to the seismological studies.

4.1. Principal Component Analysis and k -Means Clustering.

As mentioned in Section 3, we first conducted dimensionality reduction using PCA from nine-dimensional source parameters of the Okada model to the 2D space. Figure 5 illustrates the explained variance ratio for the number of samples used for PCA fitting. As described in Section 3, the data samples were presorted in the ascending order of the RMSE values. The result presents that the samples with high RMSE values interfere with the dimensionality reduction. The result also shows that the number of samples equals 193 produces the highest explained variance ratio, i.e., 96.3%. It means that the projection results with only 193 samples maintain most of the variance of the original data while keeping a low RMSE value, so the relation between those samples in the 2D space can be directly evaluated.

Table 4 shows the principal components (PCs) of each parameter, and Figure 6 presents the data projected into 2D space, respectively. With PCA, we can easily visualize the distribution of samples with only one 2D scatter plot. As shown in Figure 6, the samples with low RMSE values are clustered to the left side, while samples with high RMSE values are distributed from the center of the plot to the right side.

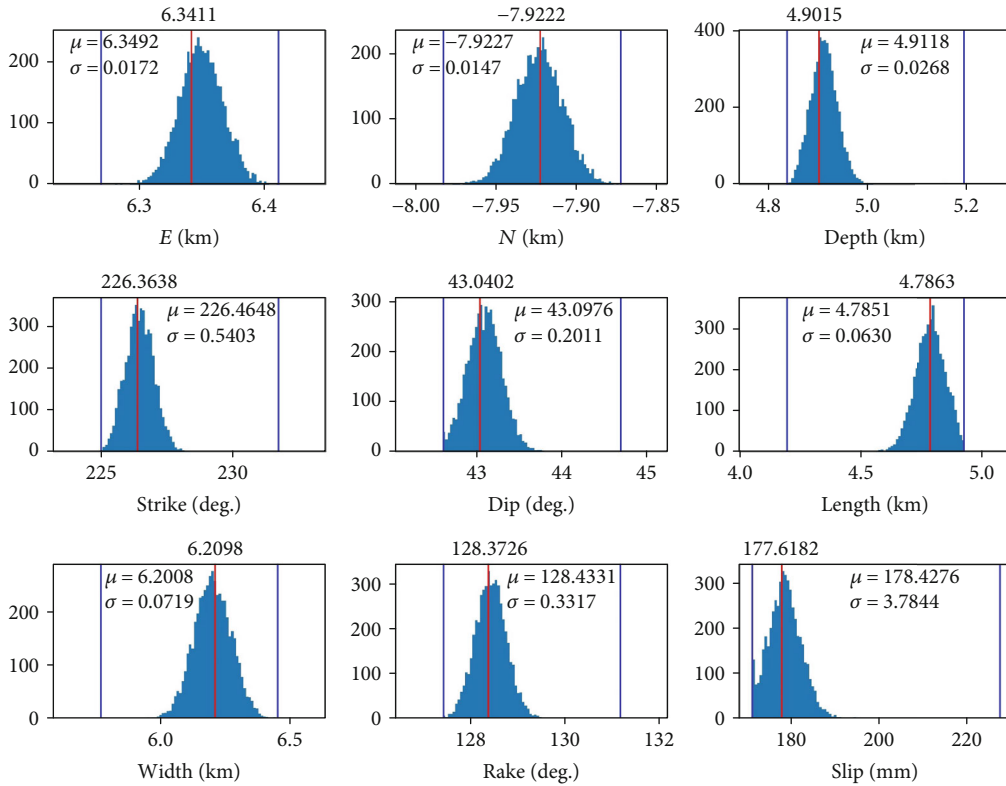


FIGURE 9: Histogram of determined parameters.

Then, we performed k -means clustering for $k = 2 \sim 5$ on the 2D hyperplane. Figures 7(a)–7(d) illustrate the k -means clustering results for each k , respectively. It was observed that the result for $k = 2$ could not distinguish the samples in the upper center with medium RMSE from those in the bottom right side with high RMSE. The results for $k = 4, 5$ also divide low-RMSE samples in the left corner. As a result, in our test case, data samples are best partitioned when $k = 3$. Table 5 also shows the mean RMSE and standard deviation of each cluster obtained for $k = 3$.

4.2. Search Space Reduction. We then reduced the search space using the minimum and maximum values of data in each cluster. Tables 6 and 7 show the search space derived from the seismological studies [37] and three reduced search spaces determined from our proposed algorithm for each source parameter. Figure 8 illustrates the results that compare the reduced search spaces derived from our approach with the empirically defined search space referring to seismological approaches. In Figure 8, the red bar shows empirically defined search space, and the others illustrate the search spaces determined by the three clusters. The blue horizontal line denotes the determined best-fit source parameters acquired by Monte-Carlo restarts of ten thousand iterations in the empirically defined search space. Figure 8 shows that, for all the parameters, the reduced search spaces for Cluster 1 include the source parameters determined from the Monte-Carlo restarts on the empirical parameter search. In other words, the reduced search space having the lowest mean RMSE is in good agreement with the conventional

research results based on seismological studies while effectively reducing the search space. For this reason, the rest of the experiments were conducted using the search space from Cluster 1.

Table 8 summarizes the reduction of search space for each source parameter after applying the proposed approach. From the results, the search space size for each parameter decreased by about 55.1~98.1%, compared with the search space from the conventional approach.

4.3. Determination of Source Parameters. As the final step, we conducted additional five thousand iterations of Monte-Carlo restarts in the reduced search space based on Cluster 1 and determined the source parameters as the best-fit parameter of the results of the Monte-Carlo restarts with a 95% confidence interval. Table 9 summarizes the source parameters determined and RMSE of our approach and the conventional approach. The results from two approaches seem similar to each other. However, it can be remarked that our machine learning approach produces more reliable results than the conventional one because a 95% confidence interval of our approach is much smaller than the conventional approach. The interval size reduction was about 60~80.5% for each parameter.

Figure 9 illustrates the distribution of the Monte-Carlo results derived from reduced search space using our machine learning-based approach. Two blue vertical lines of each histogram represent the lower and upper bounds of each parameter, and the red vertical line indicates the best-fit source parameter. In the reduced search space, all parameters

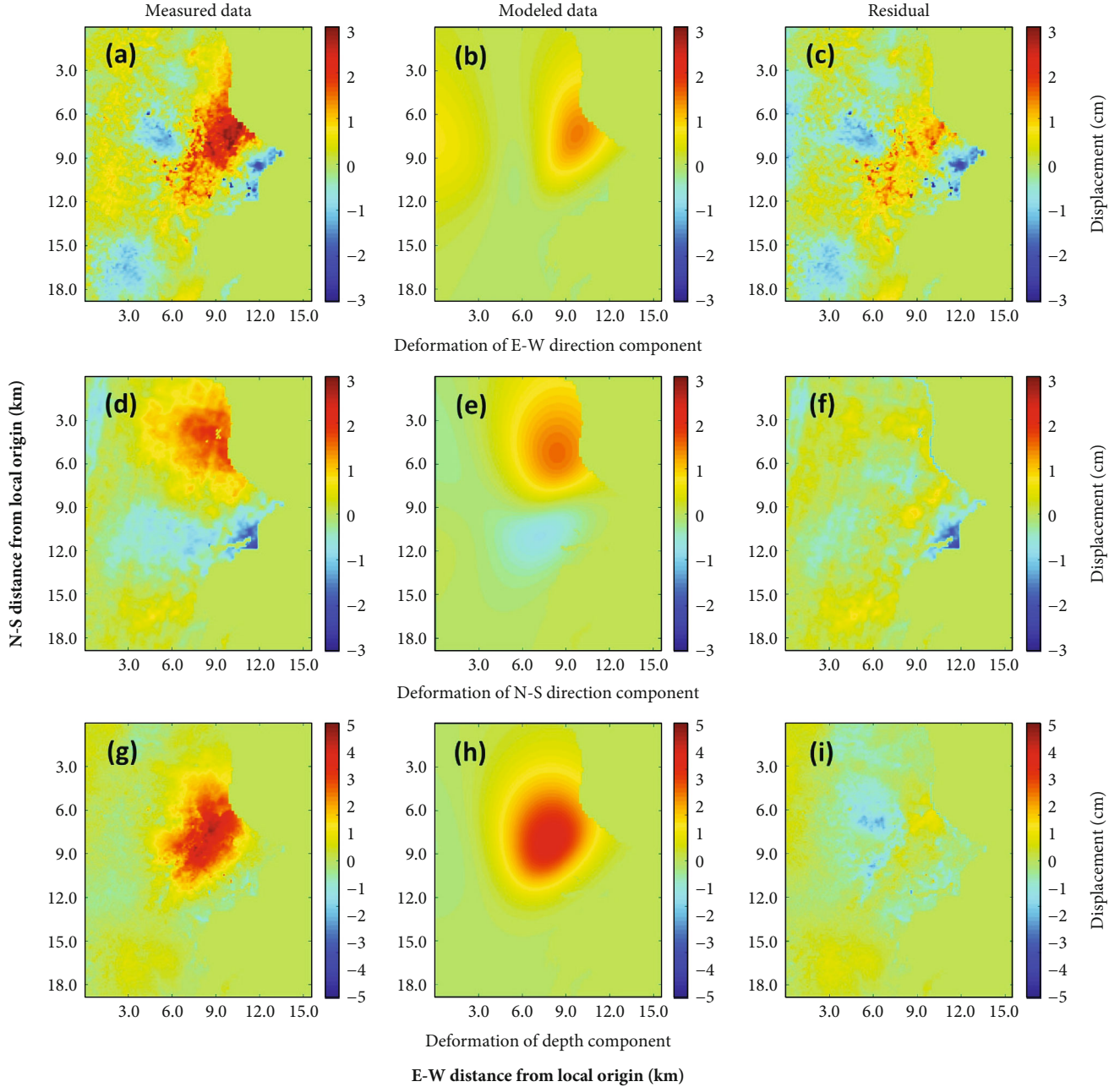


FIGURE 10: Final results of source modeling.

TABLE 10: Average number of iterations for nonlinear least square from the Monte-Carlo restarts.

Conventional approach	Our approach
4.93	2.25

feature a unimodal histogram fit in the search space appropriately with small standard deviations.

Figure 10 compares the synthetic model's deformation maps using the parameters acquired from our approach with those obtained from InSAR measurements. Each row of Figure 10 describes the deformation of E-W, N-S, and depth

direction component. The first column and the second column in Figure 10 show the measured surface deformation map for each direction component using InSAR and modeled deformation map from the Okada model. The third column shows a residual between the measured displacements and the modeled displacements. It can be concluded that the determination of source parameters using the Okada model and our reduced search space fit the measured displacement correctly because RMSE values for E-W, N-S, and depth components were 0.1863, 0.1544, and 0.0829 cm.

4.4. Computational Efficiency. For the computational efficiency validation, the average iterations of a nonlinear least

square from the Monte-Carlo restarts were also compared. As shown in Table 10, the average iterations of a nonlinear square for the reduced search space produced by our approach are much smaller than that of the conventional approach.

5. Conclusions and Future Works

This paper presented a new search space reduction algorithm using machine learning techniques for the earthquake source parameter determination. Our algorithm proceeded in the order of the Monte-Carlo restarts, PCA, k -means clustering, and search space reduction. PCA leads to better visualization of the Monte-Carlo results using only one 2D scatter plot, so we can find the hyperparameter k of the k -means clustering that groups data for them to have similar RMSE and close location to each other.

We compared the proposed approach with the conventional approach that defines the parameter search space by referring to seismological studies in terms of size of the search space, RMSE, a 95% confidence interval of Monte-Carlo results, and the average iteration number of nonlinear least square from the Monte-Carlo restarts. The experimental results for the 2017 Pohang earthquake test case showed that our approach achieves significant reductions in search space size and a 95% confidence interval size for all source parameters, i.e., about 55.1~98.1% and 60~80.5%, respectively. Also, the average iteration number of a nonlinear least square from our reduced search space was much smaller than that of a conventionally defined search space. Consequently, the results demonstrated that the proposed approach effectively reduces the parameter search space size and the computational burden.

Our future challenge is to develop the automatic search space reduction algorithm that uses the hierarchical clustering algorithm for determining the number of clusters.

Data Availability

The surface deformation maps were provided by Prof. Jung and Mr. Baek. They have generated the maps from the ALOS PALSAR-2 interferometric data that were provided through the JAXA's ALOS-2 research program (RA4, PI No. 1412).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the 2019 sabbatical year research grant of the University of Seoul.

References

- [1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathan, "A tutorial on synthetic aperture radar," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 1, pp. 6–43, 2013.
- [2] K. L. Feigl, "Estimating earthquake source parameters from geodetic measurements," in *International Handbook of Earthquake and Engineering Seismology*, vol. 81A, pp. 1–14, Elsevier, 2002.
- [3] Y. Okada, "Surface deformation due to shear and tensile faults in a half-space," *Bulletin of the Seismological Society of America*, vol. 75, no. 4, pp. 1135–1154, 1985.
- [4] X.-M. Yang, P. M. Davis, and J. H. Dieterich, "Deformation from inflation of a dipping finite prolate spheroid in an elastic half-space as a model for volcanic stressing," *Journal of Geophysical Research: Solid Earth*, vol. 93, no. B5, pp. 4249–4257, 1988.
- [5] K. Mogi, *Relations between the eruptions of various volcanoes and the deformations of the ground surfaces around them*, Earthquake Research Institute, 1958.
- [6] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems*, Elsevier, 2nd edition, 2018.
- [7] M. Sambridge and K. Mosegaard, "Monte Carlo methods in geophysical inverse problems," *Reviews of Geophysics*, vol. 40, no. 3, 2002.
- [8] W. Qu, B. Zhang, Z. Lu et al., "Source parameter estimation of the 2009 Ms6.0 Yao'an Earthquake, Southern China, using InSAR observations," *Remote Sensing*, vol. 11, no. 4, p. 462, 2019.
- [9] M.-J. Jo, H.-S. Jung, and J.-S. Won, "Detecting the source location of recent summit inflation via three-dimensional InSAR observation of Kilauea volcano," *Remote Sensing*, vol. 7, no. 11, pp. 14386–14402, 2015.
- [10] Y. Kang, C. Zhao, Q. Zhang, Z. Lu, and B. Li, "Application of InSAR techniques to an analysis of the Guanling landslide," *Remote Sensing*, vol. 9, no. 10, p. 1046, 2017.
- [11] B. Hu, Y. Wu, X. Zhang et al., "Monitoring the thaw slump-derived thermokarst in the Qinghai-Tibet Plateau using satellite SAR interferometry," *Journal of Sensors*, vol. 2019, Article ID 1698432, 8 pages, 2019.
- [12] Y. Sun, L. Jiang, L. Liu, Q. Sun, H. Wang, and H. Hsu, "Mapping glacier elevations and their changes in the western Qilian Mountains, northern Tibetan Plateau, by bistatic InSAR," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 1, pp. 68–78, 2016.
- [13] J. Hu, Z. W. Li, X. L. Ding, J. J. Zhu, L. Zhang, and Q. Sun, "Resolving three-dimensional surface displacements from InSAR measurements: a review," *Earth-Science Reviews*, vol. 133, pp. 1–17, 2014.
- [14] M. Furuya and T. Yasuda, "The 2008 Yutian normal faulting earthquake (Mw 7.1), NW Tibet: non-planar fault modeling and implications for the Karakax fault," *Tectonophysics*, vol. 511, no. 3–4, pp. 125–133, 2011.
- [15] H. S. Jung, Z. Lu, J. S. Won, M. P. Poland, and A. Miklius, "Mapping three-dimensional surface deformation by combining multiple-aperture interferometry and conventional interferometry: application to the June 2007 eruption of Kilauea volcano, Hawaii," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 1, pp. 34–38, 2011.
- [16] M.-J. Jo, H.-S. Jung, and S.-H. Yun, "Retrieving precise three-dimensional deformation on the 2014 M6.0 South Napa earthquake by joint inversion of multi-sensor SAR," *Scientific Reports*, vol. 7, no. 1, p. 5485, 2017.
- [17] G. Diczelis, M. Assumpção, J. Kellogg, P. Pedraza, and F. Dias, "Estimating the 2008 Quetame (Colombia) earthquake source parameters from seismic data and InSAR measurements,"

- Journal of South American Earth Sciences*, vol. 72, pp. 250–265, 2016.
- [18] V. De Novellis, R. Castaldo, C. De Luca et al., “Source modeling of the 2015 Wolf volcano (Galápagos) eruption inferred from Sentinel 1-A DInSAR deformation maps and pre-eruptive ENVISAT time series,” *Journal of Volcanology and Geothermal Research*, vol. 344, pp. 246–256, 2017.
- [19] M. Bagnardi and A. Hooper, “Inversion of surface deformation data for rapid estimates of source parameters and uncertainties: a Bayesian approach,” *Geochemistry, Geophysics, Geosystems*, vol. 19, no. 7, pp. 2194–2211, 2018.
- [20] R. Dutta, S. Jónsson, T. Wang, and H. Vasyura-Bathke, “Bayesian estimation of source parameters and associated Coulomb failure stress changes for the 2005 Fukuoka (Japan) earthquake,” *Geophysical Journal International*, vol. 213, no. 1, pp. 261–277, 2017.
- [21] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, “Machine learning in geosciences and remote sensing,” *Geoscience Frontiers*, vol. 7, no. 1, pp. 3–10, 2016.
- [22] G. Chen, S. Li, L. D. Knibbs et al., “A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information,” *Science of the Total Environment*, vol. 636, pp. 52–60, 2018.
- [23] L. Han, G. Yang, H. Dai et al., “Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data,” *Plant Methods*, vol. 15, no. 1, p. 10, 2019.
- [24] P. T. Noi and M. Kappas, “Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery,” *Sensors*, vol. 18, no. 1, p. 18, 2017.
- [25] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O’Reilly Media, 2019.
- [26] A. Agarwal, T. El-Ghazawi, H. El-Askary, and J. Le-Moigne, “Efficient hierarchical-PCA dimension reduction for hyperspectral imagery,” in *2007 IEEE International Symposium on Signal Processing and Information Technology*, pp. 353–356, Giza, Egypt, 2007.
- [27] X. Kang, X. Xiang, S. Li, and J. A. Benediktsson, “PCA-Based edge-preserving features for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 12, pp. 7140–7151, 2017.
- [28] Z. Wu, Y. Huang, and K. Zhang, “Remote sensing image fusion method based on PCA and curvelet transform,” *Journal of the Indian Society of Remote Sensing*, vol. 46, no. 5, pp. 687–695, 2018.
- [29] M. Zuo, G. Dai, L. Peng, L. Chen, X. Chen, and Z. Song, “Global optimisation of multiple gravity assist spacecraft trajectories based on search space exploring and PCA,” in *2016 IEEE Congress on Evolutionary Computation (CEC)*, pp. 2655–2660, Vancouver, BC, Canada, 2016.
- [30] S. Karimi, N. Samani, and Z. Mohammadi, “Characterization of Semnan thermal springs using principal component analysis and geochemical inverse modeling,” *Arabian Journal of Geosciences*, vol. 12, no. 24, p. 777, 2019.
- [31] A. K. Jain, “Data clustering: 50 years beyond K-means,” *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.
- [32] K.-H. Kim, J.-H. Ree, Y. H. Kim, S. Kim, S. Y. Kang, and W. Seo, “Assessing whether the 2017Mw5.4 Pohang earthquake in South Korea was an induced event,” *Science*, vol. 360, no. 6392, pp. 1007–1009, 2018.
- [33] T. F. Coleman and Y. Li, “An interior trust region approach for nonlinear minimization subject to bounds,” *SIAM Journal on Optimization*, vol. 6, no. 2, pp. 418–445, 1996.
- [34] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, Springer, 2013.
- [35] R. Bro and A. K. Smilde, “Principal component analysis,” *Analytical Methods*, vol. 6, no. 9, pp. 2812–2831, 2014.
- [36] F. Pedregosa, G. Varoquaux, A. Gramfort et al., “Scikit-learn: machine learning in Python,” *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [37] K.-K. Lee, *Final report of the Korean Government Commission on relations between the 2017 Pohang earthquake and EGS project*, The Geological Society of Korea, 2019.

Research Article

Adaptive Magnetic Anomaly Detection Method with Ensemble Empirical Mode Decomposition and Minimum Entropy Feature

Liming Fan ^{1,2}, Chong Kang,³ Huigang Wang,¹ Hao Hu,¹ and Mingliang Zou⁴

¹Qingdao Research Institute, Northwestern Polytechnical University, Qingdao 266200, China

²Key Laboratory of Marine Intelligent Equipment and System of Ministry of Education, Shanghai Jiao Tong University, Shanghai 200240, China

³College of Science, Harbin Engineering University, Harbin 150001, China

⁴Changsha Uranium Geology Research Institute, CNNC, Changsha 41001, China

Correspondence should be addressed to Liming Fan; limingfan@nwpu.edu.cn

Received 21 April 2020; Revised 30 June 2020; Accepted 6 August 2020; Published 28 August 2020

Academic Editor: Sang-Hoon Hong

Copyright © 2020 Liming Fan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the fast attenuation of the magnetic field along with the distance, the magnetic anomaly generated by the remote magnetic target is usually buried in the magnetic noise. In order to improve the performance of magnetic anomaly detection (MAD) with low SNR, we propose an adaptive method of MAD with ensemble empirical mode decomposition (EEMD) and minimum entropy (ME) feature. The magnetic data is decomposed into the multiple intrinsic modal functions (IMFs) with different scales by EEMD. According to a defined criterion, the magnetic noise and magnetic signal are reconstructed based on IMFs, respectively. Entropy feature of reconstructed magnetic signal is extracted based on the probability density function (PDF) of the noise which is updated by the reconstructed magnetic noise. Compared to the traditional minimum entropy method, the entropy feature extracted by the proposed method is more obvious. The magnetic anomaly is detected whenever the entropy feature drops below the threshold. Thus, it is effective for revealing the weak magnetic anomaly by the proposed method. The measured magnetic noise is used to validate the performance of the proposed method. The results show that the detection probability of the proposed method is higher with low input SNR.

1. Introduction

The magnetic field generated by a magnetic target, which changes the distribution of the ambient magnetic field around it, is called the magnetic anomaly. Magnetic anomaly detection is a passive method of detecting the ferromagnetic objects. It has been used in many areas, such as unexploded ordnance detection [1–3], underwater pipeline or cable detection [4, 5], and human medical investigation [6, 7]. Due to fast attenuation of the magnetic field along with the distance, the magnetic anomaly generated by the remote target is usually buried in the magnetic noise [8, 9]. Thus, we need the effective method to improve the performance of magnetic anomaly detection.

Recently, researchers have proposed several methods of magnetic anomaly detection [10–19]. These methods can be divided into two categories. One category is to improve the SNR of the anomaly signal by the noise suppression [20–23]. Zhou et al. [20] proposed a trend filtering based on empirical mode decomposition for MAD. Wavelet denoising and the Karhunen-Loeve expansion have been proposed to detect the magnetic anomaly. Liu et al. [8] proposed an adaptive coherent noise suppression method using coherence. However, those methods are not sufficient for magnetic anomaly detection when $\text{SNR} < 1$. For this situation, Ginzburg et al. [22] proposed a method of MAD using three orthonormal basis functions (OBF), which relies on the matched filtering. Fan et al. [24] proposed an improved

method using four orthonormal basis functions, which can not only detect magnetic anomaly but also determine the orientation of the target. These methods are optimal for detecting the magnetic anomaly buried in the Gaussian white noise. In some situations, the geomagnetic noise is considered the noise with a power spectral density (PSD) of $1/f^\alpha$, where $0 < \alpha < 2$. In order to handle the noise with PSD of $1/f^\alpha$, Sheinker et al. [25] designed a whitening filter, which improved the performance of MAD. This method can be considered an autoregressive (AR) process and work effectively with a high order of filter. The other category is to reveal the magnetic anomaly by the transformation. The minimum entropy method transforms the measured data into entropy value to detect the anomaly [12]. Because the most common-mode geomagnetic noise can be removed by the differential signal, the method based on the information entropy of the differential signal was more appropriate for the target detection in the complex environment [17]. The above methods need the PDF of the magnetic noise. The high-order crossing method is an alternative method for spectral analysis using zero-crossing count [14]. The advantages of these methods are easy to implement and low computational complexity. However, the detection performance of these methods may be limited by low SNR of input signal. In addition, some methods based on machine learning are proposed for magnetic anomaly detection [26–30].

In order to improve the detection performance in the case of low SNR and the complex magnetic environment, we propose an adaptive method of MAD with ensemble empirical mode decomposition and minimum entropy (EEMD-ME) feature in this paper. Meanwhile, the similar methods with empirical mode decomposition and minimum entropy (EMD-ME) feature and complete ensemble empirical mode decomposition with adaptive noise and minimum entropy (CEEMDAN-ME) feature are also implemented. In the methods, the measured magnetic data is decomposed into multiple IMFs with different scales. According to the characteristics of the IMFs, the parameters of PDF of magnetic noise are updated in real time using the reconstructed magnetic noise. Then, the entropy feature of the reconstructed magnetic signal is calculated in a moving window with the updated PDF of the noise. Finally, the magnetic anomaly is detected whenever the entropy feature drops below the threshold. In brief, there are three advantages of the proposed method: (1) the parameters of PDF of the noise are updated in real time; (2) SNR of input signal is improved; and (3) low computational complexity is implemented in the detection. Thus, the proposed method has better robustness, higher detection probability in the weak magnetic anomaly detection.

2. Detection Theory

2.1. EMD, Ensemble EMD, and Complete Ensemble EMD with Adaptive Noise. Empirical mode decomposition is pioneered by Huang for adaptively processing the nonlinear and nonstationary signals [31]. It decomposes the signal into a series of IMFs with different time scales, which must satisfy two conditions: (1) the number of extreme points in IMF is

equal to or not more than that of one zero-crossing point; (2) the mean of the envelope determined by the maximum and minimum values is zero. The algorithm can be described as follows:

Step 1. Set $k=0$ and find all the maxima (minima) points of the original signal $r_0 = x(t)$. Interpolate between maxima (minima) points to obtain the upper (lower) envelope $r_k^+(t)$ ($r_k^-(t)$)

Step 2. Calculate the mean envelope $m_k(t) = (r_0^+(t) + r_0^-(t))/2$

Step 3. Obtain the IMF candidate $c_k(t) = r_k(t) - m_k(t)$

Step 4. Determine whether the candidate $c_k(t)$ satisfy two conditions:

- (i) Yes, save $d_{k+1}(t)$ as an IMF and calculate the residual $r_{k+1}(t) = x(t) - \sum_{i=1}^k c_i$. Do $k = k + 1$ and treat $r_{k+1}(t)$ as the input signal in Step 2
- (ii) No, treat $d_{k+1}(t)$ as the input signal in Step 2.

Step 5. Repeat Steps 2-4 until the residual $r_{k+1}(t)$ satisfies the stopping criterion.

In order to alleviate the “mode mixing” phenomenon in EMD, noise-assisted versions have been proposed, such as EEMD [32, 33] and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [34–36]. The main idea of the noise-assisted versions is that the different finite variance white noise $n_a(t)$ plus the original signal and the average of the corresponding IMFs is calculated as the final IMFs.

The algorithm of EEMD is described as follows:

Step 1. Add the finite variance white noise $n^i(t)$ ($i = 1, \dots, I$) to the original signal with I times and obtain the noise-added signal: $x^i(t) = x(t) + \varepsilon n_a^i(t)$, where ε denotes the standard deviation of added noise amplitude and $\varepsilon > 0$

Step 2. Decompose the noise-added signal $x^i(t)$ into IMFs $c_k^i(t)$ and a residual $r^i(t)$ by EMD

Step 3. Calculate the average of each IMF and the residual as the final IMF and residual: $c_k(t) = 1/I \sum_{i=1}^I c_k^i(t)$ and $r(t) = 1/I \sum_{i=1}^I r^i(t)$.

For CEEMDAN, let $E_k(\bullet)$ be the k th mode of a given signal decomposed by EMD. The algorithm of CEEMDAN [36] is given as follows:

Step 1. Add the finite variance white noise $n_a^i(t)$ ($i = 1, \dots, I$) to the original signal with I times and obtain the noise-added signal: $x^i(t) = x(t) + \varepsilon_0 n_a^i(t)$

Step 2. Decompose the noise-added signal $x^i(t)$ by EMD and obtain the first IMF modes and residual: $c_1(t) = 1/I \sum_{i=1}^I c_1^i(t)$ and $r_1(t) = x(t) - c_1(t)$

Step 3. Decompose the signal $r_1(t) + \varepsilon_1 E_1(n_a^i(t))$ by EMD to obtain the first IMF mode. Define the second IMF modes as $c_2(t) = 1/I \sum_{i=1}^I E_1(r_1(t) + \varepsilon_1 E_1(n_a^i(t)))$

Step 4. For $k = 2, \dots, K$, calculate the residual as $r_k(t) = r_{k-1}(t) - c_k(t)$

Step 5. Decompose the signal $r_k(t) + \varepsilon_k E_k(n_a^i(t))$ by EMD to obtain the first IMF mode. Define the $(k+1)$ th modes as: $c_{k+1}(t) = 1/I \sum_{i=1}^I E_k(r_k(t) + \varepsilon_k E_k(n_a^i(t)))$

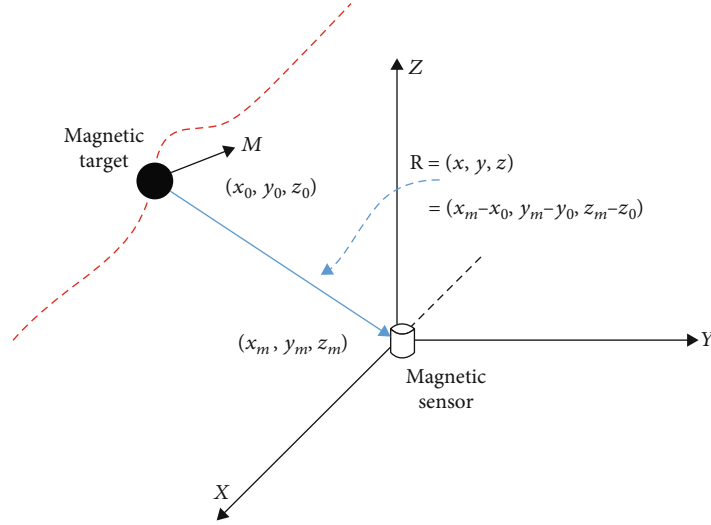


FIGURE 1: Magnetic anomaly detection model.

Step 6. Repeat Steps 4-5 until the residual $r_{k+1}(t)$ satisfies the stopping criterion.

Thus, the signal $x(t)$ can be described by the IMF components and a residual component as follows:

$$x(t) = \sum_{k=1}^K c_k(t) + r(t) \quad (1)$$

where $c_k(t)$ denotes the k th IMF component and $r(t)$ denotes the residual component.

Generally, the IMF components are sorted from high frequency to low frequency. Therefore, there will be a component indexed by h , which can be considered the division of energy distribution from noise to signal. After the h th IMF component, the dominant energy of the magnetic anomaly signal is distributed among these IMF components. The criterion based on continuous mean square error (CMSE) [34] is proposed to find the dividing point. The CMSE is defined as

$$\text{CMSE} = \frac{1}{N} \sum_{j=1}^N (c_k(j) - c_{k+1}(j))^2, \quad k = 1, \dots, K-1, \quad (2)$$

where N is the length of each IMF components and K is the number of IMF components.

If the first significant change of CMSE is $(h-1)$ th IMF component, the dividing point is h . It means that the energy components of adjacent IMF components are significantly different. Thus, the dividing point can be used to split the raw signal. In the magnetic data, the components before c_h are considered the main energy of the noise. The components after c_h contain the main energy of the magnetic signal.

2.2. Minimum Entropy Feature of Magnetic Anomaly. A magnetic target can be considered a magnetic dipole when the distance between the target and the magnetometer is more than three times the target largest dimension. For detecting

a magnetic target, the MAD model is shown in Figure 1. The magnetic field generated by the target can be expressed as [37]

$$\mathbf{B}_a = \frac{\mu_0}{4\pi} \left(\frac{3(\mathbf{M} \cdot \mathbf{R})\mathbf{R}}{R^5} - \frac{\mathbf{M}}{R^3} \right), \quad (3)$$

where $\mu_0 = 4\pi \times 10^{-7}$ H/m is the permeability of free space \mathbf{M} is the magnetic moment of the target, and \mathbf{R} is the distance between the target and the magnetic sensor.

In practice, the magnetic field \mathbf{B} measured by the magnetic sensor is consisted of the ambient magnetic field \mathbf{B}_e and magnetic field \mathbf{B}_a generated by the target when the target is present. When the target is far from the sensor, $|\mathbf{B}_a|$ is much less than $|\mathbf{B}_e|$. When using the total field magnetometers to detect the target, we can obtain the scalar value of the measurement as [13, 16]

$$B_m \approx B_e \left(1 + \frac{\mathbf{B}_e \cdot \mathbf{B}_a}{B_e^2} \right) = B_e + \mathbf{u} \cdot \mathbf{B}_a, \quad (4)$$

where \mathbf{u} denotes the unit vector of the ambient field.

In the measurement, the magnetic anomaly generated by the target can change the magnetic noise pattern. In the information theory, entropy is used as a measurement of information. In Ref. [12], entropy is used to detect the changes in the magnetic noise pattern. It means that the entropy feature of the magnetic field will dramatically change when the ambient magnetic field contains the magnetic anomaly. Thus, entropy value can be considered a feature of the magnetic anomaly. The value of entropy can be calculated in a moving window of L samples. The expression is given as follows:

$$F_{en}(x_i) = - \sum_{n=i-L+1}^i p(x_n) \log p(x_n), \quad (5)$$



FIGURE 2: One-dimensional sensor array with two magnetometers.

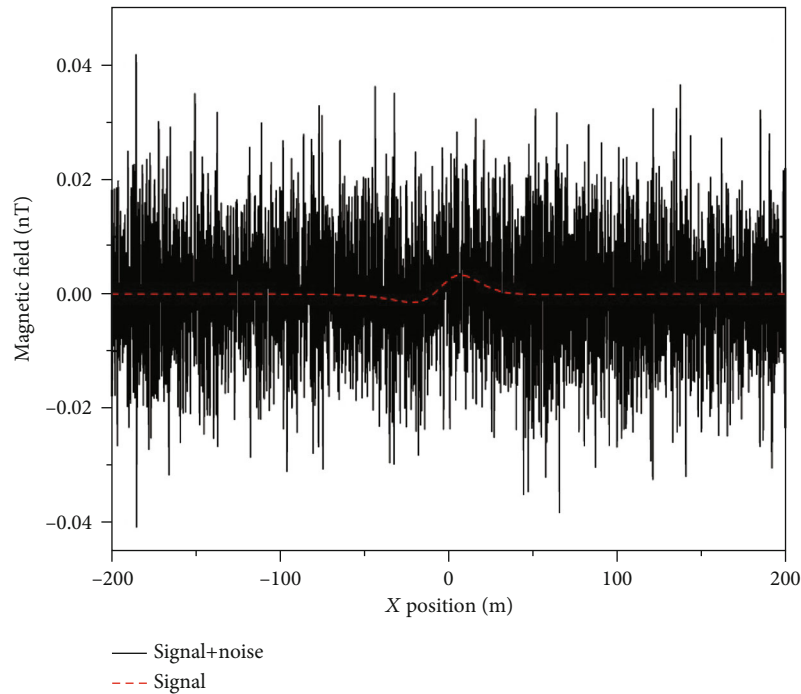


FIGURE 3: Synthetic signal of magnetic field.

where $p(x_n)$ denotes the probability density function of ambient magnetic field. The detection occurs whenever the entropy value drops below a threshold value.

In the ME method, a priori condition of PDF of magnetic noise is needed when using the entropy feature for the target detection. The PDF is acquired by using statistical methods based on the measured magnetic noise. However, when the environmental conditions change, it is necessary to update the PDF by the new data [17]. It is hard to acquire the new background noise in real time when the magnetometers mounted on the platform are moving. Thus, the detection performance of ME method will be affected in this situation.

2.3. Magnetic Anomaly Detection with EMD and Minimum Entropy Feature. In order to overcome the problems in the ME method in Ref. [12], we propose an adaptive MAD method with ensemble empirical mode decomposition and minimum entropy feature in this paper. Based on EEMD, the SNR of the reconstructed signal can be improved according to the dividing point. Meanwhile, the reconstructed noise can be approximated as magnetic background noise which is considered the Gaussian distribution [12]. The reconstructed noise can be used to update the parameters of the PDF of the noise in real time. Thus, the proposed method not only overcomes the problem of updating the PDF of the ambient magnetic noise in the ME method but also improves the detection

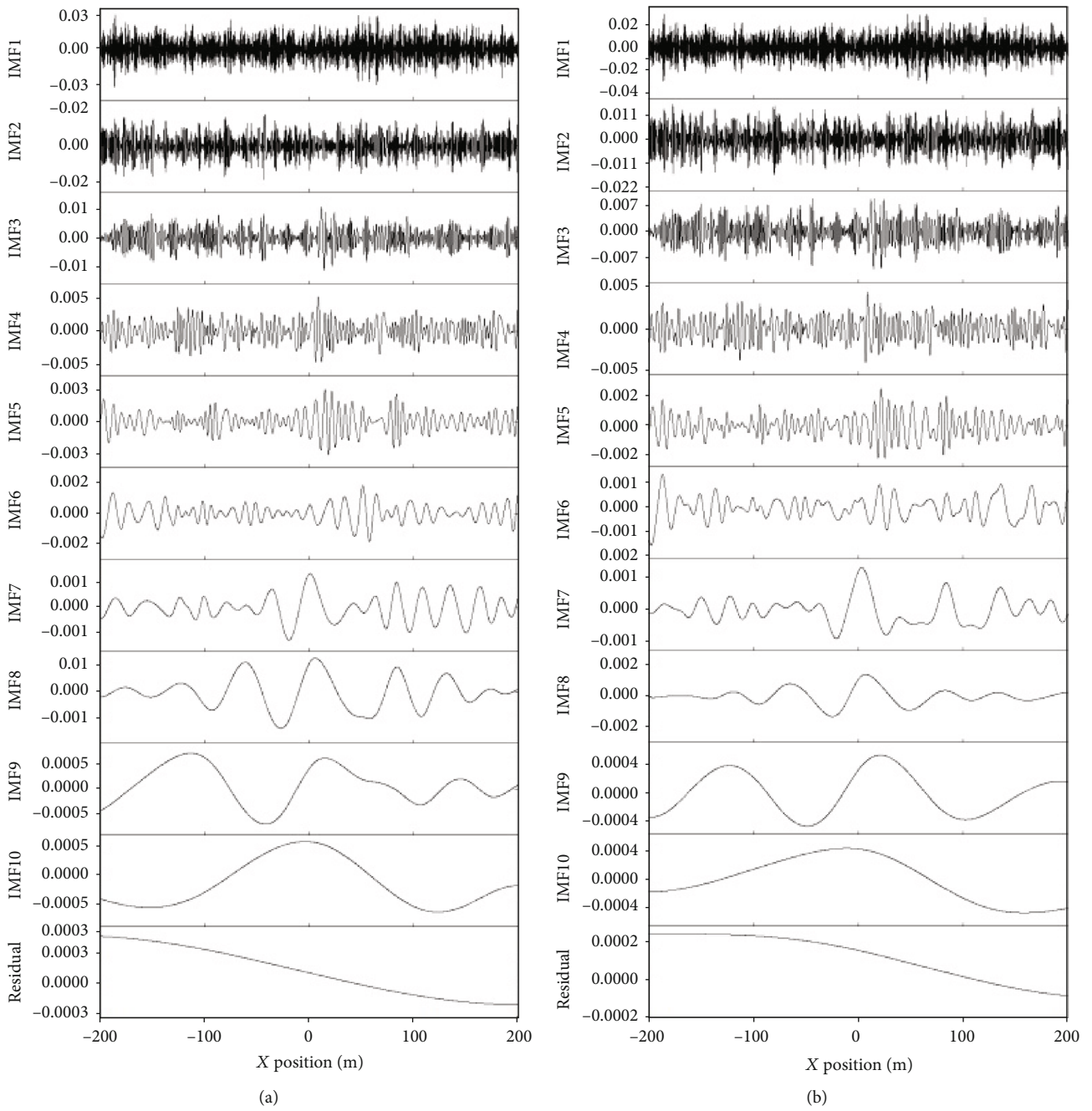


FIGURE 4: Continued.

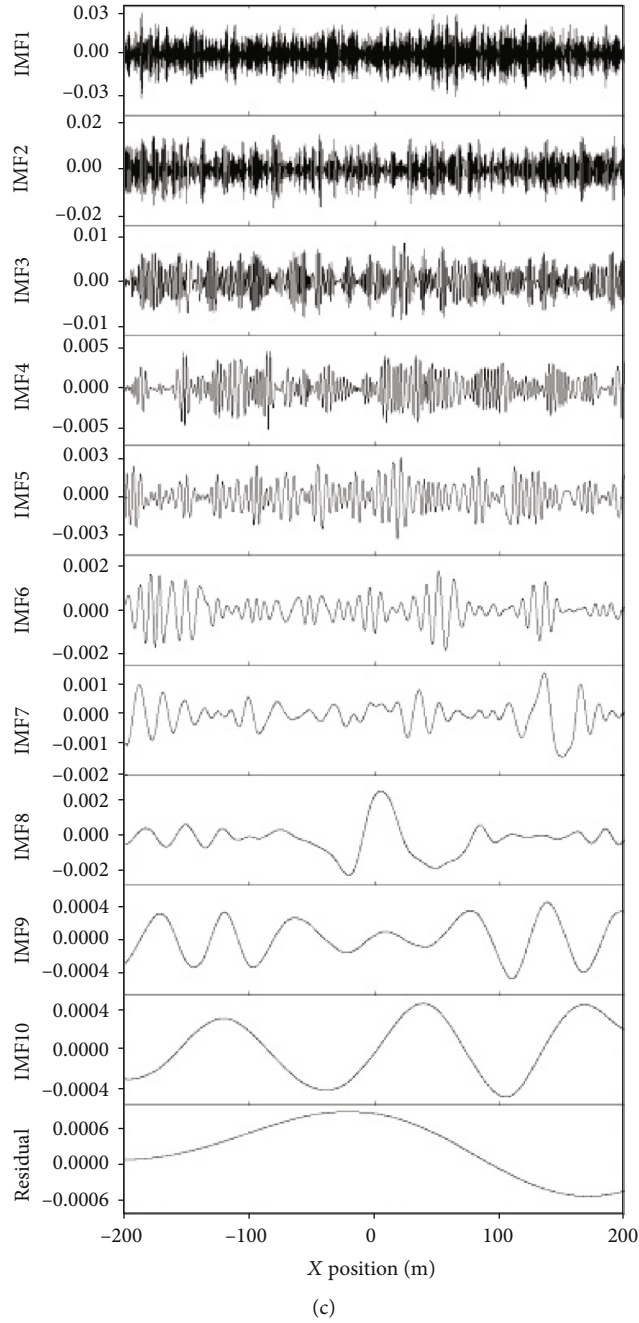


FIGURE 4: IMFs obtained from synthetic magnetic signal. (a) IMF decomposed by EMD. (b) IMF decomposed by EEMD. (c) IMF decomposed by CEEMDAN.

probability of the weak magnetic anomaly. Similarly, the EMD-ME and CEEMDAN-ME methods are also implemented. The steps of the proposed method are as follows:

Step 1. Decompose the original signal into a series of IMFs components and a residual component

Step 2. Reconstruct the signal and noise according to the dividing point

Step 3. Update the parameters of PDF of noise based on the reconstructed noise

Step 4. Calculate the entropy feature of the reconstructed signal with the PDF.

Step 5. According to the threshold, determine whether the detection occurs.

3. Experiment

3.1. Experiment Design. In order to reduce the influence of temporal magnetic variations on magnetic anomaly detection, we designed a one-dimensional sensor array (shown in Figure 2) to collect the magnetic background noise. The magnetic sensor was the optically pumped magnetometer with high sensitivity, of which the intrinsic noise was about

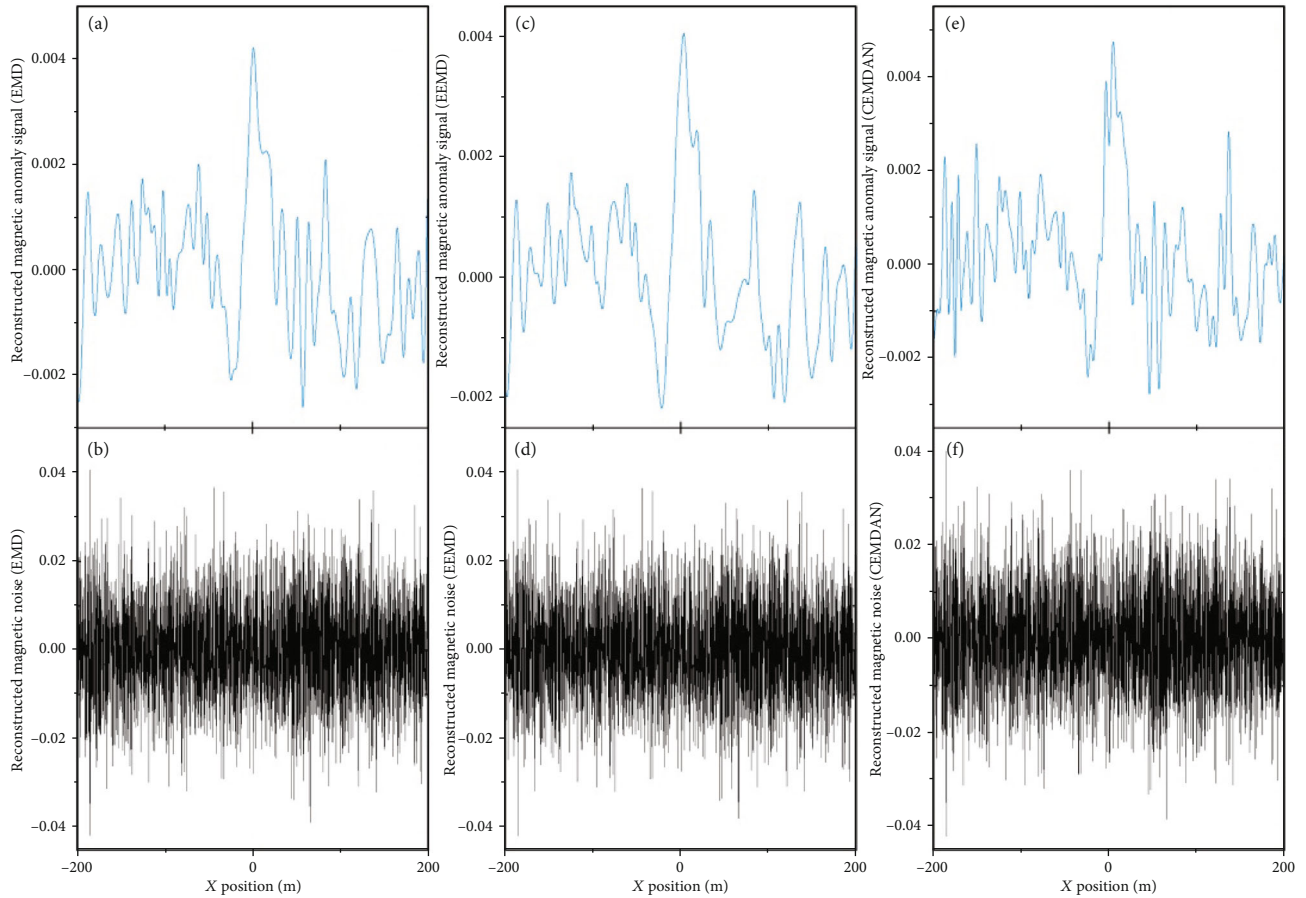


FIGURE 5: Reconstruction based on the IMFs. (a) Reconstructed signal using IMFs of EMD. (b) Reconstructed noise using IMFs of EMD. (c) Reconstructed signal using IMFs of EEMD. (d) Reconstructed noise using IMFs of EEMD. (e) Reconstructed signal using IMFs of CEEMDAN. (f) Reconstructed noise using IMFs of CEEMDAN.

$0.6\text{pT}/\sqrt{\text{Hz}}@1\text{Hz}$. We acquired the real magnetic noise samples in the Jinshatan Wetland Park in Harbin city, China, where the ambient magnetic activity was very low. The local inclination and declination of the geomagnetic field were 63.3° and -10.1° , respectively. The amplitude of the geomagnetic field was about $55,200\text{ nT}$.

The typical magnetic anomaly signal was generated by a simulated target. The target was moving in parallel to the x -axis with a constant velocity of 2 m/s , starting from $(-200, -35, -3)\text{ m}$ and ending at $(200, -35, -3)\text{ m}$. The magnetic moment of the target was unchanged during the movement. The magnitude of the moment was 30 A m^2 , and the unit direction vector was $(-0.65, -0.71, -0.26)$. The typical magnetic anomaly can be obtained by the simulation. Then, the real magnetic noise measured by sensor array was added to the simulated signal. Finally, the synthetic magnetic signal with real-world noise and the simulated anomaly signal can be obtained.

3.2. Results. The synthetic signal of magnetic field was obtained, shown in Figure 3. It can be seen that the magnetic anomaly signal was completely buried by the noise. The SNR of the synthetic signal was about -10.47 dB . The EEMD-ME method was applied to the synthetic signal, in which the

parameters of EEMD were set as follows: the standard deviation of added noise amplitude $\varepsilon = 0.1$, and the ensemble member $I = 100$. By the EEMD, the synthetic signal was decomposed into a set of IMFs, shown in Figure 4(b). Similarly, the EMD-ME and CEEMDAN-ME methods were also applied to the synthetic signal. In the CEEMDAN-ME method, the parameters were the same as that of the EEMD-ME method. The decomposition results by EMD and CEEMDAN were shown in Figures 4(a) and 4(c). From Figure 4, we can see that different IMFs reveal the different time scales of the magnetic field signal. According to formula (2), the dividing point was calculated as $h = 5$. Thus, IMF1-IMF5 represented the high-frequency components, in which the main energy of the noise was contained. IMF6-Residual represented the low-frequency components, in which the main energy of the signal was contained. Based on the IMFs, the reconstructed magnetic anomaly signal and the reconstructed magnetic noise were obtained, shown in Figure 5.

After the reconstruction, an approximated PDF of magnetic noise was obtained by the reconstructed magnetic noise. The parameters of PDF of the noise were updated based on the reconstructed noise. The entropy feature of the reconstructed signal can be calculated with the PDF of the noise. The outputs of the methods are shown in

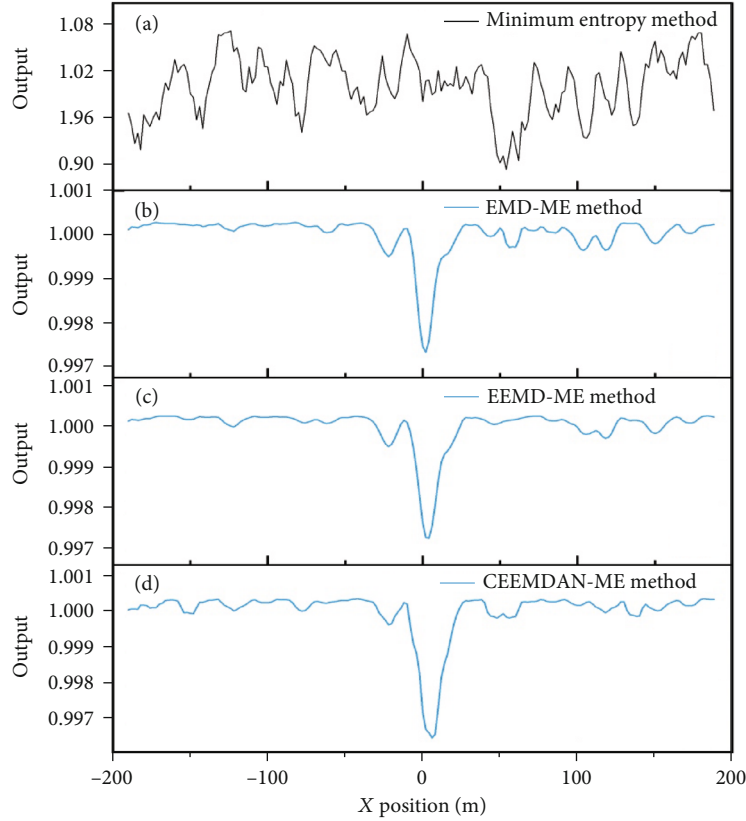


FIGURE 6: Output of the method. (a) Output of minimum entropy method. (b) Output of EMD-minimum entropy method. (c) Output of EEMD-minimum entropy method. (d) Output of CEEMDAN-minimum entropy method.

TABLE 1: Detection performance of the proposed methods.

Method	CPA							
	$P_y = 20 \text{ m}, P_z = 3 \text{ m}$		$P_y = 25 \text{ m}, P_z = 3 \text{ m}$		$P_y = 30 \text{ m}, P_z = 3 \text{ m}$		$P_y = 35 \text{ m}, P_z = 3 \text{ m}$	
	ACC (%)	time (s)	ACC (%)	time (s)	ACC (%)	time (s)	ACC (%)	time (s)
EMD-ME	100	0.17	100	0.13	91.0	0.12	57.4	0.11
EEMD-ME	100	1.44	100	1.12	99.5	0.98	90.0	0.90
CEEMDAN-ME	100	3.56	100	3.01	97.7	2.78	79.6	2.64

Figures 6(b)–6(d). It can be seen that the magnetic anomaly was obviously revealed by the proposed methods. In addition, the traditional ME method was also applied to the synthetic signal. The output of the ME method is shown in Figure 6(a). From the results, the detection performance of the proposed method was better than that of the ME method in low input SNR.

Compared to the traditional ME method, there are two advantages of the proposed method in weak magnetic anomaly detection. (1) The approximated PDF of the noise can be updated in real time by the reconstructed magnetic noise. It overcomes the problem that the PDF of magnetic noise is difficult to update in the changing movement. (2) The input SNR can be improved by the reconstruction signal based on IMFs. Compared to the SNR of the original signal with -10.47 dB , the SNR of the reconstructed signal is

improved to 9.37 dB in the EEMD-ME method, 8.16 dB in the EMD-ME method, and 7.14 dB in the CEEMDAN-ME method, respectively. In addition, the execution time of updating the parameters of PDF of the noise is about 0.87 s in the EEMD-ME method, 0.07 s in the EMD-ME method and 2.51 s in the CEEMDAN-ME method, respectively. It can be considered the update of PDF in real time, especially in the EMD-ME and EEMD-ME methods.

The mentioned SNR is calculated by the following equation:

$$\text{SNR} = 20 \log_{10} \frac{\max(x(1:N))}{\text{std}(n(1:N))}, \quad (6)$$

where x denotes the signal, n denotes the noise, and N is the length of the signal.

In order to evaluate the effectiveness of the proposed method, the Monte Carlo simulation was carried out. In the simulation, the simulated target was moving in parallel to the x -axis with a constant velocity of 2/m, the range of X position in (-200, 200)m. The magnitude of magnetic moment of the target was 30 A m^2 . The orientation (inclination and declination of the moment) of the target was randomly varied. According to the Neyman–Pearson criterion [21], the threshold values of EEMD-ME, EMD-ME, and CEEMDAN-ME were set to 0.9991, 0.9955, and 0.9984 with the false alarm rate 3.4% based on the real magnetic noise. The closest proximity approach (CPA) from the magnetic sensors to the moving track of the target was changed to evaluate the performances of the methods. Table 1 shows the detection performance versus with different CPA.

From Table 1, the magnetic anomaly generated by the target can be detected by the EMD-ME, EEMD-ME, and CEEMDAN-ME methods, and the detection accuracy was 100% when CPA was less than 25.18 ($\sqrt{P_y^2 + P_z^2}$). As the CPA increased, the detection accuracy decreased. When CPA was 30.14, the detection accuracy of EEMD-ME was 99.5%, which was higher than that of the CEEMDAN-ME and EMD-ME methods. Similarly, when the CPA was increased to 35.12, the detection accuracy of EEMD-ME dropped to 90.5%. Under the same condition, the detection accuracy of CEEMDAN-ME and EMD-ME methods dropped to 79.6% and 57.4%, respectively. From the result, it is noted that the detection performance of EEMD-ME is better than of EMD-ME and CEEMDAN-ME methods in weak magnetic anomaly detection. Meanwhile, the execution time of the EEMD-ME method is about 1 s. Thus, the EEMD-ME method is better in the weak magnetic anomaly detection.

4. Conclusions

In this paper, we propose the adaptive method with ensemble empirical mode decomposition and minimum entropy feature, which can improve the performance of magnetic anomaly detection with low SNR. In the method, the raw magnetic data is processed by EEMD, and IMFs are obtained. According to the properties of IMFs, the reconstructed magnetic noise is used to update the parameters of PDF of it. The reconstructed magnetic signal is used to extract the entropy feature of the magnetic anomaly using the updated PDF. Whenever the entropy feature drops below the threshold, the magnetic anomaly is detected. Compared to the traditional ME method, the entropy feature extracted by the proposed method is more obvious and useful to reveal the weak magnetic anomaly. The execution time of EEMD-ME method is about 1 s. Meanwhile, the EMD-ME and CEEMDAN-ME methods are implemented and used to detect the weak magnetic anomaly signal. The results show that when the CPA is increased from 30.14 to 35.12, the detection accuracy of EEMD-ME method drops from 99.5% to 90.5%. However, the detection accuracy of EMD-ME and CEEMDAN-ME methods drops from 91.0% and 97.7% to 57.4% and 79.6%,

respectively. Therefore, the detection performance of EEMD-ME method is better in the weak magnetic anomaly detection.

Data Availability

The magnetic field data used to support the findings of this study were supplied by Geomagnetic Sensing and Application Lab, Harbin Engineering University, under license and so cannot be made freely available. Requests for access to these data should be made to Prof. Chong Kang, kangchongheu@163.com.

Conflicts of Interest

The authors declare that they have no conflict of interests.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 51909216), Natural Science Basic Research Program of Shaanxi (Program No. 2020JQ-151), and Fundamental Research Funds for the Central Universities (No. 31020190QD033).


References

- [1] S. L. Tantum, Y. Yu, and L. M. Collins, “Bayesian mitigation of sensor position errors to improve unexploded ordnance detection,” *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 1, pp. 103–107, 2008.
- [2] Z. Zalevsky, Y. Bregman, N. Salomonski, and H. Zafrir, “Resolution enhanced magnetic sensing system for wide coverage real time uxu detection,” *Journal of Applied Geophysics*, vol. 84, pp. 70–76, 2012.
- [3] Y. Sui, K. Leslie, and D. Clark, “Multiple-order magnetic gradient tensors for localization of a magnetic dipole,” *IEEE Magnetics Letters*, vol. 8, pp. 1–5, 2017.
- [4] Z. Guo, D. Liu, Q. Pan, Y. Zhang, Y. Li, and Z. Wang, “Vertical magnetic field and its analytic signal applicability in oil field underground pipeline detection,” *Journal of Geophysics and Engineering*, vol. 12, no. 3, pp. 340–350, 2015.
- [5] C. Xu, J. Chen, D. Yan, and J. Ji, “Review of underwater cable shape detection,” *Journal of Atmospheric and Oceanic Technology*, vol. 33, no. 3, pp. 597–606, 2016.
- [6] J. A. Baldoni and B. B. Yellen, “Magnetic tracking system: monitoring heart valve prostheses,” *IEEE Transactions on Magnetics*, vol. 43, no. 6, pp. 2430–2432, 2007.
- [7] J. E. McGary, “Real-time tumor tracking for four-dimensional computed tomography using squid magnetometers,” *IEEE Transactions on Magnetics*, vol. 45, no. 9, pp. 3351–3361, 2009.
- [8] D. Liu, X. Xu, C. Huang et al., “Adaptive cancellation of geomagnetic background noise for magnetic anomaly detection using coherence,” *Measurement Science and Technology*, vol. 26, no. 1, article 015008, 2015.
- [9] L. Fan, C. Kang, X. Zhang, and S. Wan, “Real-time tracking method for a magnetic target using total geomagnetic field intensity,” *Pure and Applied Geophysics*, vol. 173, no. 6, pp. 2065–2071, 2016.
- [10] J. Ge, S. Wang, H. Dong et al., “Real-time detection of moving magnetic target using distributed scalar sensor based on hybrid algorithm of particle swarm optimization and gauss-

- newton method," *IEEE Sensors Journal*, vol. 20, no. 18, pp. 10717–10723, 2020.
- [11] C. Wan, M. Pan, Q. Zhang, D. Chen, H. Pang, and X. Zhu, "Performance improvement of magnetic anomaly detector using karhunen-loeve expansion," *IET Science, Measurement and Technology*, vol. 11, no. 5, pp. 600–606, 2017.
- [12] A. Sheinker, N. Salomonski, B. Ginzburg, L. Frumkis, and B.-Z. Kaplan, "Magnetic anomaly detection using entropy filter," *Measurement science and technology*, vol. 19, no. 4, article 045205, 2008.
- [13] A. Sheinker, A. Shkalim, N. Salomonski, B. Ginzburg, L. Frumkis, and B.-Z. Kaplan, "Processing of a scalar magnetometer signal contaminated by $1/f\alpha$ noise," *Sensors and Actuators A: Physical*, vol. 138, no. 1, pp. 105–111, 2007.
- [14] A. Sheinker, B. Ginzburg, N. Salomonski, P. A. Dickstein, L. Frumkis, and B.-Z. Kaplan, "Magnetic anomaly detection using high-order crossing method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1095–1103, 2012.
- [15] C. Wan, M. Pan, Q. Zhang, F. Wu, L. Pan, and X. Sun, "Magnetic anomaly detection based on stochastic resonance," *Sensors and Actuators A: Physical*, vol. 278, pp. 11–17, 2018.
- [16] L. Fan, X. Kang, Q. Zheng et al., "A fast linear algorithm for magnetic dipole localization using total magnetic field gradient," *IEEE Sensors Journal*, vol. 18, no. 3, pp. 1–1038, 2017.
- [17] Y. Tang, Z. Liu, M. Pan et al., "Detection of magnetic anomaly signal based on information entropy of differential signal," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 4, pp. 512–516, 2018.
- [18] L. Yang, L. Zhongyan, P. Mengchun et al., "Magnetic anomaly signal space analysis and its application in noise suppression," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 1, pp. 130–134, 2019.
- [19] Y. Shen, J. Wang, J. Shi, S. Zhao, and J. Gao, "Interpretation of signature waveform characteristics for magnetic anomaly detection using tunneling magnetoresistive sensor," *Journal of Magnetism and Magnetic Materials*, vol. 484, pp. 164–171, 2019.
- [20] H. Zhou, Z. Pan, and Z. Zhang, "Magnetic anomaly detection with empirical mode decomposition trend filtering," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E100.A, no. 11, pp. 2503–2506, 2017.
- [21] C. Wan, M. Pan, Q. Zhang, D. Chen, H. Pang, and X. Zhu, "Performance improvement of magnetic anomaly detector using Karhunen-Loeve expansion," *IET Science, Measurement & Technology*, vol. 11, no. 5, pp. 600–606, 2017.
- [22] B. Ginzburg, L. Frumkis, and B. Z. Kaplan, "Processing of magnetic scalar gradiometer signals using orthonormalized functions," *Sensors and Actuators A: Physical*, vol. 102, no. 1-2, pp. 67–75, 2002.
- [23] W. Zhao, X. Zhang, and H. Wang, "Characteristics and noise combination suppression of wide-angle ocean bottom seismography (obs) data in shallow water: a case study of profile obs 2016 in the South Yellow Sea," *Chinese Journal of Geophysics*, vol. 63, no. 6, pp. 2415–2433, 2020.
- [24] L. Fan, C. Kang, H. Hu et al., "Gradient signals analysis of scalar magnetic anomaly using orthonormal basis functions," *Measurement Science and Technology*, 2020.
- [25] A. Sheinker, A. Shkalim, N. Salomonski, B. Ginzburg, L. Frumkis, and B.-Z. Kaplan, "Processing of a scalar magnetometer signal contaminated by $1/f\alpha$ noise," *Sensors and Actuators A: Physical*, vol. 138, no. 1, pp. 105–111, 2007.
- [26] S. Liu, J. Hu, P. Li et al., "Magnetic anomaly detection based on full connected neural network," *IEEE Access*, vol. 7, pp. 182198–182206, 2019.
- [27] H. Zhao, J. Zheng, W. Deng, and Y. Song, "Semi-supervised broad learning system based on manifold regularization and broad network," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 3, pp. 983–994, 2020.
- [28] W. Deng, H. Liu, J. Xu, H. Zhao, and Y. Song, "An improved quantum-inspired differential evolution algorithm for deep belief network," *IEEE Transactions on Instrumentation and Measurement*, p. 1, 2020.
- [29] H. Zhao, J. Zheng, J. Xu, and W. Deng, "Fault diagnosis method based on principal component analysis and broad learning system," *IEEE Access*, vol. 7, pp. 99263–99272, 2019.
- [30] W. Deng, J. Xu, and H. Zhao, "An improved ant colony optimization algorithm based on hybrid strategies for scheduling problem," *IEEE Access*, vol. 7, pp. 20281–20292, 2019.
- [31] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [32] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2011.
- [33] Z. Ma, G. Wen, and C. Jiang, "EEMD independent extraction for mixing features of rotating machinery reconstructed in phase space," *Sensors*, vol. 15, no. 4, pp. 8550–8569, 2015.
- [34] S. Nalband, A. Prince, and A. Agrawal, "Entropy-based feature extraction and classification of vibroarthrographic signal using complete ensemble empirical mode decomposition with adaptive noise," *IET Science, Measurement and Technology*, vol. 12, no. 3, pp. 350–359, 2018.
- [35] Y. Li, L. Wang, X. Li, and X. Yang, "A novel linear spectrum frequency feature extraction technique for warship radio noise based on complete ensemble empirical mode decomposition with adaptive noise, duffing chaotic oscillator, and weighted-permutation entropy," *Entropy*, vol. 21, no. 5, p. 507, 2019.
- [36] Y. Li, X. Chen, and J. Yu, "A hybrid energy feature extraction approach for ship-radiated noise based on ceemdan combined with energy difference and energy entropy," *Processes*, vol. 7, no. 2, p. 69, 2019.
- [37] T. Nara, S. Suzuki, and S. Ando, "A closed-form formula for magnetic dipole localization by measurement of its magnetic field and spatial gradients," *IEEE Transactions on Magnetics*, vol. 42, no. 10, pp. 3291–3293, 2006.

Research Article

Measurement of Cotton Canopy Temperature Using Radiometric Thermal Sensor Mounted on the Unmanned Aerial Vehicle (UAV)

Anjin Chang,¹ Jinha Jung,² Murilo M. Maeda,³ Juan A. Landivar,⁴ Henrique D. R. Carvalho,⁵ and Junho Yeom ⁶

¹School of Engineering and Computing Science, Texas A&M University-Corpus Christi, Corpus Christi, TX 78414, USA

²Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA

³Texas A&M AgriLife Extension, Lubbock, TX 79403, USA

⁴Texas A&M AgriLife Research, Corpus Christi, TX 78406, USA

⁵Department of Soil and Crop Science, Texas A&M University, College Station, TX 77843, USA

⁶Department of Civil Engineering, Gyeongsang National University, Jinju, South Gyeongsang Province 52828, Republic of Korea

Correspondence should be addressed to Junho Yeom; junho.yeom@gnu.ac.kr

Received 30 June 2020; Revised 6 August 2020; Accepted 13 August 2020; Published 19 August 2020

Academic Editor: Sang-Hoon Hong

Copyright © 2020 Anjin Chang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Canopy temperature is an important variable directly linked to a plant's water status. Recent advances in Unmanned Aerial Vehicle (UAV) and sensor technology provides a great opportunity to obtain high-quality imagery for crop monitoring and high-throughput phenotyping (HTP) applications. In this study, a UAV-based thermal system was developed to directly measure canopy temperature, skipping the traditional radiometric calibration process which is time-consuming and complicates data processing. Raw thermal imagery collected over a cotton field was converted to surface temperature using the Software Development Kit (SDK) provided by the sensor company. Canopy temperature map was generated using Structure from Motion (SfM), and Thermal Stress Index (TSI) was calculated for the test site. UAV temperature measurements were compared to ground measurements acquired by net radiometers and thermocouples. Temperature differences between UAV and ground measurements were less than 5%, and UAV measurements proved to be more stable. The proposed UAV system was successful in showing temperature differences between the cotton genotype. In conclusion, the system described in this study could possibly be used to monitor crop water status in a field setting, which should prove helpful for precision agriculture and crop research.

1. Introduction

Canopy temperature is an important indicator of water availability, water stress, and irrigation status in agriculture [1]. Measuring crop canopy temperature can help establishing relationships with harvest yield, as well as support water management decisions [2]. Remotely sensed data, which are acquired by sensors on space-borne, air-borne, or ground-based platforms, have been widely used in agriculture to estimate crop parameters such as vegetation indices and Leaf Area Index (LAI) [3, 4]. However, limitations of traditional remote sensing technologies include low spatial and temporal resolutions for time-series analysis, as well as high cost and low efficiency [5, 6].

Unmanned Aerial Vehicle (UAV) and sensor technology are quickly evolving and offering a great opportunity for the development of precision agriculture and high-throughput phenotyping (HTP) systems for a variety of applications. Most studies using UAV for agriculture have focused on red-green-blue (RGB) and/or multispectral sensors to calculate vegetation indices and monitor crop development for yield forecasting. Advanced UAV systems can provide fine spatial and high temporal resolution data at relatively low cost so that crop traits such as height, canopy morphology, and greenness can be estimated [2, 6–8]. Vegetation indices were calculated by remotely sensed data from a multispectral sensor mounted on a UAV to estimate LAI, which is one of the key parameters determining photosynthesis, respiration,

and transpiration of vegetation [2, 9]. Chang et al. [6] proposed a method to monitor the growth of sorghum using a commercially-available UAV, while Anthony et al. [7] used a Micro-UAV system with a laser scanner to measure crop height. Furthermore, Patrick and Li [8] generated 3D models of blueberry bushes from UAV data to extract morphological traits for genotype selection and found a strong relationship between traditional growth indices and image-derived bush volume. The UAV platforms with hyperspectral sensors have been used to extract phenotypes and to predict biomass for sorghum [10–12]. Ramanurthy et al. [10] extracted features from hyperspectral and RGB images for predictive modeling of sorghum plants. Zhang et al. [11] developed nonlinear regression models to predict sorghum biomass from multi-temporal UAV-based hyperspectral and RGB data, while Ali et al. [12] adopted hyperspectral, LiDAR, and RGB data for sorghum biomass prediction.

Although canopy temperature may be used to detect crop stress [13], there are challenges associated with the ability to accurately measure canopy temperature using thermal cameras, whether mounted on a ground-based platform [13], manned and unmanned aerial platforms [14–17] or satellite [18]. Data from each platform could be useful for different purposes, however, as long as their different spatial resolution output is accounted for. Postharvest and quality evaluation operations of fruits and vegetables have been conducted using sensors and ground-based platforms [19–21]. Bulanon et al. [22] studied fruit recognition using thermal imaging to enhance the robotic harvesting of citrus. Berni et al. [14], Zarco-Tejada et al. [15], and Gonzalez-Dugo et al. [16] used a UAV platform to collect multispectral, hyperspectral, and thermal data to calculate vegetation indices, water stress, and canopy temperature of fruit tree species. Remotely acquired thermal images from aerial platform have also been used to derive the Crop Water Stress Index (CWSI) and map canopy conductance in olive orchards [17]. Thermal imagery collected from fixed-wing platforms at altitudes of 150 m or higher to cover large agricultural areas complicates image preprocessing methods including radiometric calibration and atmospheric correction [14, 15]. Ribeiro-Gomes et al. [23] proposed an uncooled calibration algorithm for thermal camera used in UAV applications for agriculture, while Berni et al. [14] performed laboratory calibration using a blackbody source to estimate stabilization procedure and absolute temperature shifts for the radiometric sensor calibration. Atmospheric correction methods have been applied to thermal imagery based on the MODTRAN radiative transfer model to calculate surface temperatures [15, 16]. Single-layer atmosphere (uniform conditions including air temperature, relative humidity, and barometric pressure) corrections of thermal imagery for fixed-wing UAV have been proposed [14]. Additionally, thermal sensors can be affected by camera tilt angles caused by the platform's orientation, viewing angle, and directional effects [17]. Rotary-wing UAV systems, on the other hand, could be a good alternative to overcome limitations of fixed-wing platforms. Although rotary-wing UAVs cover smaller areas at lower altitudes (<100 m) usually, these platforms tend to be more stable and provide reliable data with higher spatial resolution (<10 cm) [6]. In

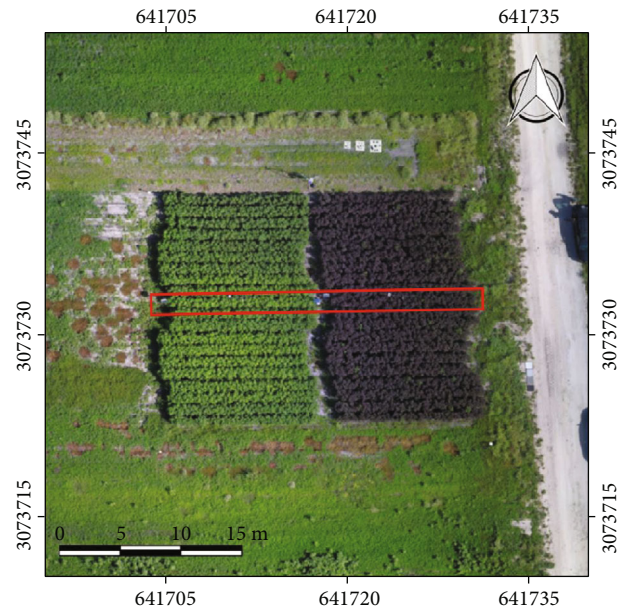


FIGURE 1: Image of study area taken by a Phantom 4 Pro on September 8th, 2016. The coordinate system is WGS84 UTM 14N. The red rectangle indicates the placement of ground sensors (net radiometer and thermocouples), in the 11th row from North. Meteorological data were collected at the study site by a weather station located in the middle of the field between the two cotton genotypes.

recent days, a multirotor UAV with a thermal camera used to estimate the adaptive CWSI for precision agriculture [24]. UAV-based thermal system provided relevant instantaneous and seasonal variations of water status [25].

In this study, we developed a UAV-based thermal system using a quad-copter platform and a radiometric thermal sensor. The framework of data collection and processing is proposed to directly measure crop canopy temperatures without the traditional radiometric calibration and/or atmospheric correction as well as field-based measurement. Thermal images can be collected over a cotton field to generate a canopy temperature and the crop stress index map. Canopy temperature measured by the UAV-mounted sensor was compared with ground-based measurements to evaluate the performance of the proposed system.

2. Materials and Methods

2.1. Study Area and Data Acquisition. The test site was located at the Texas A&M AgriLife Research and Extension Center in Corpus Christi, Texas, USA (-97°33'E, 27°47'N) (see Figure 1). Two genotypes of cotton (*Gossypium hirsutum* L.) contrasting in leaf pigmentation were used. The line TAMU4920 (Red leaves) has a marked presence of anthocyanins, whereas DP1044 (Green leaves) do not. Plot size for each line was limited to 20 rows that were 12.2 m long, spaced 0.96 m apart, and followed East-West orientation. The materials were planted on June 20, 2016. Irrigation tapes were installed on July 3, 2016, and plots were irrigated as needed throughout the season, to promote adequate vegetative

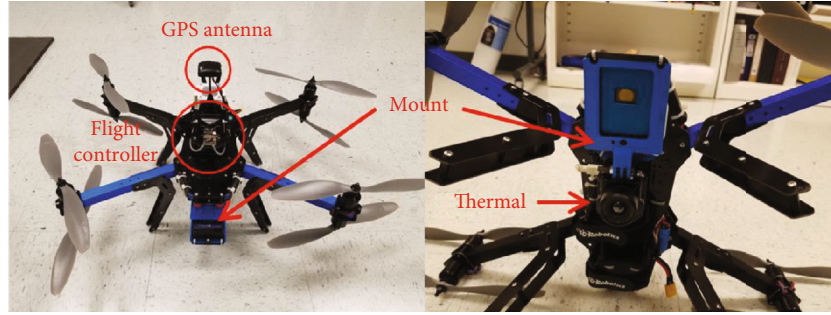


FIGURE 2: The integrated UAV platform, radiometric thermal sensor, and other components such as GPS and 3D printed mount.

growth. The crop height between red and green cotton was similar at UAV data collection.

Ground-based leaf temperature measurements were collected in each plot using ten type-T thermocouples (OMEGA Engineering, Bridgeport, NJ), placed on row 11 (Figure 1). The average of ten measurements was used to estimate the leaf temperature of the plots. The sensors were installed at the centre of the abaxial surface of sunlit main-stem leaves using clear surgical tape with the sensor wires secured around the leaf petioles by zip ties. Two 4-channel net radiometers (model CNR1, Kipp & Zonen, Delft, Netherlands) were used to measure net radiation (R_n) over the plots. The instruments were installed at the centre of each field at a height of 2 m above the soil surface. The radiometer measured four components of the surface radiation balance separately: direct incoming shortwave radiation (SW_{in}), reflected shortwave radiation (SW_{out}), longwave radiation from sky (LW_{in}), and longwave radiation emitted by the surface (LW_{out}). LW_{out} was converted to surface temperature [26, 27] as

$$T_c = \left(\frac{LW_{out} - (1 - \varepsilon)LW_{in}}{\varepsilon\sigma} \right)^{0.25} - 273.15, \quad (1)$$

where ε is the emissivity of the crop, assumed to be 0.96 [28], and σ is the Stefan-Boltzmann constant ($5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$).

The sensors were controlled by dedicated data-loggers (model CR1000, Campbell Scientific, Logan, UT). The thermocouples were multiplexed. For TAMU4920, an AM25T (Campbell Scientific, Logan, UT) multiplexer was used, while for DP1044, an AM16/32B (Campbell Scientific, Logan, UT) multiplexer was used. The data-loggers were programmed to scan the sensors every 30 seconds and compute 10-minute averages.

The UAV system and thermal sensor used in this study were a 3DR X8 octocopter system, (3D Robotics, Berkeley, USA), and FLIR Vue Pro R 640 radiometric thermal camera (FLIR, Wilsonville, USA), respectively (Figure 2). A PixHawk flight controller (3DR, Berkeley, USA) was used for platform and sensor integration; the sensor was mounted to the platform using a custom 3D-printed mount. The thermal sensor was equipped with a 9 mm lens, producing images with 640×512 pixels and a spectral response in the range of $7.5\text{-}13.5 \mu\text{m}$. The detectable temperature range of the sensor

TABLE 1: The summary of flight plan and data collection.

Flight parameters	Value
Altitude	40 m above ground
Overlap	75%
Date and time	September 8, 2016 11:10~20 AM

is $-20\text{-}50^\circ\text{C}$. The raw images were saved as uncompressed 14-bit radiometric images with telemetry in standard metadata fields. The camera was factory-calibrated to calculate temperatures with 5% measurement accuracy. The sensor manufacturer, FLIR, provides a Matlab library for conversion of pixel values in the raw image output to surface temperatures using radiometric metadata information and parameters such as emissivity, atmosphere temperature, relative humidity, reflective temperature, and distance of the sensor to target. Raw images were converted to canopy temperature before applying the Structure from Motion (SfM) algorithm to generate the orthomosaic image and canopy temperature map.

The UAV flight parameters were determined based on the field size and sensor specification (see Table 1). We adopted a grid-style flight pattern and nadir view for data collection, as described by Chang et al. [6]. A total of 89 geo-tagged thermal images were collected for the study site.

2.2. Conversion of Thermal Imagery and Image Mosaicking. A Matlab Software Development Kit (SDK) is currently being offered by FLIR as an ‘‘Add on Tool’’ to support the use of FLIR thermal sensors. Although it is not open-source, the user can install it for free and use the tools to view, analyze, and capture data from FLIR thermal sensors directly in Matlab. Five parameters, including emissivity, atmosphere temperature, relative humidity, reflective temperature, and sensor distance to target, have to be entered in order to convert raw imagery to surface temperatures. Emissivity is defined as the ratio of infrared energy emitted by the object, compared to that emitted by an ideal blackbody. We used an emissivity value of 0.96 for cotton [29]. Reflective temperature is any thermal radiation originating from other objects that reflects off the target [30]. We used the temperature measured by a weather station in the field as a reflective temperature. Atmosphere temperature and relative humidity were selected in the National Centers for Environmental Information (NCEI) provided by NOAA [31]. Flight altitude

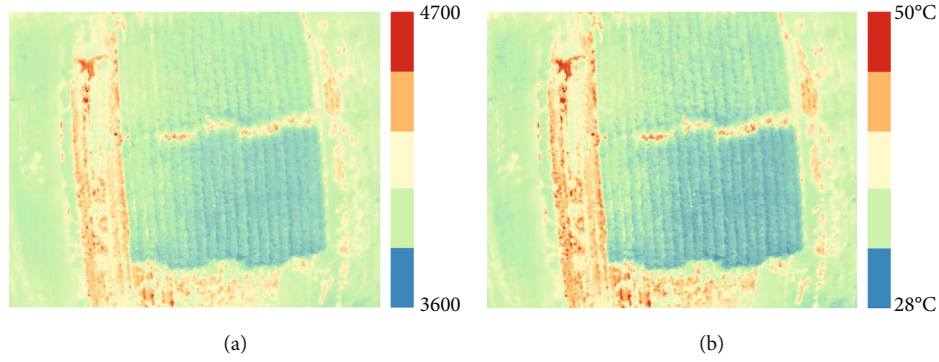


FIGURE 3: Examples of (a) raw image and (b) temperature-converted result.

was used as the sensor's distance to target. The digital number (DN) of each pixel in raw thermal images was input to the function of the SDK with the parameters to calculate surface temperatures. Figures 3(a) and 3(b) show individual raw (digital number) and converted surface temperature images, respectively.

After image conversion, surface temperature images were processed using the Agisoft Photoscan Pro software to generate an orthomosaic image of the study area. Photoscan Pro adopts the Structure from Motion (SfM) algorithm, which is a photogrammetric imaging technique to estimate 3D structures from 2D images [2, 6, 32, 33]. Generally, UAV images can be stitched using tie-points between the images for alignment and geo-referenced with GPS coordinate of image location determined by GPS module equipped in the UAV platform. In this study, the flight controller, PixHaw, recorded GPS coordinates (X, Y, Z) when a trigger signal was sent to the thermal sensors to capture images. Although the single GPS module was used, all images included a GPS coordinate in its metadata. Geo-tagged information was input into Photoscan Pro to generate a geo-referenced orthomosaic canopy temperature map. Image alignment to generate a sparse point cloud, the point optimization, dense point cloud generation, and DSM/Orthomosaic generation were conducted sequentially.

2.3. Thermal Stress Index (TSI). Canopy temperature has been considered as a proxy for monitoring crop water status [16]. Although CWSI can be calculated from thermal images, upper and lower baseline temperature of air and canopy should be measured over the whole growing season. However, Thermal Stress Index, which was optimized for cotton, was defined as Equation (2) as below.

$$TSI = \frac{\max(T_f, T_b) - T_b}{T_b}, \quad (2)$$

where T_f is foliage temperature and T_b is the biochemically determined baseline temperature. The crop-specific biochemical temperature optimum was suggested as baseline temperature. In the literature, The 27.5°C midpoint temperature of the Thermal Kinetic Window, which means the temperature range for which the value of the apparent Michaelis

constant remained within 200% of the minimum observed value, for cotton was examined as a baseline temperature (T_b) for the TSI [34]. The values of TSI range from zero to some positive limit. The biochemical-based TSI and the physically based CWSI were highly correlated for cotton across a range of environmental conditions. In this study, the temperature value in the orthomosaic temperature map was considered as the foliage temperature (T_f) to calculate TSI.

3. Results and Discussion

3.1. Canopy Temperature & TSI Map. A canopy temperature map was created from thermal images collected with the UAV-based system using the SfM algorithm and geo-tagged temperature images (see Figure 4(a)). The final canopy temperature map had a 7 cm spatial resolution. In this study, an orthomosaic image with finer spatial resolution could be generated since rotary-wing UAVs can fly at lower altitude with fairly stable orientation, when compared to a fixed-wing system. The temperature of bare soil and road was higher than those of vegetation. Temperature differences were also visible between green and red cotton, with the green genotype exhibiting lower canopy temperatures.

A TSI map was computed using Equation (2) and thermal data (see Figure 4(b)). As expected, similar to what was seen for the surface temperatures, TSI values for the green cotton genotype were lower and ranged from 0.2 to 0.3, while the red cotton genotype exhibited TSI values approximately 0.1 higher. In the TSI map, it was found crops in the northern area had more pressure of water. It could be supposed that water was not enough due to the irrigation system, terrain slope, or vegetation cover.

3.2. Evaluation of Canopy Temperature. The canopy temperature from UAV and ground-based sensors was compared (see Figure 5). The average of pixels on 11 rows in orthomosaic temperature map was adopted as UAV measurement. As indicated by ground and UAV measurements, the red plants were warmer than the green ones. The temperature of the red cotton measured by the net radiometer, thermocouples, and UAV was 33.69, 33.77, and 34.65°C, respectively. For the green cotton, the net radiometer, thermocouples, and UAV

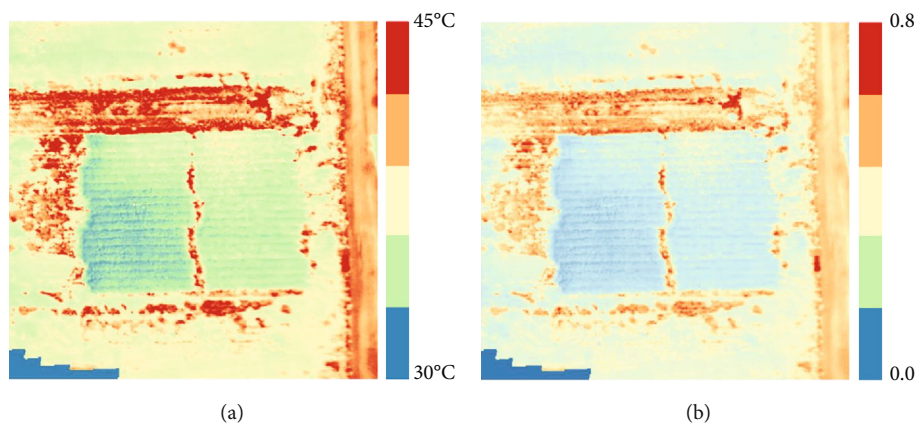


FIGURE 4: Orthomosaic image of (a) canopy temperature and (b) Thermal Stress Index (TSI) for the study area.

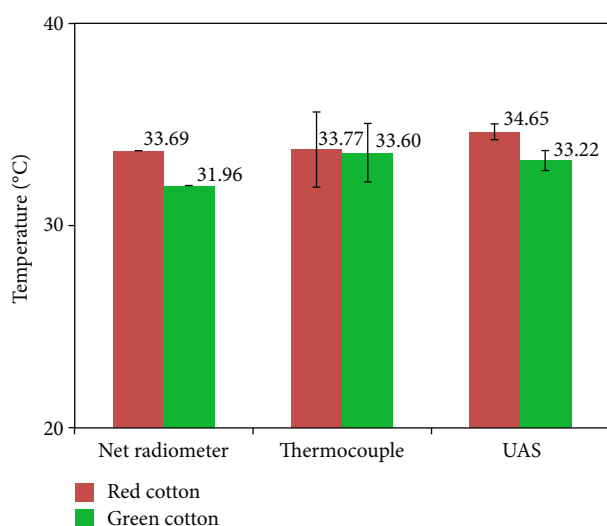


FIGURE 5: Canopy and leaf temperatures of green and red cotton genotypes. Bars indicate \pm one standard deviation.

measured 31.96, 33.60, and 33.22°C, respectively. For both plots, the UAV measurements slightly overestimated the temperature of the plants. For the red cotton, the UAV overestimated the temperature with respect to the net radiometer and thermocouples by 2.85% and 2.60%, respectively, while for the green cotton, it overestimated by 3.94% and underestimated by 1.1%. The trend can be explained by the measuring location of temperature and atmosphere effect. The temperature at the top of the plant canopy should be higher than at the middle layer of the plant. Since UAV measures of the canopy surface, especially the top area, UAV overestimated the temperature. Although UAV thermal imagery was collected at the lower altitude (40 m) than the conventional platform, the higher atmosphere temperature (34.11°C) than canopy affected the thermal camera capturing the energy in the wavelength range. The leaf level measurements showed a greater variability than the canopy temperature derived by the UAV as shown by the standard deviations in Table 1. It is well known that leaf temperature measurements tend to be rather variable, since they are strongly influenced by the

angle of incidence of SW_{in} [35]. Although the temperature measured by the thermocouple was closer to it of the net radiometer, the UAV measurements were in agreement with the ground sensors and were successful in showing temperature differences between the plots (see Table 2).

4. Conclusions

In this study, a UAV-based thermal sensor system was developed to measure canopy temperature using a radiometric calibrated thermal camera. Geo-tagged thermal imagery was collected over a cotton field including two different cotton genotypes exhibiting red and green leaves. Raw thermal data was converted to surface temperature using Matlab SDK provided by the sensor manufacturer. Radiometric calibration was performed using environmental parameters such as emissivity and weather conditions. Canopy temperature from UAV measurements was compared with that of net radiometer and thermocouples. The results show that the UAV system slightly overestimated canopy temperature when compared to the ground sensors. However, the errors did not exceed 5%, thus showing that the deviations were small and not significant for practical purposes. Additionally, the UAV system was successful in showing temperature differences between the plots. The proposed method showed the advantages of measuring canopy temperature and generating crop stress index map without field-based measurement. In the future, UAV and ground-based multiple thermal datasets will be collected for a newly designed plot to verify the proposed UAV-based thermal system for precision agriculture.

The proposed methodology could be applied to various agriculture fields to monitor crops for water stress and possibly the development of precision irrigation management applications in the future. However, there are a number of challenges that need to be addressed in order to make UAV technology practical for large commercial operations. One of the current limitations is the difficulty to generate precise geo-referenced orthomosaic thermal image without ground control points. Timely processing of a large volume of UAV data is also an issue that needs attention. Especially for crop

TABLE 2: The average temperature and standard deviation of ground and UAV system in green and red cotton.

Sensors	Cotton	Average (°C)	Standard deviation	Difference (°C)
Net radiometer	Red	33.69	—	1.73
	Green	31.96	—	
Thermocouple	Red	33.77	1.84	0.17
	Green	33.60	1.44	
UAV	Red	34.65	0.39	1.43
	Green	33.22	0.49	

precision management applications, the time between data collection and the output of actionable information needs to be drastically reduced.

Data Availability

The data used to support the findings of this study have not been made available because it is only available for researcher and collaborators of Texas A&M AgriLife Research and Extension.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This research was supported by Texas A&M AgriLife Research.

References

- [1] R. D. Jackson, S. B. Idso, R. J. Reginato, and P. J. Pinter, "Canopy temperature as a crop water stress indicator," *Water Resources Research*, vol. 17, no. 4, pp. 1133–1138, 1981.
- [2] X. Yao, N. Wang, Y. Liu et al., "Estimation of wheat LAI at middle to high levels using unmanned aerial vehicle narrow-band multispectral imagery," *Remote Sensing*, vol. 9, no. 12, article 1304, 2017.
- [3] C. Atzberger, "Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs," *Remote Sensing*, vol. 5, no. 2, pp. 949–981, 2013.
- [4] J. Liu, E. Pattey, and G. Jégo, "Assessment of vegetation indices for regional crop green LAI estimation from Landsat images over multiple growing seasons," *Remote Sensing of Environment*, vol. 123, pp. 347–358, 2012.
- [5] J. Enciso, M. Maeda, J. Landivar, J. Jung, and A. Chang, "A ground based platform for high throughput phenotyping," *Computers and Electronics in Agriculture*, vol. 141, pp. 286–291, 2017.
- [6] A. Chang, J. Jung, M. M. Maeda, and J. Landivar, "Crop height monitoring with digital imagery from Unmanned Aerial System (UAS)," *Computers and Electronics in Agriculture*, vol. 141, pp. 232–237, 2017.
- [7] D. Anthony, S. Elbaum, A. Lorenz, and C. Detweiler, "On crop height estimation with UAVs," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4805–4812, Chicago, IL, USA, September 2014.
- [8] A. Patrick and C. Li, "High throughput phenotyping of blueberry bush morphological traits using unmanned aerial systems," *Remote Sensing*, vol. 9, no. 12, article 1250, 2017.
- [9] A. S. Laliberte, M. A. Goforth, C. M. Steele, and A. Rango, "Multispectral remote sensing from unmanned aircraft: image processing workflows and applications for rangeland environments," *Remote Sensing*, vol. 3, no. 11, pp. 2529–2551, 2011.
- [10] K. Ramamurthy, Z. Zhang, and A. M. Thompson, "Predictive modeling of Sorghum phenotypes with airborne image features," in *Proceeding of KDD Workshop on Data Science for Food, Energy, and Water*, San Francisco, CA, USA, August 2016.
- [11] A. Zhang, A. Masjedi, J. Zhao, and M. M. Crawford, "Prediction of sorghum biomass based on image based features derived from time series of UAV images," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 6154–6157, Fort Worth, TX, USA, July 2017.
- [12] A. Masjedi, J. Zhao, A. M. Thompson et al., "Sorghum biomass prediction using UAV-based remote sensing data and crop model simulation," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 7719–7722, Valencia, Spain, July 2018.
- [13] R. Vadivambal and D. S. Jayas, "Applications of thermal imaging in agriculture and food industry—a review," *Food and Bioprocess Technology*, vol. 4, no. 2, pp. 186–199, 2011.
- [14] J. A. J. Berni, P. J. Zarco-Tejada, L. Suarez, and E. Fereres, "Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 722–738, 2009.
- [15] P. J. Zarco-Tejada, V. González-Dugo, and J. A. J. Berni, "Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera," *Remote Sensing of Environment*, vol. 117, pp. 322–337, 2012.
- [16] V. Gonzalez-Dugo, P. Zarco-Tejada, E. Nicolás et al., "Using high resolution UAV thermal imagery to assess the variability in the water status of five fruit tree species within a commercial orchard," *Precision Agriculture*, vol. 14, no. 6, pp. 660–678, 2013.
- [17] J. A. J. Berni, P. J. Zarco-Tejada, G. Sepulcre-Cantó, E. Fereres, and F. Villalobos, "Mapping canopy conductance and CWSI in olive orchards using high resolution thermal remote sensing imagery," *Remote Sensing of Environment*, vol. 113, no. 11, pp. 2380–2388, 2009.
- [18] S. Barbagallo, S. Consoli, and A. Russo, "A one-layer satellite surface energy balance for estimating evapotranspiration rates and crop water stress indexes," *Sensors*, vol. 9, no. 1, pp. 1–21, 2009.
- [19] P. Baranowski, W. Mazurek, B. Witkowska-Walczak, and C. Sławiński, "Detection of early apple bruises using pulsed-

- phase thermography,” *Postharvest Biology and Technology*, vol. 53, no. 3, pp. 91–100, 2009.
- [20] A. Manickavasagan, D. S. Jayas, N. D. G. White, and J. Paliwal, “Technical Note: Wheat class identification using thermal imaging: a potential innovative technique,” *Transactions of the ASABE*, vol. 51, no. 2, pp. 649–651, 2008.
- [21] G. Ginesu, D. D. Giusto, V. Märgner, and P. Meinlschmidt, “Detection of foreign bodies in food by thermal image processing,” *IEEE Transactions on Industrial Electronics*, vol. 51, no. 2, pp. 480–490, 2004.
- [22] D. M. Bulanon, T. F. Burks, and V. Alchanatis, “Study on temporal variation in citrus canopy using thermal imaging for citrus fruit detection,” *Biosystems Engineering*, vol. 101, no. 2, pp. 161–171, 2008.
- [23] K. Ribeiro-Gomes, D. Hernández-López, J. Ortega, R. Ballesteros, T. Poblete, and M. Moreno, “Uncooled thermal camera calibration and optimization of the photogrammetry process for UAV applications in agriculture,” *Sensors*, vol. 17, no. 10, article 2173, 2017.
- [24] S. Park, D. Ryu, S. Fuentes, H. Chung, E. Hernández-Montes, and M. O’Connell, “Adaptive estimation of crop water stress in nectarine and peach orchards using high-resolution imagery from an unmanned aerial vehicle (UAV),” *Remote Sensing*, vol. 9, no. 8, p. 828, 2017.
- [25] L. G. Santesteban, S. F. di Gennaro, A. Herrero-Langreo, C. Miranda, J. B. Royo, and A. Matese, “High-resolution UAV-based thermal imaging to estimate the instantaneous and seasonal variability of plant water status within a vineyard,” *Agricultural Water Management*, vol. 183, pp. 49–59, 2017.
- [26] L. Morillas, M. García, H. Nieto et al., “Using radiometric surface temperature for surface energy flux estimation in Mediterranean drylands from a two-source perspective,” *Remote Sensing of Environment*, vol. 136, pp. 234–246, 2013.
- [27] D. Neukam, H. Ahrends, A. Luig, R. Manderscheid, and H. Kage, “Integrating wheat canopy temperatures in crop system models,” *Agronomy*, vol. 6, no. 1, p. 7, 2016.
- [28] S. B. Idso, R. D. Jackson, W. L. Ehrlner, and S. T. Mitchell, “A method for determination of infrared emittance of leaves,” *Ecology*, vol. 50, no. 5, pp. 899–902, 1969.
- [29] J. Monteith and M. Unsworth, *Principles of Environmental Physics*, Academic Press, Cambridge, USA, 3rd edition, 2008.
- [30] “Radiometric Temperature Measurements for sUAS,” June 2020, <https://www.flir.com/discover/suas/radiometric-temperature-measurements-for-suas/>.
- [31] National Centers for Environmental Information June 2020, <https://www.ncdc.noaa.gov/>.
- [32] S. Harwin, A. Lucieer, and J. Osborn, “The impact of the calibration method on the accuracy of point clouds derived using unmanned aerial vehicle multi-view stereopsis,” *Remote Sensing*, vol. 7, no. 9, pp. 11933–11953, 2015.
- [33] F. J. Mestas-Carrascosa, J. Torres-Sánchez, I. Clavero-Rumbao et al., “Assessing optimal flight parameters for generating accurate multispectral orthomosaics by UAV to support site-specific crop management,” *Remote Sensing*, vol. 7, no. 10, pp. 12793–12814, 2015.
- [34] J. J. Burke, J. L. Hatfield, and D. F. Wanjura, “A thermal stress index for cotton,” *Agronomy Journal*, vol. 82, no. 3, pp. 526–530, 1990.
- [35] C. B. Tanner, “Plant temperatures1,” *Agronomy Journal*, vol. 55, no. 2, pp. 210–211, 1963.

Research Article

Prediction Model of Thermal Thawing Sensibility and Thaw Depth for Permafrost Embankment along the Qinghai-Tibet Engineering Corridor Using MODIS Data

Fuqing Cui ¹, Jianbing Chen ², Zhiyun Liu ¹, Wu Zhu ¹, Wei Wang ¹,
and Wei Zhang ¹

¹College of Geological Engineering and Geomatics, Chang'an University, Xi'an, Shaanxi 710054, China

²State Key Laboratory of Road Engineering Safety and Health in Cold and High-Altitude Regions, CCCC First Highway Consultants Co. Ltd, Xi'an, Shaanxi 710065, China

Correspondence should be addressed to Zhiyun Liu; dcdgx33@chd.edu.cn

Received 23 May 2020; Revised 22 June 2020; Accepted 23 July 2020; Published 17 August 2020

Academic Editor: Sang-Hoon Hong

Copyright © 2020 Fuqing Cui et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The aim of this paper was to reveal the distribution law of permafrost thermal thawing sensibility and thaw depth caused by road construction in Qinghai-Tibet engineering corridor (QTEC). The prediction models of permafrost thermal thawing sensibility and thaw depth have been developed by incorporating the MODIS and in situ soil temperature observation data. The comprehensive earth-atmosphere-coupled numerical models of different embankment structures have been utilized to calculate the thaw depth of the underlying permafrost foundation. Finally, using the given data and above developed prediction models, the distribution maps of permafrost thermal thawing sensibility and thaw depth in QTEC are obtained by grid calculation. The results show the following: (1) Insensitive permafrost of QTEC mainly distributes in the large-scale mountain and high latitude area, and highly sensitive permafrost is located in the perennial river bed, flood plain, and terrace regions. (2) Road construction has a strong thermal disturbance to underlying permafrost, and the proportion of large thaw depth area of separate embankment is obviously smaller than that of 26 m full-width embankment. (3) Increase of subgrade interval reduces the proportion of large thaw depth areas, and the application of separate embankment structure is an effective engineering means for the Qinghai-Tibet expressway.

1. Introduction

With the construction of the Qinghai-Tibet Railway, a series of communication pipelines, oil and gas pipelines, and power grid transmission projects have plan to build in QTEC [1]. The construction of the Qinghai-Tibet expressway has also been put on the agenda, which is included in the “National Highway Network Planning” and “National Highway Network Planning (2013~2030).” Therefore, ensuring the safety of major projects, avoiding engineering instability and reducing the occurrence of geological disasters will be the major issues that need to be solved in the engineering construction of the Qinghai-Tibet plateau permafrost regions [2–4]. Per-

mafrost is a special natural and ecological product of the Qinghai-Tibet Plateau, and its spatial distribution characteristics directly affect the safety and stability of engineering constructions [5–7]. Influenced by human activities or engineering structures, the frozen soil at the permafrost table will melt. The different ground temperature, surface characteristics, and ice-containing conditions will cause different thermal respond sensibility to external thermal disturbance and make the difference of thaw depth. The thermal thawing sensibility (S_e) can be used as the indicator of frozen soil thermal responds speed [8], and it is also one of the key influencing factors of permafrost characteristics. Thus, the research on the permafrost thermal thawing sensibility and thaw depth

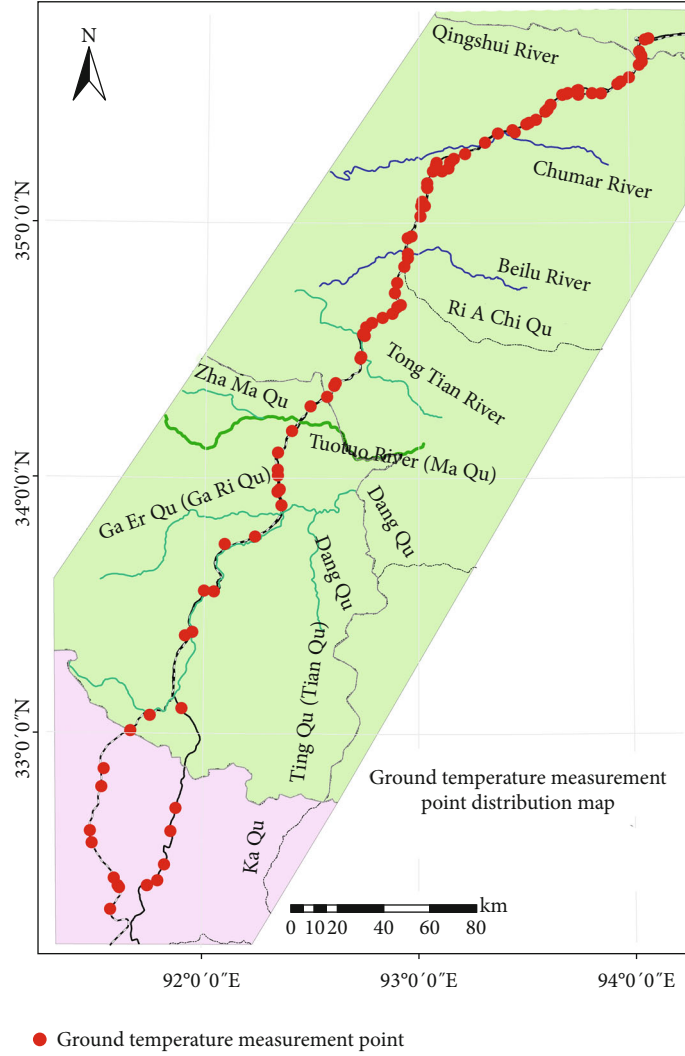


FIGURE 1: Soil temperature monitoring sections along the QTEC.

TABLE 1: Parameter estimation and test of each variable in regression equation.

Model	w_i	α_i	S.E.	Wals	Sig
Sunny slope	Constant	-218.244	105.897	4.247	0.039
	Latitude	4.009	1.612	6.187	0.013
	Elevation	18.814	11.759	2.560	0.110
	LST	-3.821	1.606	5.659	0.017
Shady slope	Constant	-455.593	176.390	6.671	0.010
	Latitude	9.036	3.342	7.312	0.007
	Elevation	43.661	17.419	6.282	0.012
	Equivalent latitude	-1.769	0.739	5.735	0.017
	NDVI	83.510	33.750	6.123	0.013

distribution in QTEC is not only a prerequisite for environmental monitoring, management, and protection but also a necessary requirement for the engineering structures design and construction [9–12].

The coupled heat transfer processes between permafrost and the external environment are quite complex and are affected by many factors such as latitude, longitude, elevation, vegetation coverage, soil characteristics, and climate environment [13, 14]. So distribution characteristics of thermal thawing sensibility are very regional and special [6, 8, 15]. Wu et al. [8] proposed that the permafrost thermal thawing sensibility has a strong correlation with frozen soil temperature and is also related to the seasonal thaw depth (active layer thickness) of frozen soil. It is essential to obtain the distribution of above two influence factors for thermal thawing sensibility determination. Earlier studies on the regionalization of permafrost temperature mainly used artificial drilling and monitoring methods [16]. But the vast area of Qinghai-Tibet Plateau and harsh natural climatic conditions prevented the researchers from obtaining complete, accurate, and timely geothermal observation data, which can usually only be determined the permafrost temperature distribution in a certain area. In recent years, with the development of remote sensing technology and the access to related data, the quantitative studies on the permafrost

TABLE 2: Comparison of permafrost logistic identification model and measured data.

Model	Area category	Observation number	Permafrost judgment		Prediction accuracy, %
			Yes	No	
Sunny slope	Permafrost	65	61	4	93.8
	Talik	9	1	8	88.9
Shady slope	Permafrost	39	36	3	92.3
	Talik	8	1	7	87.5

TABLE 3: Estimation and statistical test of the coefficients of MAGT prediction model.

Parameters	Coefficient values	t	Sig	Collinear statistics	
				Tolerance	VIF
Constant	56.871	10.930	0	—	—
φ	-0.952	-9.928	0	0.307	3.259
H	-4.736	-9.233	0	0.469	2.134
φ'	-0.067	-3.574	0.001	0.843	1.186
N	-5.425	-2.322	0.022	0.611	1.636

TABLE 4: Estimation and statistical test of the coefficients of ALD prediction model.

Model	Parameters	Coefficient values	S.E.	t	Sig
$T_g < -1^\circ\text{C}$	Constant	14.142	6.000	2.357	0.020
	T_g	0.157	0.035	4.544	0
	φ	-0.140	0.064	-2.192	0.03
	N	-0.970	0.602	-1.612	0.109
$T_g \geq -1^\circ\text{C}$	Constant	11.617	5.300	2.192	0.03
	T_g	0.579	0.089	6.522	0
	φ	-0.108	0.057	-1.908	0.058
	N	-1.195	0.606	-1.972	0.05

TABLE 5: Partial correlation analysis between the permafrost thermal thawing sensibility and the MAGT and ALD.

Parameters	Analysis type	MAGT	ALD	Se
MAGT	Pearson correlation (0.01 level)	1.0	0.733	0.952
	Significance	—	0	0
ALD	Pearson's correlation (0.01 level)	0.733	1.0	0.811
	Significance	0	—	0

distribution, soil temperature, thermal thawing sensibility distribution based on digital elevation model (DEM), and MODIS data have been rapidly developed [17, 18]. Meanwhile, an automatic permafrost temperature monitoring system based on the integrated Global System for Mobile Communications-Railway (GSM-R) and General Packet

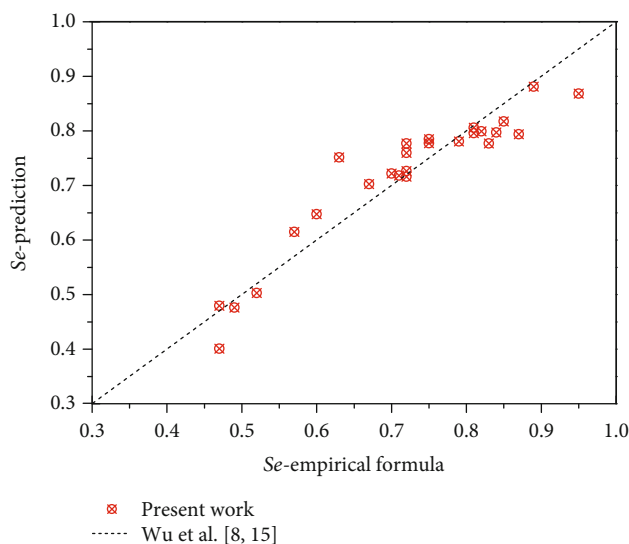
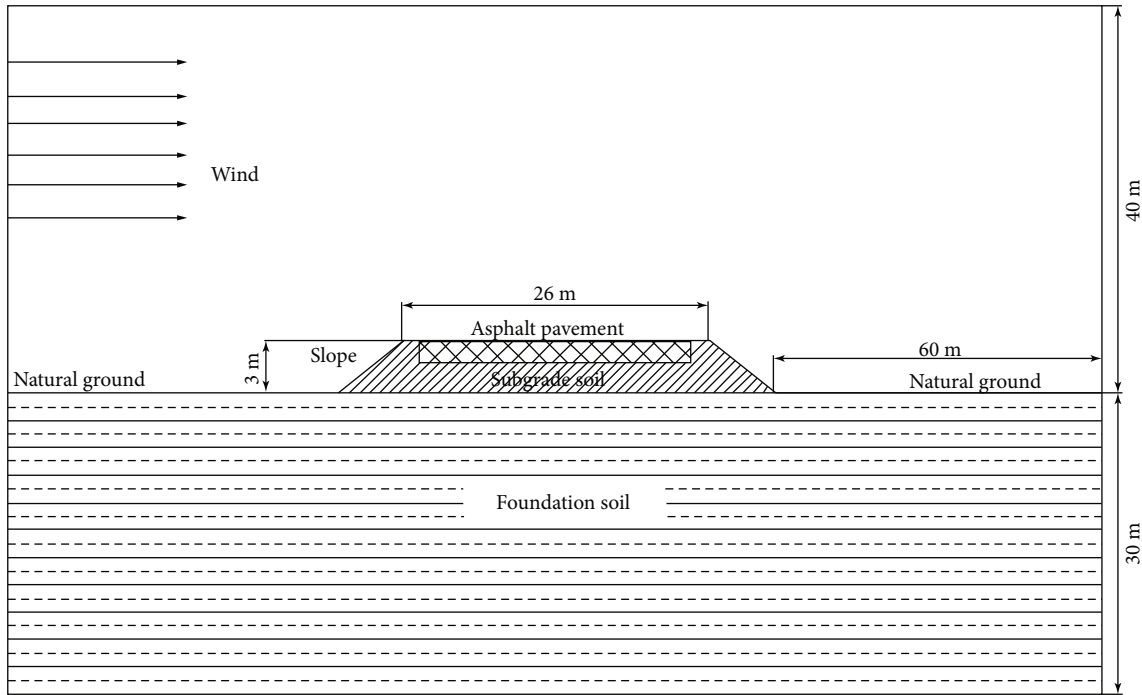


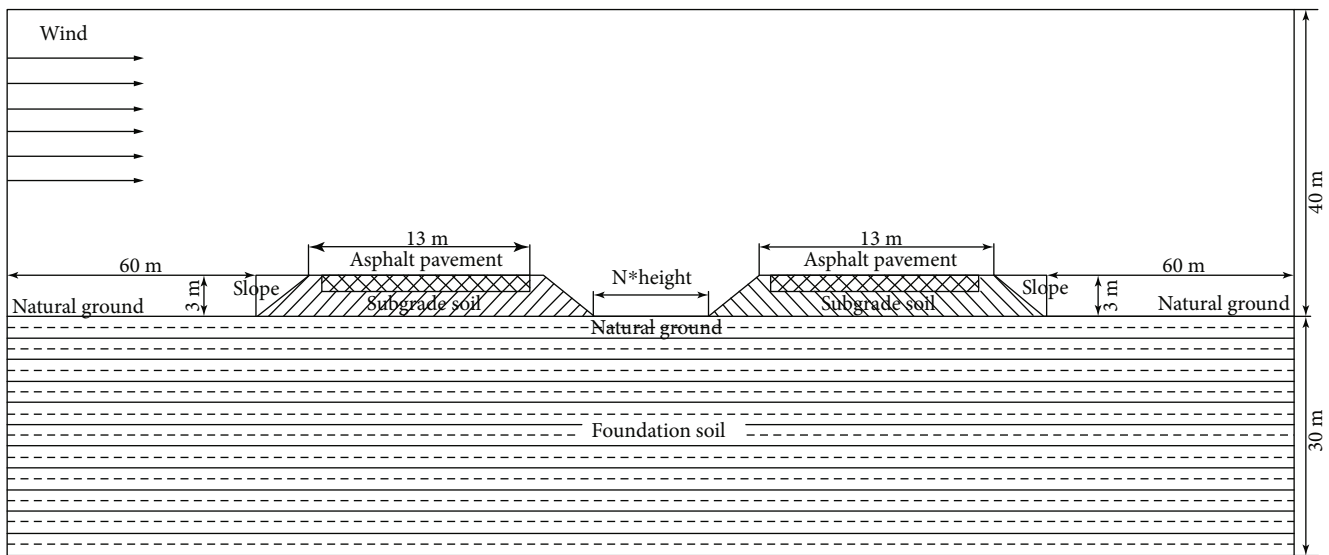
FIGURE 2: Comparison of the thermal thawing sensibility calculation results between the prediction model and the empirical formula.

Radio Service technology (GPRS) has also been applied continuously, thereby realizing permafrost temperature's real-time monitoring and analyzing [16, 19, 20].

For the permafrost thaw depth research, Nelson and Shiklomanov and Nelson [21, 22] used the Stefan method to map the high-precision distribution of active layer in the Kuparuk watershed of northern Alaska. Klene et al. [23] considered the thermal performance and topography of the melt index, soil, water, and surface coverage characteristics of the study area, used the Stefan formula to calculate the active layer thickness of the urbanized area of arctic, and developed a regional probability stochastic model of active layer thickness. Pang et al. [24] considered the influence of vegetation and soil properties on the active layer thickness and obtained active layer thickness distribution map of Qinghai-Tibet Plateau using the Kudryavtsev formula. Wu et al. [25] analyzed the ground temperature data of 10 field observation sites in the Beiluhe region from 2002 to 2012, and the results showed that the average increase of active layer thickness was 4.26 cm/year, and the main reason was the increased rainfall in summer. Chaves et al. [26] monitored the soil temperature of the Keller peninsula in Antarctica, obtained the characteristics of the interannual variability of active layer thickness and near-surface soil heat flow under the climate warming condition, and proved that the thermal regime of patterned



(a) 26 m full-length expressway embankment model



(b) 13 m width separate embankment model

FIGURE 3: Schematic of the permafrost embankment computation model.

ground soils was sensitive to air temperatures and atmospheric variations.

With the climate warming, the degradation of the underlying permafrost and the increasing thaw depth caused by engineering construction in QTEC are becoming more serious, and the accompanying engineering risks are also increasing. Therefore, the distribution law of thermal thawing sensibility and thaw depth are the important research topics for permafrost engineering. However, in most of recent studies on the thermal thawing sensibility and thaw depth, the research works are based on the long-term soil temperature monitoring and required a relatively complex

mathematical calculation process to obtain the permafrost thermal thawing sensibility [8]. Furthermore, restricted by environmental and economic factors, it cannot arrange a sufficient soil temperature monitoring site along the QTEC, which makes it impossible to obtain the distribution map of permafrost thermal thawing sensibility and thaw depth in the entire corridor. In the present work, to reveal the distribution law of permafrost thermal thawing sensibility and thaw depth caused by road construction in QTEC, the corresponding prediction models have been developed by incorporating the MODIS and in situ soil temperature observation data. The comprehensive earth-atmosphere-

TABLE 6: Computation cases considering for the distribution pattern of top layer permafrost and external environment conditions in QTEC.

Case indicator	Permafrost type	Moisture content	Air temperature, °C
1	Clay	0.2	-4.5
2	Clay	0.15	-5.5
3	Clay	0.25	-3.5
4	Clay	0.3	-3.0
5	Gravel	0.07	-4.5
6	Gravel	0.13	-5.5
7	Gravel	0.17	-3.0
8	Gravel	0.22	-3.5
9	Mudstone	0.2	-4.5
10	Mudstone	0.06	-3.5
11	Mudstone	0.1	-3.0
12	Mudstone	0.14	-5.5

coupled numerical models of 26 m full-width and 13 m separate embankment are used to calculate the thaw depth of the underlying permafrost foundation since road construction. Finally, using the given data and above developed prediction models, the distribution maps of permafrost thermal thawing sensibility and thaw depth in QTEC are obtained. The research results of this paper can contribute to the thermal design and maintenance of engineering structure in permafrost regions.

2. Methodology

2.1. Data and Data Processing

2.1.1. Remote Sensing Data. The Qinghai-Tibet engineering corridor is chosen as study area in this study, which is located in the hinterland of the Qinghai-Tibet plateau, and permafrost and frozen ground are widely distributed there. Three types of remote sensing data are collected: (1) Moderate-resolution Imaging Spectrometer (MODIS) Land Surface Temperature (LST) data with a spatial resolution of 1000×1000 m and a temporal resolution of 8 days that covered the period from 2000 to 2016; (2) MODIS Normalized Difference Vegetation Index (NDVI) data with a spatial resolution of 250×250 m and a temporal resolution of 16 days that covered the period from 2000 to 2016; and (3) Shuttle Radar Topography mission (SRTM) Digital Elevation Model (DEM) data with a spatial resolution of 90×90 m. The MODIS LST and NDVI datasets (MOD11A2, MOD13Q1) were downloaded from the Land Processes Distributed Active Archive Center (LPDAAC) of NASA (available at <https://lpdaac.usgs.gov/>), and the DEM SRTM3 dataset was downloaded from the Consortium for Spatial Information (CSI) of CGIAR (available at <http://srtm.csi.cgiar.org/>) [27, 28].

The daily LST is calculated by the average value of the 11:30 am and 11:30 pm MODIS LST data product. As the cloud, snow, or sensors sometimes induce outliers, NASA

adopts the mask processing or outlier identification for the daily LST data during product distribution. The average LST is obtained by the band calculation method using the effective value of daily LST.

The vegetation index (VI) reflects the difference between the reflection of vegetation in visible light, near-infrared band, and soil background, which is widely utilized for indicating the status of vegetation growth [19]. In our study, we calculated the annual NDVI values to analyze the trends in vegetation change. The NDVI is expressed on a scale of -1 to +1. The spatial distribution characteristics of surface vegetation and nonvegetation can be determined by reasonable selection of thresholds. In the present study, the average vegetation index, N , is calculated by the arithmetic mean of NDVI which is greater than 0.

Slope is an important index to describe terrain information, and it mainly indicates the inclination of the ground surface. Using the SRTM3-DEM data, surface analysis has been carried out in the ARCGIS 10.2 Software to generate the slope and aspect distribution of the study area. And the equivalent latitude is calculated and applied for determining the sunny or shady of slope direction [29].

$$\varphi' = \sin^{-1}(\sin k \cos l \cos \varphi + \cos k \sin \varphi), \quad (1)$$

where φ is the latitude, k is the slope of location, φ' is the equivalent latitude, and l is the aspect of location. When the equivalent latitude is less than or equal to the latitude, it is the sunny slope, and vice is shady.

2.1.2. In Situ Observation Data. To investigate the long-term changes of permafrost along the QTEC, a series of soil temperatures monitoring sections are built along the Qinghai-Tibet railway and highway, which is of 127 mean annual ground temperature (MAGT) monitoring points and 371 active layer depth (ALD) monitoring points. As shown in Figure 1, the monitoring sections are located from Xidatan to the southern foot of Tanggula Mountain, which is also the main permafrost distribution region in QTEC. Soil temperatures are measured at depths of 0.5 m to 30.0 m using thermal probes with a precision of $\pm 0.02^\circ\text{C}$, and these measurements are made twice a month. The ALD of a certain year is defined as the maximum depth at which the ground temperature is zero in this year. The MAGT is defined as the soil temperature of certain depth basically does not change with time. Thus, the ALD and MAGT of 2016 are calculated from the ground temperature observations.

2.2. Prediction Model of Thermal Thawing Sensibility

2.2.1. Logistic Probability Identification of Permafrost. Identifying the permafrost distribution in QTEC is the primary work for the research of the permafrost thermal thawing sensitivity. The logistic probability identification model of permafrost has been utilized in the identification of permafrost distribution. Considering the environment and terrain factors related to the permafrost distribution, the calculation formula of the probability of permafrost

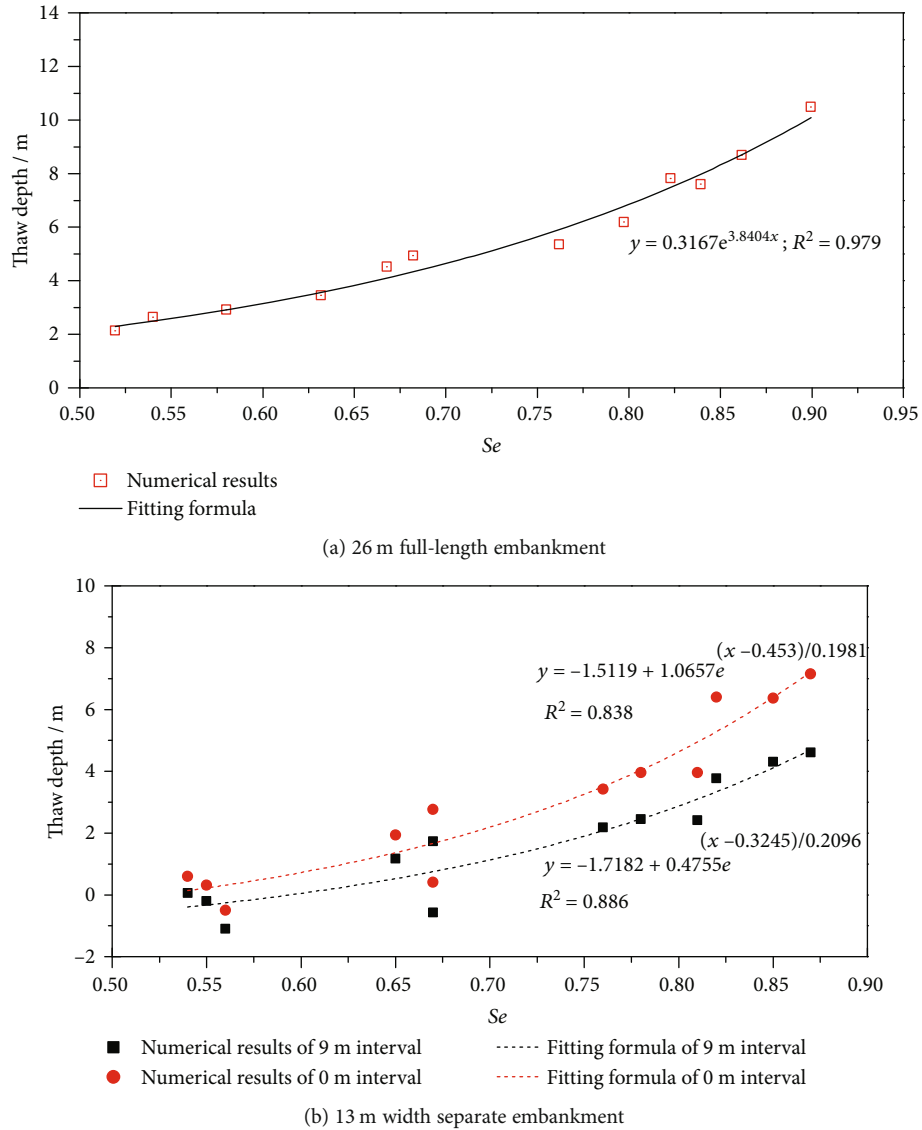


FIGURE 4: Underlying permafrost thaw depth for different thermal thawing sensibility since the embankment construction finished 20 years.

identification is

$$P = \frac{e^{w_0 + \sum_{i=1}^n \alpha_i w_i}}{1 + e^{w_0 + \sum_{i=1}^n \alpha_i w_i}}, \quad (2)$$

where w_0 is the constant of fitting formula, (w_1, w_2, \dots, w_n) are the quantitative indicators of environment, terrain, and other factors related to the permafrost distribution, and $(\alpha_1, \alpha_2, \dots, \alpha_n)$ are the coefficients for corresponding multiple regression equation. To achieve better fitting goodness, the prediction accuracy of different combinations of fitting parameters of sunny and shady slope models has been compared. It is found that the influence of LST is significantly greater than equivalent latitude and NDVI for the sunny slope model, and vice versa. Therefore, the optimal parameter combinations of sunny and shady slope models are given in Table 1. Furthermore, statistical test on parameters in the regression model has also been done. It can be seen that the

Wals value of every parameters are large and the Sig value is almost less than 0.05, which means each variable in the equation has a high significance and the regression model is considered to pass the test.

According to filed survey data of the ratio of permafrost and talik region in QTEC and comparative analysis of prediction accuracy of different identification probability, a fixed value of 0.7 is determined as the identification probability of permafrost:

$$\text{Mark} = \begin{cases} 1 & \text{if } P \geq 0.7 \\ 0 & \text{if } P < 0.7. \end{cases} \quad (3)$$

Table 2 is the comparison results of permafrost logistic identification model and measured data. It can be inferred that for the sunny and shady slopes, the corresponding prediction accuracy of permafrost is 93.8% and 92.3%,

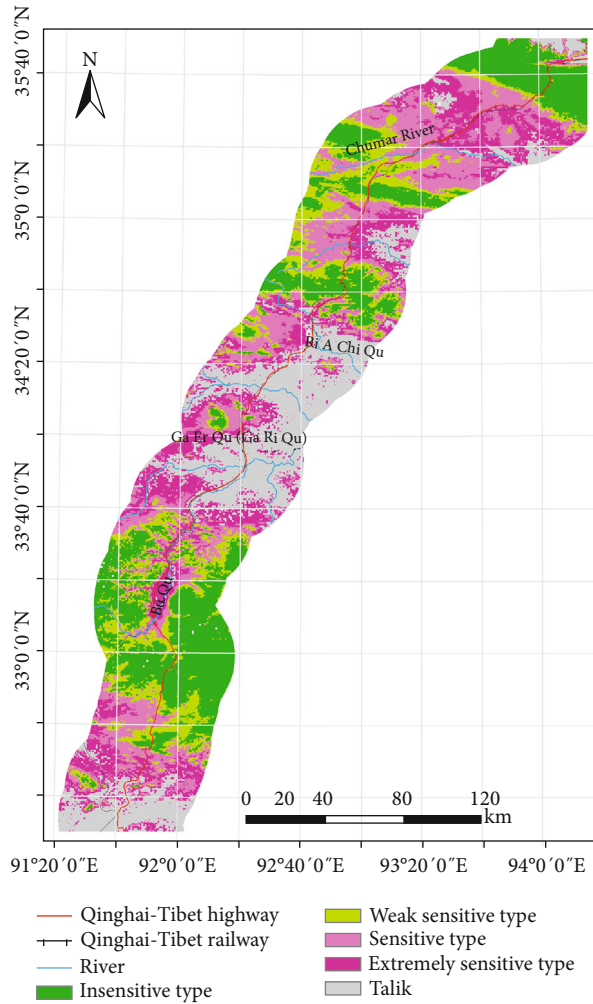


FIGURE 5: Distribution map of permafrost thermal thawing sensibility in QTEC.

respectively, and the prediction accuracy of talik is 88.9% and 87.5%, which validates accuracy of probability model.

2.2.2. Multiple Linear Regression Models of MAGT and ALD. The prediction of MAGT adopts the stepwise regression analysis strategy. The parameters are incorporated into multiple regression models one by one. The selected parameters are tested for significance and partial regression square value. During the fitting process, the significance level value of F is used as the criterion of the stepwise regression method, and the probability of selecting and excluding independent parameters are set to 0.05 and 0.01, respectively. Taking MAGT as the dependent variable and remote sensing data as the independent variables, the stepwise regression analysis is performed to obtain the multiple linear regression model of MAGT (T_g):

$$T_g = 56.871 - 0.952\varphi - 4.736H - 0.067\varphi' - 5.425N, \quad (4)$$

where φ is the latitude, H is the elevation, φ' is the equivalent latitude, and N is the arithmetic mean of vegetation index

TABLE 7: Classification criteria of permafrost thermal thawing sensibility.

Type	Insensitive	Weakly sensitive	Sensitive	Extremely sensitive
Range	≤ 0.54	0.54~0.66	0.66~0.8	> 0.8

TABLE 8: Prediction of the proportion of different permafrost thermal thawing sensibility regions in QTEC.

Type	Insensitive	Weakly sensitive	Sensitive	Extremely sensitive
Proportion	22.2%	13.8%	21.1%	16.5%

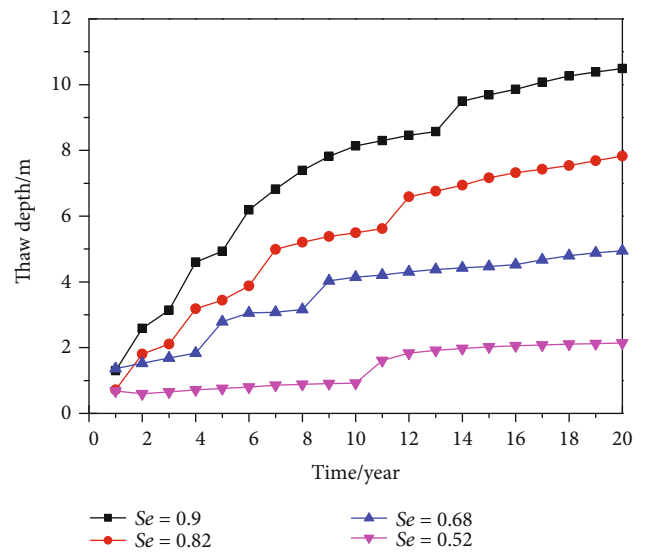


FIGURE 6: Simulated yearly thaw depth for the embankment with different thermal thawing sensibility.

greater than 0. The statistical test of MAGT prediction model is shown in Table 3. It shows that the Sig values of all fitting parameters are less than 0.05, which satisfy the significance requirements. From the perspective of collinearity, the parameters' tolerance value of prediction model is relatively large, and the VIF value is small, which can be concluded that the collinearity among various fitting parameters is excluded.

Considering the ALD is mainly influenced by MAGT, geographical location, surface characteristics, slope, etc., the above multiple regression method is used to obtain the ALD prediction model either. Using a total of 371 in situ observation data along QTEC, the multiple linear regression model of ALD (h) is calculated as follows:

$$h = B - \sum_{i=1}^n \beta_i x_i, \quad (5)$$

where B is the constant of fitting formula, (x_1, x_2, \dots, x_n) are the multivariate fitting parameters, and $(\beta_1, \beta_2, \dots, \beta_n)$ are the coefficients for corresponding multiple regression equation. The specific values of the coefficients of regression equation under different MAGT conditions are given in Table 4.

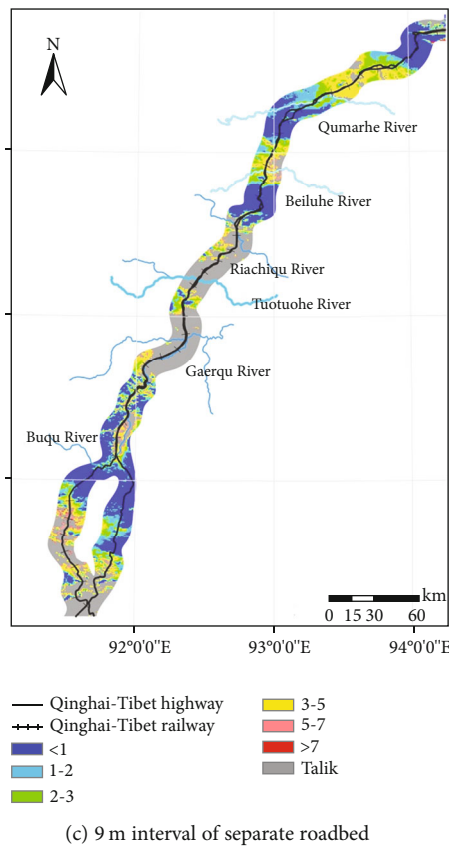
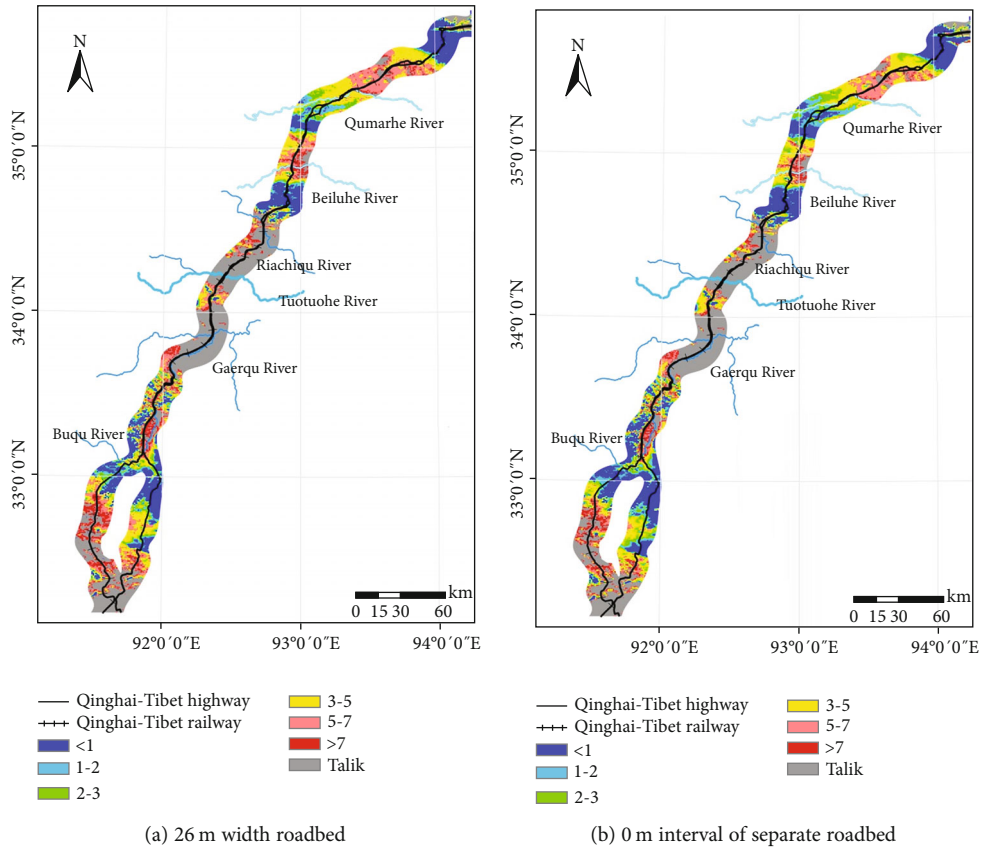


FIGURE 7: The thaw depth distribution of different embankment structures in QTEC.

TABLE 9: Prediction of the proportion distribution of thaw depth for different embankment structures in QTEC.

Thaw depth, m	Embankment structure		
	26 m full width, %	13 m separate 0 m interval, %	13 m separate 9 m interval, %
<1	16.1	19.3	27.8
1~2	5.2	7.8	11.0
2~3	6.4	8.1	11.3
3~5	17.9	16.5	14.7
5~7	15.8	11.1	3.4
>7	7.2	5.7	0.3
Talik	31.4	31.5	31.5

2.2.3. *Fitting Formula of Thermal Thawing Sensibility.* The partial correlation analysis between the permafrost thermal thawing sensitivity and the MAGT and ALD is shown in Table 5.

It can be seen from Table 5 that the permafrost thermal thawing sensitivity is significantly correlated with the MAGT and ALD. Through linear regression, the multiple linear regression model of the permafrost thermal thawing sensitivity is obtained as follows:

$$Se = 0.806 + 0.107T_g + 0.036h. \quad (6)$$

Figure 2 is the comparison of thermal thawing sensibility calculation results using the prediction model and the empirical formula. It can be seen that the predicted values agrees well with the calculated results of the Kudryavtsev formula using the soil temperature monitoring data [8, 15] ($R^2 = 0.935$), which testifies reasonableness and reliability of the model.

2.3. Prediction Model of Thaw Depth for Permafrost Embankment

2.3.1. *Numerical Calculation Model of Thaw Depth for Permafrost Embankment.* The numerical method is used to calculate the time variation of thaw depth of the underlying permafrost foundation after road construction. Based on the analysis of the complex heat transfer processes among the surrounding environment, embankment, and underlying permafrost, the earth-atmosphere-coupled open numerical model is developed [30, 31], and two corresponding physical models for numerical calculation are built: (i) the 26 m width expressway subgrade model with adjacent natural permafrost and air environment and (ii) the separate 13 m width subgrade model with different intervals and comprehensive earth-atmosphere coupled system. The schematic of the above two numerical models are shown in Figure 3.

Considering the distribution pattern of top layer permafrost and external environment conditions in QTEC, 12 computation cases for the above two models have been built to investigate the thaw depth of different soil types, moisture content, and yearly average air temperature [27] (see in Table 6). A 2D unsteady model has been developed in this work to predict the spatial and temporal thermal exchange

processes between the earth-atmosphere-coupled system. The standard κ - ϵ model is used for the air region turbulence simulation. The complex boundary conditions of air regions and source term of ground surface are imported into the FLUENT by the UDF program. The air temperature, wind speed, wind direction, solar radiation intensity, and other data are taken from the meteorological monitoring data, and the annually temperature ascending speed is defined as $0.022^\circ\text{C}/\text{year}$. The thermal thawing sensitivity of the above cases can be calculated from the temperature field computation results of the nonroadbed model in 100th year.

2.3.2. *Fitting Formula of Thaw Depth.* Using the developed numerical model, the underlying permafrost thaw depth of the 26 m full-length and separate embankment in 20th year of different thermal thawing sensitivity conditions is calculated (as shown in Figure 4). It can be seen that as the sensitivity of thermal ablation increases, the thaw depth, Δh (amount of the artificial permafrost table to natural permafrost table), increases as well, and the growth trend follows the exponential law. It infers that the thaw depth of permafrost embankment can be predicted with thermal thawing sensitivity.

For the formula of the fitting thaw depth and thermal thawing sensitivity of 26 m full-length and separate embankments, the results are as follows:

$$\begin{aligned} \Delta h_{w-26m} &= 0.3167e^{3.8404Se}, \\ \Delta h_{s-0m} &= -1.5119 + 1.0657e^{\frac{Se-0.453}{0.1981}}, \\ \Delta h_{s-9m} &= -1.7182 + 0.4755e^{\frac{Se-0.3245}{0.2096}}, \end{aligned} \quad (7)$$

where Δh_{w-26m} , Δh_{s-0m} , and Δh_{s-9m} are the thawing depth of 26 m full-width embankment, 0 m interval separate embankment, and 9 m interval separate embankment in 20th year, respectively. The determined coefficients (R^2) of three fitting formulas are 0.979, 0.838, and 0.886, which exhibit relative better fitness of the calculation formula.

3. Results

3.1. *Permafrost Thermal Thawing Sensibility Distribution.* Importing the LST, NDVI, and SRTM3-DEM data into ARCGIS 10.2 Software, the grid calculation has been done based on the thermal thawing sensibility prediction model (Eqs. (2)–(6)), and the permafrost thermal thawing sensibility distribution map can be obtained (as shown in Figure 5). The criteria for classification the thermal ablation sensitivity are shown in Table 7 [8, 15].

The results show that the present insensitive permafrost mainly distributes in the large-scale mountain and high latitude area, such as the Kunlun mountain, Fenghuo mountain, and Tanggula mountain. The highly sensitive permafrost is located in the perennial river bed, flood plain, and terrace regions, such as the Chumaer river, Tuotuohe river, and Tongtian river. Meanwhile, as shown in Table 8, sensitive and extremely sensitive permafrost account for 38% for the entire QTEC region based on the current data, and weakly sensitive and insensitive account for 13.8% and 22.2%,

respectively. It can be expected that permafrost in the QTEC will become more fragile, and the proportion of sensitive-type permafrost will also upsurge with global warming and intensified human engineering activities.

3.2. Thaw Depth Distribution of Wide and Separate Embankment. Figure 6 is the interannual variation of thaw depth of 26 m full-width embankment with different thermal thawing sensibility within 20 years. It can be seen that the thaw depth of the underlying permafrost foundation has a strong correlation with its thermal thawing sensitivity. With the increase of the thermal thawing sensitivity, the permafrost is more sensitive to the external thermal disturbance. Under the same thermal disturbance (3 m height embankment), permafrost thaw depth increases with thermal thawing sensitivity and time. It also should be noticed that speed of thawing becomes slower in the later period (>10 years), indicating that the external thermal disturbance has gradually penetrated into the interior of permafrost foundation, and the whole heat-exchange process gradually tends to a new dynamic equilibrium.

In order to further clear the thermal impact of separated roadbed and full width embankment on underlying permafrost in QTEC, the thaw depth distribution maps of different embankment structures have been calculated with the above proposed thaw depth prediction model (see in Figure 7). The figure shows that road construction has a strong thermal disturbance to the underlying permafrost. Similar to thermal thawing sensitivity distribution, the relatively small thaw depth areas are mainly distributed in the large-scale mountain and high latitude area, such as the Kunlun mountain, Fenghuo mountain, and Tanggula mountain. Their thaw depth is roughly less than 3 m, and some are even below 1 m. The large thaw depth area is mainly distributed in the perennial river bed, floodplain, and river terrace area of the Chumal River, Tuotuo River, or Tongtian River. The thaw depth is generally between 5 and 10 m. Furthermore, it also can be seen that the large thaw depth area of separate embankment is obviously smaller than that of 26 m full-width embankment, especially in the Chumar River, Beilu River, and Buqu River Valley regions. Therefore, it can be concluded that the heat disturbance of separate embankment to the underlying permafrost is weaker than that of the full-width embankment.

Table 9 shows the proportion distribution of thaw depth for different embankment structures in QTEC. It can be seen that the proportion of the talik zone in QTEC is basically unchanged (about 31%) for all three kinds of adopted embankment structure. However, the proportions of thaw depth between 5~7 m and >7 m of separate embankment are much less than that of 26 m full-width embankment, and the range of 0~2 m, vice versa, especially for the separate embankment with 9 m interval (best spacing). For the 26 m full-width embankment, the area ratio of large thaw depth (>5 m) is 22.9%, and the small thaw depth (<3 m) area proportion is 21.3%. Comparatively, large and small area proportion of separate embankment with 9 m interval is only 3.7% and 38.8%, respectively. The increase of subgrade interval also reduces the proportion of large thaw depth areas. In summary, the

use of separate embankment structure is an effective engineering means for the Qinghai-Tibet expressway.

4. Conclusion

The distribution law of permafrost thermal thawing sensibility along QTEC, the degradation of the underlying permafrost, and the increasing thaw depth caused by road construction are the important research topics of road engineering in permafrost regions. In the present work, the prediction models of permafrost thermal thawing sensibility and thaw depth have been developed by incorporating the remote sensing and in situ observation data. The comprehensive earth-atmosphere-coupled numerical models of 26 m full-width and 13 m separate embankment are used to calculate the thaw depth of the underlying permafrost foundation since road construction. Distribution maps of permafrost thermal thawing sensibility and thaw depth in QTEC are obtained by grid calculation in the ARCGIS 10.2 Software. The research results show that

- (1) The insensitive permafrost mainly distributes in the large-scale mountain and high latitude area, and highly sensitive permafrost is located in the perennial river bed, flood plain, and terrace regions. With climate warming and intensified human engineering activities, the permafrost in the QTEC will become more fragile, and the proportion of sensitive-type permafrost will also increase
- (2) The road construction has a strong thermal disturbance to the underlying permafrost, and the proportion of a large thaw depth area of separate embankment is obviously smaller than that of 26 m full-width embankment
- (3) The increase of subgrade interval reduces the proportion of large thaw depth areas. Therefore, the application of separate embankment structure is an effective engineering means for the Qinghai-Tibet expressway.

Data Availability

The raw LST, NDVI, and SRTM3 data used to support the findings of this study were downloaded from the Land Processes Distributed Active Archive Center of NASA and Consortium for Spatial Information of CGIAR, which are freely available at <https://lpdaac.usgs.gov/> and <http://srtm.csi.cgiar.org/> respectively.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Science Foundation of China (Grant Nos. 41502292, 51574037), the Natural Science Basic Research Plan in Shaanxi Province of China (2018JQ4031), the Fundamental Research Funds for the

Central Universities, CHD (Nos. 300102269207, 300102269303), and the Applied Fundamental Research Project of China Communications Construction Co, Ltd (Nos. 2018-ZJKJ-PTJS03, 2016-ZJKJ-02). We are grateful to the anonymous reviewers for their constructive comments to improve this manuscript.

References

- [1] F. Han, W. Yu, X. Yi, D. Hu, and Y. Lu, "Thermal regime of paved embankment in permafrost regions along the Qinghai-Tibet Engineering Corridor," *Applied Thermal Engineering*, vol. 108, pp. 330–338, 2016.
- [2] F. Yu, J. Qi, X. Yao, and Y. Liu, "Degradation process of permafrost underneath embankments along Qinghai-Tibet Highway: an engineering view," *Cold Regions Science and Technology*, vol. 85, pp. 150–156, 2013.
- [3] W. Yu, T. Zhang, Y. Lu, F. Han, Y. Zhou, and D. Hu, "Engineering risk analysis in cold regions: State of the art and perspectives," *Cold Regions Science and Technology*, vol. 171, article 102963, 2020.
- [4] Y. Lai, X. Xu, Y. Dong, and S. Li, "Present situation and prospect of mechanical research on frozen soils in China," *Cold Regions Science and Technology*, vol. 87, pp. 6–18, 2013.
- [5] C. Ya-ling, Y. Sheng, and W. Ma, "Study on the effect of the thermal regime differences in roadbed slopes on their thawing features in permafrost regions of Qinghai-Tibetan plateau," *Cold Regions Science and Technology*, vol. 53, no. 3, pp. 334–345, 2008.
- [6] Y. M. Lai, M. Y. Zhang, and S. Y. Li, *Theory and Application of Cold Regions Engineering*, Science Press, Beijing, 2009.
- [7] Q. Wu, X. Dong, Y. Liu, and H. Jin, "Responses of permafrost on the Qinghai-Tibet Plateau, China, to climate change and engineering construction," *Arctic, Antarctic and Alpine Research*, vol. 39, no. 4, pp. 682–687, 2007.
- [8] W. Qingbai, Z. Yuanlin, and L. Yonzhi, "Evaluation model of permafrost thermal stability and thawing sensibility under engineering activity," *Cold Regions Science and Technology*, vol. 34, no. 1, pp. 19–30, 2002.
- [9] P. P. Overduin, D. L. Kane, and W. K. P. van Loon, "Measuring thermal conductivity in freezing and thawing soil using the soil temperature response to heating," *Cold Regions Science and Technology*, vol. 45, no. 1, pp. 8–22, 2006.
- [10] M. Zhang, W. Pei, X. Zhang, and J. Lu, "Lateral thermal disturbance of embankments in the permafrost regions of the Qinghai-Tibet engineering corridor," *Natural Hazards*, vol. 78, no. 3, pp. 2121–2142, 2015.
- [11] Q. Yu, K. Fan, Y. You, L. Guo, and C. Yuan, "Comparative analysis of temperature variation characteristics of permafrost roadbeds with different widths," *Cold Regions Science and Technology*, vol. 117, pp. 12–18, 2015.
- [12] W. Pei, L. Jin, M. Zhang, S. Li, and Y. Lai, "Study of the time-dependent thermal behavior of the multilayer asphalt concrete pavement in permafrost regions," *Construction and Building Materials*, vol. 193, no. 30, pp. 162–172, 2018.
- [13] F. Niu, J. Luo, Z. Lin, W. Ma, and J. Lu, "Development and thermal regime of a thaw slump in the Qinghai-Tibet plateau," *Cold Regions Science and Technology*, vol. 83–84, pp. 131–138, 2012.
- [14] F. Ling and T. Zhang, "A numerical model for surface energy balance and thermal regime of the active layer and permafrost containing unfrozen water," *Cold Regions Science and Technology*, vol. 38, no. 1, pp. 1–15, 2004.
- [15] Q. B. Wu, Y. L. Zhu, and Y. Z. Liu, "Evaluation model of environmental change of frozen soil under human engineering activities," *Science China: Earth Science*, vol. 32, no. 2, pp. 141–148, 2002.
- [16] J. Yang, Q. H. Jiang, and P. Zhi, "Qinghai-Tibet railway permafrost regions ground temperature data transformation by GSM-R/GPRS," *Railway computer application*, vol. 17, no. 2, pp. 25–27, 2008.
- [17] S. Hachem, M. Allard, and C. Duguay, "Using the MODIS land surface temperature product for mapping permafrost: an application to Northern Québec and Labrador, Canada," *Permafrost and Periglacial Processes*, vol. 20, no. 4, pp. 407–416, 2009.
- [18] M. Langer, S. Westermann, and J. Boike, "Spatial and temporal variations of summer surface temperatures of wet polygonal tundra in Siberia - implications for MODIS LST based permafrost monitoring," *Remote Sensing of Environment*, vol. 114, no. 9, pp. 2059–2069, 2010.
- [19] A. G. Rike, K. B. Haugen, M. Børresen, B. Engene, and P. Kolstad, "In situ biodegradation of petroleum hydrocarbons in frozen arctic soils," *Cold Regions Science and Technology*, vol. 37, no. 2, pp. 97–120, 2003.
- [20] R. Xing, S. Jiang, and P. Xu, "Long-term temperature monitoring of tunnel in high-cold and high-altitude area using distributed temperature monitoring system," *Measurement*, vol. 95, pp. 456–464, 2017.
- [21] N. Shiklomanov and F. E. Nelson, "Active-layer mapping at regional scales: a 13-year spatial time series for the Kuparuk region, north-central Alaska," *Permafrost and Periglacial Processes*, vol. 13, no. 3, pp. 219–230, 2002.
- [22] N. I. Shiklomanov and F. E. Nelson, "Climatic variability in the Kuparuk Region, North-central Alaska: optimizing spatial and temporal interpolation in a sparse observation network," *Arctic*, vol. 56, no. 2, pp. 136–146, 2003.
- [23] A. E. Klene, F. E. Nelson, and N. I. Shiklomanov, "The N-factor in natural landscapes: variability of air and soil-surface temperatures, Kuparuk River Basin, Alaska, U.S.A.," *Arctic Antarctic and Alpine Research*, vol. 33, no. 2, pp. 140–148, 2001.
- [24] Q. Pang, L. Zhao, S. Li, and Y. Ding, "Active layer thickness variations on the Qinghai-Tibet Plateau under the scenarios of climate change," *Environmental Earth Sciences*, vol. 66, no. 3, pp. 849–857, 2012.
- [25] Q. Wu, Y. Hou, H. Yun, and Y. Liu, "Changes in active-layer thickness and near-surface permafrost between 2002 and 2012 in alpine ecosystems, Qinghai-Xizang (Tibet) Plateau, China," *Global and Planetary Change*, vol. 124, pp. 149–155, 2015.
- [26] D. A. Chaves, G. B. Lyra, M. R. Francelino, L. D. B. Silva, A. Thomazini, and C. E. G. R. Schaefer, "Active layer and permafrost thermal regime in a patterned ground soil in Maritime Antarctica, and relationship with climate variability models," *Science of the Total Environment*, vol. 584, pp. 572–585, 2017.
- [27] Y. Song, L. Jin, and H. Wang, "Vegetation changes along the Qinghai-Tibet Plateau Engineering Corridor since 2000 induced by climate change and human activities," *Remote Sensing*, vol. 10, no. 2, p. 95, 2018.
- [28] H. Frey and F. Paul, "On the suitability of the SRTM DEM and ASTER GDEM for the compilation of topographic parameters in glacier inventories," *International Journal of Applied Earth Observation and Geoinformation*, vol. 18, pp. 480–490, 2012.

- [29] I. R. Whitson, "Equivalent latitude for prediction of soil development in a complex mapunit," *Canadian Journal of Soil Science*, vol. 95, no. 2, pp. 125–137, 2015.
- [30] Z. Y. Liu, J. B. Chen, L. Jin, Y. J. Zhang, and C. Lei, "Roadbed temperature study based on earth-atmosphere coupled system in permafrost regions of the Qinghai-Tibet plateau," *Cold Regions Science and Technology*, vol. 86, pp. 167–176, 2013.
- [31] Z.-Y. Liu, F.-Q. Cui, J.-B. Chen, L. Jin, W. Wang, and W. Zhang, "Study on the permafrost heat transfer mechanism and reasonable interval of separate embankment for the Qinghai-Tibet expressway," *Cold Regions science and Technology*, vol. 170, article 102952, 2019.

Research Article

Developing a Data Model of Indoor Points of Interest to Support Location-Based Services

Alexis Richard C. Claridades ^{1,2} and Jiyeong Lee ¹

¹Department of Geoinformatics, University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, Seoul 02504, Republic of Korea

²Department of Geodetic Engineering, University of the Philippines Diliman, Quezon City 1101, Philippines

Correspondence should be addressed to Jiyeong Lee; jlee@uos.ac.kr

Received 21 April 2020; Revised 4 June 2020; Accepted 8 June 2020; Published 1 August 2020

Academic Editor: Sang-Hoon Hong

Copyright © 2020 Alexis Richard C. Claridades and Jiyeong Lee. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Focus on indoor spatial applications has been rising with the growing interest in indoor spaces. Along with the widespread use of mobile devices and the internet, it has increased demands for indoor location-based services (LBS), demanding more efficient representation and management of indoor spatial data. Indoor points of interest (Indoor POI) data, representing both spaces and facilities located indoors, provide the infrastructure for these services. These datasets are vital in delivering timely and accurate information to users, such as in cases of managing indoor facilities. However, even though there are studies that explore its use across applications and efforts exerted towards the standardization of the data model, most POI development studies have focused on the outdoors and remain underdeveloped in the indoors. In this paper, we propose a spatial-temporal Indoor POI data model to provide direction for the establishment of indoor POI data and to address limitations in currently available data specifications. By exploring how different Indoor POIs are from its outdoor counterparts, particularly on extending its outdoor counterparts' functions on searching, sharing, and labeling, we describe the data model and its components using the Unified Modeling Language (UML). We perform an SQL-based query experiment to demonstrate the potential use of the data model using sample data.

1. Introduction

Nowadays, day-to-day human activities have been closely tied with the use of mobile devices and gadgets, most equipped with GPS receivers and cameras, and are continuously improving in terms of features and speeds while decreasing in size [1, 2]. With this, the demand for information arose through location-based services (LBS), which aim to give users relevant and timely information based on their positions [1, 3], and augmented reality (AR) applications that combine images from the real-world to virtual images in three-dimensions [4]. These services form part of the core requirement of Smart Cities, as localities around the world aim to establish seamless integration of technology to the daily life of its citizens.

Now, as interest in indoor space continues to rise [5], the demand for spatial applications and services also increases. These technologies that signal that we are now living in a

digital world spark interest in digitizing real-world indoor scenes [6].

Indoor POI is a location in indoor space where information regarding a particular place, service, facility, or event is available, in contrast to traditional POIs located in outdoor environments. A reliable Indoor POI dataset is vital to provide the fundamental infrastructure to LBS, to provide successful services to users. This approach, however, is faced with several difficulties. First, Indoor POIs are not always identified by a proper name, as most features that they represent in indoor space consists of facilities, such as ticket machines, CCTV cameras, or fire extinguishers. Second, Indoor POIs are more appropriately referred to by their type, or classification, shifting the general POI idea of a location that is identified by a unique name. The basic definitions of POI from W3C [2] and OGC [7] have stated that a name is a primary component, together with ID and location. Third, most indoor LBS are still providing viewer-level service [8],

due to existing data models being underdeveloped in the indoor aspect and not differentiated with its outdoor counterparts. Established standards regarding primarily dealing with indoor data have dealt more with navigation, such as Open Geospatial Consortium (OGC) standard Indoor Geography Markup Language (IndoorGML) [9], not on representing features and spaces for LBS, and they have no precise specifications regarding Indoor POI in their respective models.

This paper is motivated by the requirement for the formalization of Indoor POI, to expand services such as in facilities management, simulation, and monitoring. As current applications demand further classification of POIs into levels—named objects with ID and location as Level 1, locations that include unnamed real objects such as indoor or outdoor facilities as Level 2, and intangible assets or events as Level 3 [2]. Indoor POIs, for instance, take prime importance in Level 2, as this has gained interest in the increasing demand for indoor applications, including indoor LBS, indoor facility management, increase in accessibility for disabled persons, evacuation for emergencies, and even commercial or robotic applications.

In response to the difficulties faced in dealing with Indoor POI stated above, this study proposes a data model that characterizes its vital aspects as essential elements in providing spatial services. Identifying these aspects and formalizing this model is key for assuring data quality, provide prospects for validation, encourage analysis, and at the same time, promoting data sharing and integration. Furthermore, we intend to demonstrate the potential and usability of the proposed data model through an implementation.

This paper is structured as follows. The next section discusses studies on efforts on POI standardization and data model, as well as notable utilization across application domains. The third section presents the characteristics of the Indoor POI data model, followed by the proposed spatial-temporal indoor data model. We conduct an experimental implementation of the data model through a use case involving facility management to demonstrate its various aspects through a small sample dataset, and the last section focuses on conclusions and limitations of this study to be addressed by future work.

2. Related Research

The rapid growth of mobile devices and internet technology has led to the acceleration of LBS applications, with a special interest in the indoors [10–12] due to its strong influence in people's daily lives [13], and recognition that these spaces are even as dynamic as its outdoor counterparts [14]. In recent years, the indoor environment has been a target of a wide area of research ranging from data acquisition, 3D data modeling, and indoor navigation [15].

Though the progress in outdoor mapping and navigation applications encourage the interest of its indoor counterpart, the indoor case postulates problems in the direct application of technologies as they exhibit different characteristics [16]. These problems include the lack of cheap and convenient positioning systems and the availability and complexity of maps [17]. However, as the key enabler for Indoor LBS

[18], Indoor POI can help augment these shortcomings, together with other datasets. Indoor POI is essential in positioning indoors as much as how GPS has enabled this outdoors [19].

To date, organizationally agreed-upon models regarding Indoor POI are still, and a standard is yet to be acknowledged [2]. One major problem that having an Indoor POI data standard would resolve is having separate sets of POIs for every application or infrastructure. In a study for the development of a participatory collection of Indoor POI, despite crowdsourcing having the advantage of a large volume of data, there is a necessity for data quality control and assurance due to the massive number of users and variety of devices. As such, a data model would prove useful in performing data integrity checks.

Tracing the timeline of POI standard development would begin with nodes in OpenStreetMap (OSM), KML, and Places Library of Google Maps but individually lacks in aspects that are essential to the direction of applications where POIs are going [3, 20] place identifiers in ISO 19112 [21] and ISO 19155 [22]. Specifically, the World Wide Web Consortium (W3C) POI Special Working Group (SWG) has become the first step towards the standardization for the definition and exchange with a focus on web architecture and in AR applications. SWG has published the Point of Interest Core (POI Core), describing eight categories to describe POI with attributes and various location types [1, 3].

Commercial GIS providers have also observed the emerging need for Indoor POI. ArcGIS Pro by ESRI has also included a provision for Indoor POIs through ArcGIS Indoors. In this application module, Indoor POIs can represent features on an indoor map. These Indoor POIs have two levels of classification, the first one being broader classes of people, places, events, and objects [23].

In parallel, industry stakeholders, including car industry specialists and experts in mobile technology, navigation systems, and digital maps, formulated a general-purpose specification called point of interest exchange language (POIX) and submitted this as a preliminary proposal to W3C. However, this lacked in categorical, descriptive, and temporal aspects and seemed inclined towards car navigation [20]. In Korea, the Telecommunications Technology Association (TTA) established a data model and other private companies, including Naver and Daum, to establish portal and navigation services. Government agencies also partner with research and development institutions to study POIs in the context of positioning and human-oriented geographic information [2].

A spatial-temporal model based on the W3C data model was proposed by [2], based on the W3C data model, extending certain aspects to the goal of expanding the utilization of POI. This model emphasized the three major roles and characteristics of POIs in providing LBS—in searching, in display, and in sharing—and is subsequently formulated according to each of these individual functions. Time series management is also enabled by expanding definitions for temporal aspects of POI in terms of changes in location over time for a POI (feature-based) and change in the POI over a period time for a location (location-based) [24]. However, this standard tackles the first level of POI, where a unique

name refers to each feature and suggests that expansion towards the higher levels is necessary in future studies.

The concept of Indoor POIs has been used by [8] to represent facilities for an indoor LBS. Spatial relationships between the Indoor POI and the indoor spaces abstracted with topological data provided by IndoorGML are defined to provide an indoor patrol service. Also, in an indoor setting, [25] proposed a location-aware POI recommender system based on user preferences mined from social networking data. Indoor POIs have also been used to build an indoor facility information and visualization system [26], annotators to denote user visits in urban areas [27], generating large scale maps [28] and in labeling objects and spaces in AR platforms [29] and navigation systems [30–32]. These applications, however, focused on utilizing Indoor POI as a marker for objects in indoor space, rather than differentiating its identity from POIs in the outdoors.

In literature, indoor navigation is one of the major uses for Indoor POIs, such as in determining best routes for a context-aware systems for navigation [17], ubiquitous indoor navigation [33], web-based navigators [18], WiFi-assisted path planning [34], and point planning for robotic navigations [35, 36]. It is also interesting to note that studies cite Indoor POI data as an environmentally crucial component, especially in cases where navigation is critical. Several studies use them as navigation guides in indoor wayfinding systems for visually impaired situation awareness [37–40] or those physically impaired [41] as path determinants or as indicators for hazards. Most importantly, Indoor POI integration with datasets based on international data standards, such as IndoorGML, is possible to more accurately portray and perform applications in indoor space [42].

Indoor POIs also play a key role in the indoors as landmarks, not only objects to denote the location of objects or spaces but also as guides for users to form mental spatial representations of their surroundings. Especially in situations where users are unfamiliar with the surroundings, these POIs aid route decision-making and orientation [43], as well as reference points that assist in recreating physical layouts of buildings [34]. These features are highlighted as important elements of the user line of sight and as background knowledge that moderates how users perceive the indoor environment [44].

Studies have also explored the localization of and using Indoor POIs. Since these objects are distinguishable from their surroundings, they are ideal for localization. Referred to as beacons, range-only SLAM (simultaneous localization and mapping) was able to identify positions of POIs using only distances with the assistance of the strength of radio frequency signals [45]. Similarly, WiFi signals which are readily available indoors plus radio FM (frequency modulation) signals permit indoor localization of POIs by similar principles [19]. Conversely, these POIs have assisted pedestrian dead-reckoning [46] and applications on indoor look-up services [47]. On a related note on the Indoor POIs as landmarks, it is possible to achieve adequate indoor localization using only these objects through a fingerprinting-based approach [48].

Various motivations have sought the extension of POI's nature as being an entity with a location and some attributes.

OGC defines POIs as “a location (with a fixed position) where one can find a place, product or service, typically identified by name rather than by address and characterized by type, which serves as a reference or a target in an LBS request, e.g., as the destination of a route” [7]. [4] pointed out that the term itself distinguishes between what is “interesting” and what is not, depending on the available context, but the OGC definition does not consider this subjective component [20].

A resolution to this gap is providing context in the usage. In 3D environments, especially indoors, apart from being key features that are essential for specific user tasks, these data also provide insight for visualization and navigation. In this case, however, Indoor POI visualization would have to deal with issues that are not usually present in 2D, such as occlusion, perspective, and scene complexity. An approach using a cloud of interest (COI), was proposed maximizing the information that the user is receiving without context distortion, too much cluttering, and additional cognitive task of looking through multiple views at the same time and still be suitable for small viewing spaces, hence bridging the challenge of creating models for mobile 3D geovirtual environments [4]. As 3D visualization of POI in the indoor environment representing both objects and spaces are emphasized [49], this emphasizes the need for a formalized data model that enables linking with other data.

On a related note, studies have shown that semantic models based on international standards such as CityGML and IFC (Industry Foundation Classes) have enriched thematic information of each other [50–52] or of other datasets like 3D mesh data used in solar potential analysis [53]. In contrast, while studies have also shown that application flexibility is enhanced utilizing semantically enriched POIs [54], this enrichment may come from the semantic models. In both outdoors [1] and indoors, POI data is essential for ontology-based recommender systems in different applications. Studies have used Indoor POIs in recommender systems utilizing shopping trajectories to model user behavior and preferences [25, 27]. Literature also cites that having an alias database management system would increase the efficiency of POI data, that is, obtaining the same level of richness of information even with a significantly smaller size of dataset [55].

Semantic hierarchies in the indoor environment have been demonstrated by [56–59] through describing how indoor spaces are related to its subunits and how this relationship plays a role in various aspects. The concept of space subdivision and aggregation enhances how space is perceived cognitively, particularly in as an essential in providing descriptive information on location, determining functional areas within indoor space, and determining which parts are suitable for navigation. The BOT (building topology ontology) is an effort to evolve existing IFC-specified standards towards Linked Data practices for modern web-based applications. Similarly, in the context of building and construction field, BOT presents spatial 3D volumes as zones that may contain other zones in a hierarchical manner—sites that may contain buildings, buildings that may contain building levels, and building levels that may contain spaces. Furthermore, this specification also defines tangible elements that either comprise or contained within these zones [60, 61].

The lack of support for temporal data for available models has also been raised [1, 20]. Data such as opening/closing times of establishments, the amount of time that people spent, or real-time data about services enrich the attribute content of POIs, which may be crucial to many applications. Some characteristics and even the location may change over time. From tracking datasets, novel query methods of two types—snapshot (on a given time point) or interval (over a given period)—have determined frequently visited Indoor POIs [62]. Keeping multiple versions of the Indoor POI has also been suggested to maintain information content [63]. To maximize this information in studying change and patterns, the data model must incorporate these.

For most cases, a name has been an identifier for an outdoor POI. However, this identifier would not mean that an Indoor POI corresponds to one and only one exact string of text. Users may vary in keyword use, and typographical errors are not impossible, so more than one keyword may exist, called an alias. This case is especially true for Indoor POI that may be referred to with similar characters, due to it not being identified by name. In a crowdsourcing-based collection method for Indoor POI, multiple names may refer to a single location [10]. To develop a system for managing aliases, [55] classified POI alias attributes and used word similarity measurements to input and retrain an alias database containing nonofficial names for POI.

Based on the developmental direction from POI towards Indoor POI data models, as well as primary usage domains for Indoor POI evident from previous studies, namely, searching, data labeling, and sharing, and identified areas of improvement and development, we propose the spatial-temporal indoor POI data model in the next section.

3. Proposed Spatial-Temporal Indoor POI Data Model

In Section 3, we consider the considerations for the spatial-temporal data model proposed in this study. We investigate the attributes of the Indoor POI in terms of its main usage and identify critical points that the data model must define.

3.1. Characteristics of Indoor POI. Indoor POIs may represent indoor spaces such as a room, corridor, lobby, or stairwells, as well as facilities, movable or immovable, located in those spaces such as furniture, installations, or equipment. Previous studies discussed in the preceding chapter have shown extensive use of Indoor POIs across a wide range of application domains. These features are present in navigation as either targets (e.g., what is the route from my current position going to Indoor POI 1?), guides (e.g., in calculating a route from point A to point B, avoid Indoor POI 1 and pass through Indoor POI 2), or both. In terms of localization, they have been vital in both finding positions of other objects (e.g., given the coordinates of visible Indoor POI 1, 2, and 3, what is the coordinates of the user), or the targets of localization using various measurements (e.g., given the WiFi signal strength from routers A, B, and C, what is the position of Indoor POI?). Indoor POIs are essential in 3D indoor visualization as landmarks to improve users' mental recognition of

their surroundings or even as merely labeling features to increase information content. Furthermore, these features provide rich content that enables spatial and temporal queries in LBS applications.

The Indoor POI data model proposed in this paper considers the same aspects as the previous data model [2], with particular consideration to the specific cases of indoor space, as compared to outdoor space. This proposed data model does not restrict a generic set of objects to be represented as Indoor POIs. Instead, any indoor facility (in the spatial range of a room), as well as the room containing these facilities (in a corridor's spatial range), may be represented. In this regard, inclusion relationships may also form between the former and the latter. Moreover, in the indoor environment, the difference in spatial range is more apparent, the presence of Indoor POIs in say, a corridor would have a difference in contrast to those present inside a room in different aspects. Moreover, facilities and other objects represented by Indoor POIs are more mobile, i.e., movable and can change location over time, and conversely, a location may have various Indoor POIs over a specified period.

Figure 1 improves upon the developmental direction of the outdoor POI proposed in [2]. The essential POI purposes of sharing, labeling, and searching remain to be the motivation of the development of the Indoor POI data model; however, specific key characteristics shared between these purposes are differentiated according to the importance and vital differences with the case of outdoors. Between the three primary purposes are the corresponding aspects of Indoor POI that are imperative in developing the data model. First, the data model's management of aliases level of detail is critical for purposes of search and labeling. These aspects are crucial in datasets to improve searching and managing the amount of information that the screen presents to the user, respectively. More importantly, information about spatial hierarchy, which corresponds to the spatial relationships of indoor space with either other indoor spaces or objects found indoors, is critical since this is more apparent in the case of indoors. In relation to Indoor POI sharing and searching functions, the aspects of maintaining a classification scheme and handling of multitemporal information are crucial driving points. Finally, as an identifier for places to their specific location, this aspect of Indoor POI is more crucial in its role in data sharing. These characteristics are elaborated in the sections to follow.

3.2. Indoor POI Nomenclature. Even if a feature does not have a unique name, Indoor POIs can still serve its purpose as an identifier, since it still connects an "indirect" geographic reference to a specific location. Intuitively, this classification aspect would more often be the more practical or in some cases the only existing nomenclature to identify a particular Indoor POI, as most objects found that indoors, despite being tangible objects, do not have a specific name. We can only refer to them through their generic names, such as a fire extinguisher, a CCTV, or an ATM. Providing a classification scheme for Indoor POIs would provide not only a uniform method of defining and differentiating features but also an opportunity for faster queries by narrowing down POI results depending on the purpose of the user. It also improves

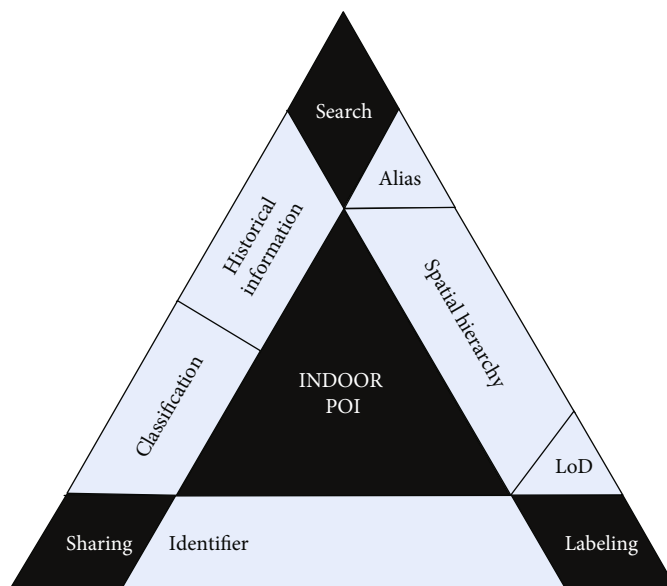


FIGURE 1: Developmental direction of the Indoor POI data model.

TABLE 1: Sample conceptual Indoor POI classification scheme.

Level 1	Level 2	Level 3
Place	Pedestrian	stairs, slope way, lobby
	Private	common room
	Relaxation	relaxation room, smoking area, men’s toilet, women’s toilet, disabled toilet, bench, rubbish bin
Things	Access Facility	door, ticket gate, access control units
	Conveyor transport	up escalator, down escalator, up moving sidewalk, down moving sidewalk, horizontal moving sidewalk, elevator, wheelchair lift
Retail & Services	Services	nursery, drug store, vending machine, ticket machine, information, lounge, guest room, covered car park, bank, cash office, cash machine or ATM, post office, billboard
	Retail	restaurant, coffee shop, clothing store, hair shop, ticket office, aquarium, bowling alley, swimming-pool, store
Safety & Security	Fire-fighting supplies	fire protection appliance, fire extinguishing system, fire extinguisher, fire alarm
	Emergency relief supplies	automated external defibrillator, life-saving trolley
	Evacuation facility	shelter, emergency call center, exit, emergency escape device, relief goods

efficiency in query-based implementations, as classifications would also enable grouping and subgrouping of similar features. In the data sharing aspect, classification provides an identifier for linking data from external sources, such as the code list from CityGML, for example.

Hence, a classification scheme to categorize Indoor POI is necessary to encourage utilization, increase query efficiency, and avoid duplications in datasets. A scheme also supports the Indoor POI functionalities of sharing and searching. In Table 1, we illustrate a sample of a classification of Indoor POIs, created based on the ESRI POI classification scheme [23], categorizing each feature in three levels of increasing specificity, and each category would correspond to a 6-character category code as an attribute for the Indoor POI.

This classification scheme does not intend to provide an exhaustive list of all possible types of objects and spaces but rather as an illustration of possible varieties of what an Indoor POI can represent. For example, “Vending Machine” and “Drug Store” may both be represented even though they are differentiated by [42] as non-navigable facilities and navigable facilities, respectively. The previous examples are actual objects located indoors, while the latter represents the spaces that may contain the former.

In developing LBS applications, and even though conventional models only require a name alongside a location to define a POI even outdoors, aliases are existing because of the nonuniformity of the keywords that users key-in for searching, and typographical errors are not impossible to occur. Even if an Indoor POI has an official identifier, be it

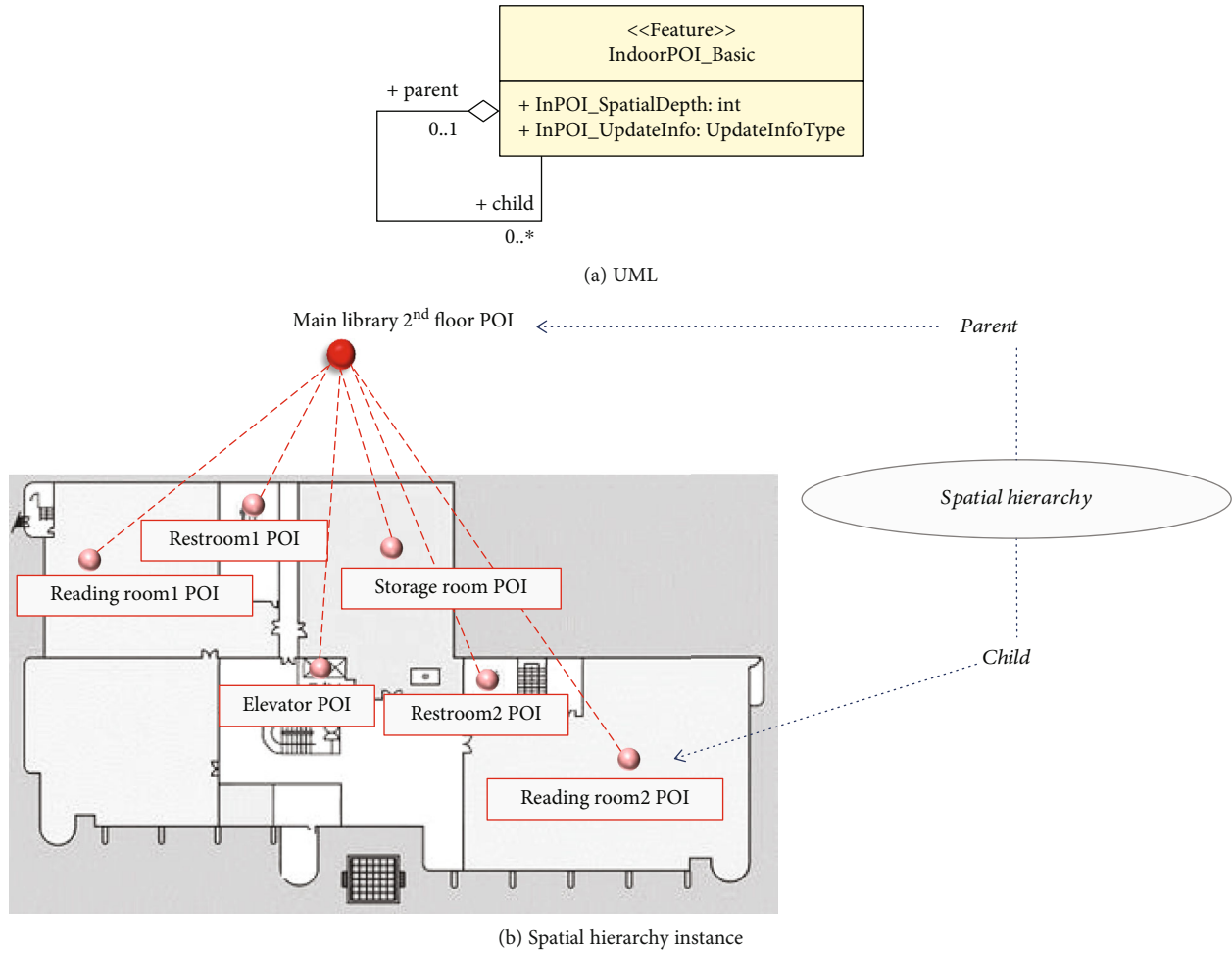


FIGURE 2: Self-aggregation for spatial hierarchy.

a name or its classification, the data model should be able to incorporate aliases. Having an alias database would improve searching while ensuring a practical and flexible, yet efficient delivery of information to users.

Managing descriptive information of the Indoor POI is also essential to maintain data integrity and quality. As data sharing is encouraged by a standardized data model, the author of the dataset must also be included to ensure efficient management, accountability, and facilitation of data reuse and updating. Similarly, successful LBS is possible if the Indoor POI can carry attributes apart from its name, classification, and location. Other descriptive information that may widely vary in data type, length, and value should be managed by the data model so rich information may be maintained and furnished to users.

3.3. Spatial Hierarchy and Spatial Relations. For Indoor POIs, spatial relationships may exist in two ways. First, a spatial hierarchical relationship may exist between an Indoor POI and another Indoor POI, as expressed in the previous research for outdoor POI [2], say, for example, an Indoor POI representing a floor level, and the Indoor POI representing the rooms in that level. This case represents the aggregation of smaller spatial units in one hierarchical level (the rooms, in

this case) towards a larger spatial unit in a higher hierarchical level (floor). Second, an inclusion relationship exists between Indoor POIs that represent space and those that represent objects located inside those spaces, say for an Indoor POI representing a library, and for Indoor POIs inside representing shelves. These relationships must be maintained in the data model to facilitate query analysis and extend into applications such as navigation, facilities management, or patrol services to fulfill its roles in searching and feature labeling properly. Also, this provides an opportunity for the improvement of data integration with other standards dealing with indoor spatial information such as IndoorGML.

We express these relationships in the data model as a self aggregation of the IndoorPOI_Basic class. Each POI instance has a 0~1 parent or a 0~ n child, as shown in Figure 2(a). This multiplicity specifies that an Indoor POI may not have a parent class, but if it does, it cannot have more than one parent having a higher spatial hierarchy. A child class for an Indoor-POI_Basic instance may not be present, but should it be, this instance may have one or more child classes having a lower spatial hierarchy. For example, a space-based hierarchical structure exists between a POI representing a floor level of a building, and the corresponding POIs representing the rooms and facilities existing within that floor level, as in Figure 2(b).

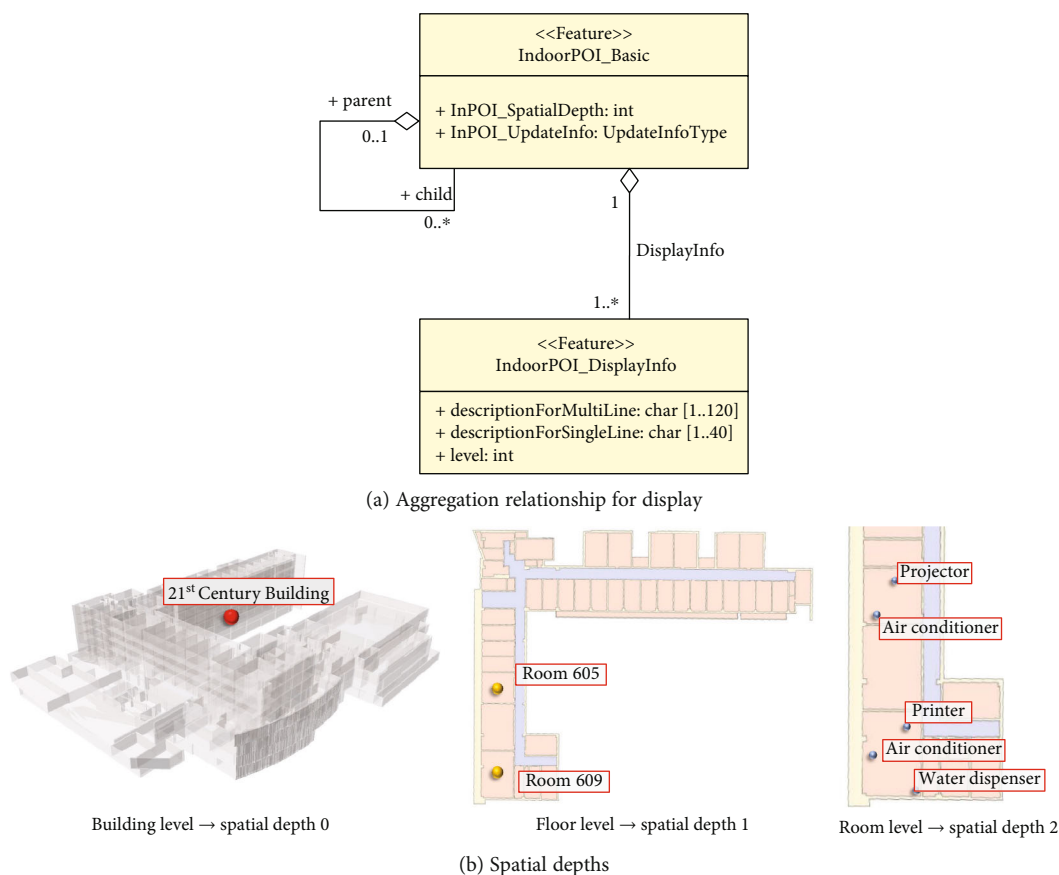


FIGURE 3: Spatial depth of indoor POI.

3.4. Spatial Depth. In the display of spatial data to users, the scale plays an important factor in how much information is visible and intelligible. As with any conventional or digital map, at varying spatial scales, Indoor POI must be expressed efficiently in a proper level of detail. Hence, a different set of Indoor POIs must be visible in larger scales compared to smaller scales. This aspect is crucial in order for it to achieve its role in feature labeling.

One of the main challenges in creating LBS platforms is screen size, due to the limitation of the devices where they run on [4]. As POIs and Indoor POIs are mainly geared towards providing LBS and as trends point toward more portable and handheld devices, the display restricts the relay of the richness of information to the users. Smaller screen size, as well as the size limitation in the indoor space themselves, command methods for efficient expression of Indoor POIs and their respective attributes.

Although closely related, spatial depth does not directly equate to spatial hierarchy. The self-aggregation for the latter refers to the relationship of an Indoor POI parent node containing another smaller spatial unit represented by an Indoor POI child node, for example, the cases of between a building (parent) and floors (child), and between the floors (parent) and rooms (child). Differently, this may also refer to the inclusion relationship between a space represented by an Indoor POI (room) and the objects inside the room (desks).

On the other hand, two Indoor POIs having the same spatial depth may belong in different hierarchical levels,

say, for example, Indoor POIs for a cinema lobby and a ticket machine. Both may be displayed at the same time (same spatial depth) even though the lobby has an inclusion relationship with the ticket machine (different spatial hierarchies). Indoor POIs at the same hierarchical levels may also belong in different display levels, for example, Indoor POIs representing a shelf and books. Both are at the same hierarchical levels below a room Indoor POI, but in an application, displaying all books might be illegible for display, unless a larger scale is visible.

POI display on the application is expressed for each scale level through the definition of spatial depth, through a user-defined `InPOI_SpatialDepth` attribute. Indoor POIs having the same integer value for this attribute would be displayed together at the same scale level. Additionally, an aggregation relationship, shown in Figure 3(a), between the `IndoorPOI_Basic` class and the `IndoorPOI_DisplayInfo` class having the expression level as the attribute, allows users to display Indoor POI descriptions in levels independent of the spatial depth, depending to the user's intent, or the importance of a POI in a particular context. It has a 1~ n child multiplicity, expressed as an Indoor POI on multiple levels, say an important facility such as an elevator used to transport between floor levels. Figure 3(b) shows an example of Indoor POIs having different spatial depths. An entire building is expressed as a single Indoor POI at spatial depth 0, while individual rooms may be expressed distinctly at Spatial Level 1. Further, Spatial Level 2 shows objects found inside the rooms at Spatial Level 2. Expanding to

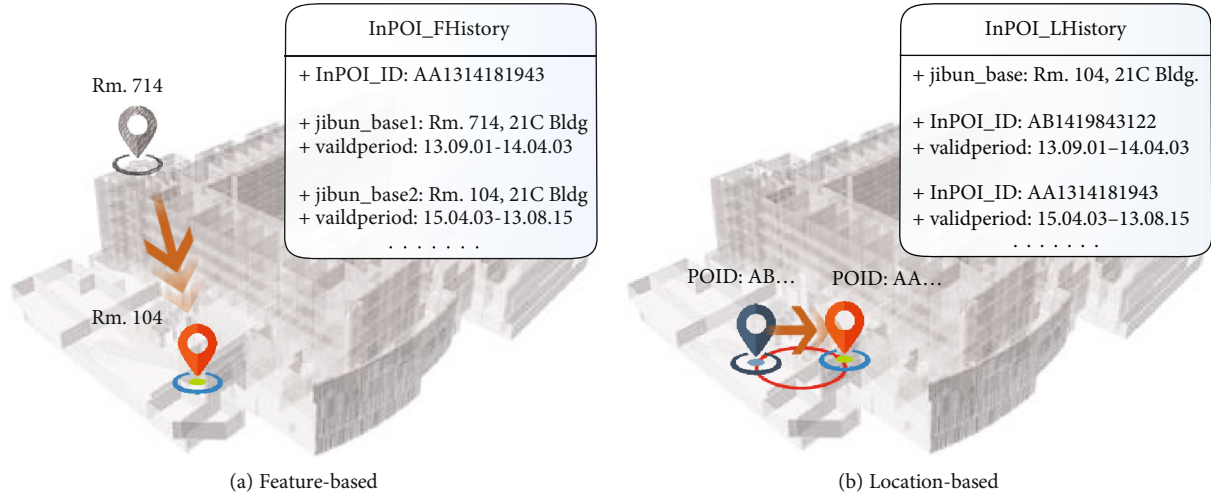


FIGURE 4: Two ways to describe Indoor POI temporal information.

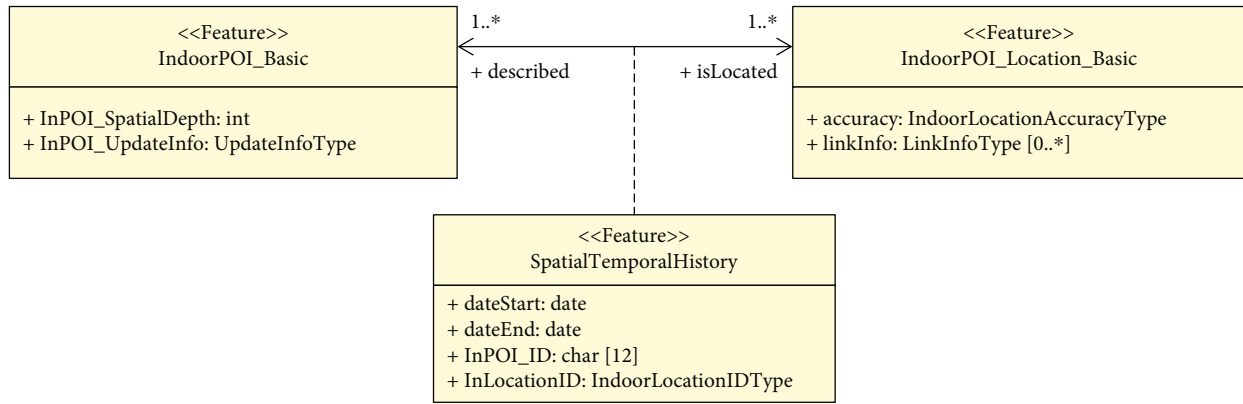


FIGURE 5: Association class for spatial-temporal Indoor POI history management.

further spatial depths is possible, if necessary, depending on the datasets and the application.

3.5. *Spatial-Temporal Information.* Despite being indoors, we cannot expect entities that are represented by Indoor POIs to be stationary, as most objects existing in those locations are mobile. For example, equipment transferred from one area of a room to another, or an entirely new location within the same building. Similarly, descriptive information (such as usage, schedule) regarding a space (such as a room) may be dynamic due to relocations, reconstructions, or maintenance. For instance, an ordinary classroom transformed into a computer laboratory, in the case of a school. These changes, that is, the change history of either the locations or the features themselves, may be permanent or temporary. Regardless, keeping track of this information must be maintained in the data model so Indoor POIs can improve its functions in information query or searching.

This proposed model introduces two ways of time series management for Indoor POI, feature-based and location-based. Feature-based management means management of the locational changes that a single Indoor POI undergoes

over periods, as shown in Figure 4(a). A certain POI object named “Student Lounge” in Room 714 of the 21st Century Building from 13 September 2001 to 14 April 2003 was moved to Room 104 of the same building from 15 April 2003 to 13 August 2015.

Time series management of Indoor POIs based on location means monitoring the Indoor POI located at that fixed position over periods, as shown in Figure 4(b). For example, on the location of Room 104 of the 21st Century Building, a specific Indoor POI for a Photocopy Room exists from, 13 September 2001 to 14 April 2003 but was changed to the Student Lounge starting 15 April 2003 to 13 August 2015. If applications manage temporal information like this, both management of feature changes and locational changes for Indoor POI and positions are possible, enabling historical search in various forms and implementations.

An association class SpatialTemporalHistory was added in the association relationship of IndoorPOI_Basic class and IndoorPOI_Location_Basic class to accommodate these time series management concepts to the Indoor POI data model. Figure 5 illustrates this relationship. In the generic data model from [2], these classes had a one-to-one

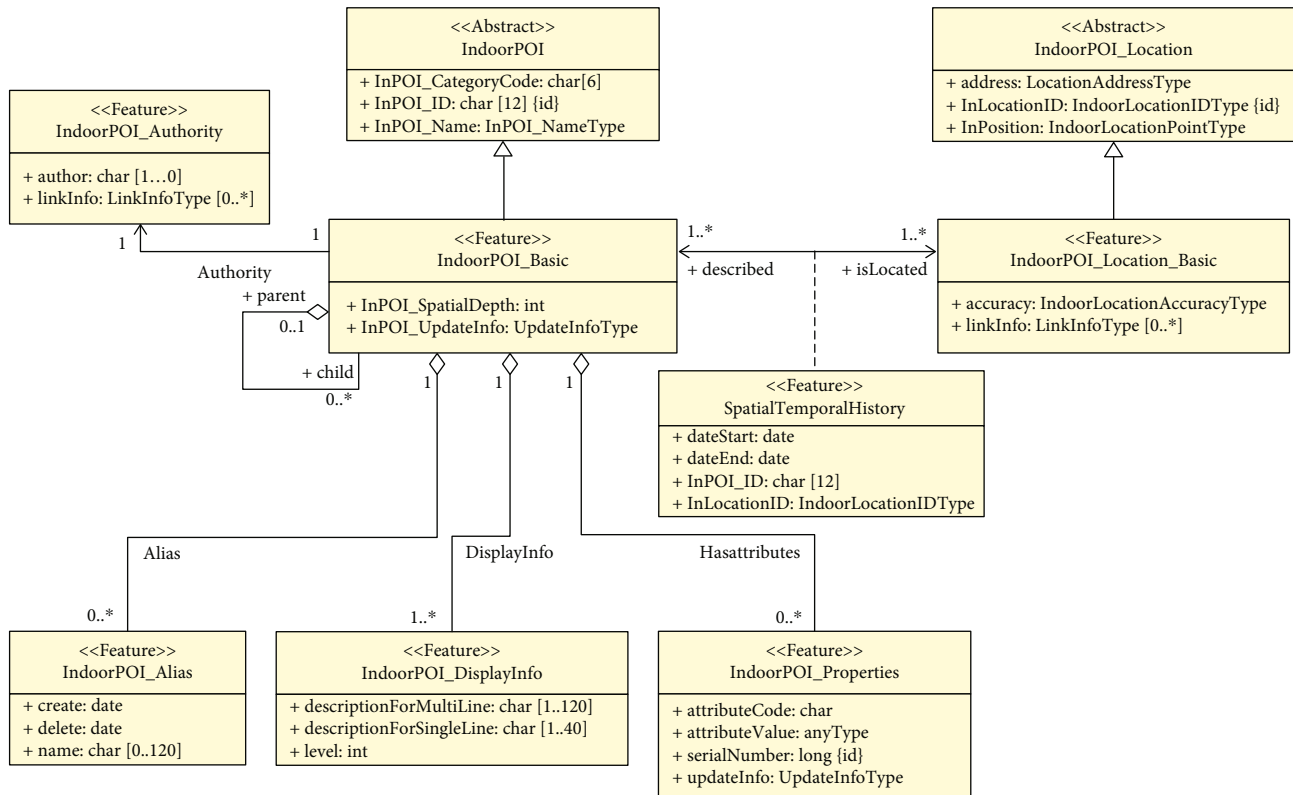


FIGURE 6: UML diagram of spatial-temporal Indoor POI data model.

association, but considering a feature-based time series management, the IndoorPOI_Basic now has a one-to-many association relationship with the IndoorPOI_Location_Basic class through the SpatialTemporalHistory Association class. On the other hand, a one-to-many association from IndoorPOI_Location_Basic class to the IndoorPOI_Basic class through the SpatialTemporalHistory class corresponds to the location-based management approach.

3.6. Spatial-Temporal Indoor POI Data Model. In this chapter, we discuss the structure of the spatial-temporal Indoor POI data model, from the considered characteristics in the previous section based on primary Indoor POI functions of searching, sharing, and labeling and is built upon the generic POI data model by [2]. The classes reflected in the data model represent the essential aspects of an Indoor POI based on the discussions in the preceding sections, as guided by previously proposed models for conventional POI models and considering the case of the indoors. We designed this data model to align implementations towards establishing actual Indoor POI data, while addressing current limitations of available data models.

The data model consists of 9 classes, namely, IndoorPOI and IndoorPOI_Location abstract classes, IndoorPOI_Basic class, IndoorPOI_Location_Basic class, IndoorPOI_Authority class, IndoorPOI_Alias class, IndoorPOI_Properties class, IndoorPOI_DisplayInfo class, and the SpatialTemporalHistory Association classes. The UML Diagram in Figure 6 describes each of these classes and their respective relationships.

The IndoorPOI_Basic class is the class that expresses the Indoor POI object, which may represent facilities or indoor spaces, which characterizes spatial hierarchy through its self-aggregation and multiple expressions across varying spatial depths through the one-to-many aggregation with child class IndoorPOI_Basic. The IndoorPOI_Basic class has a one-to-one relationship with the IndoorPOI_Location_Basic class, a one-to-many relationship with attribute information through the IndoorPOI_Properties class, and copyright information with its one-to-one association with the IndoorPOI_Authority class. To increase effectiveness and efficiency in managing the Indoor POI information, it implements alias management through the IndoorPOI_Alias class.

The IndoorPOI_Basic class and the IndoorPOI_Location_Basic classes inherit from the abstract classes IndoorPOI, and IndoorPOI_Location, respectively, thus obtaining their attributes for each instantiation. First, the IndoorPOI_Basic class has an ID, Name, CategoryCode, UpdateInfo, spatial depth, and child (parent) as attributes. The ID is a unique 12-digit character combining the information from the linked authorized agency and the serial number of the object. The attributes include a name, as an InPOI_NameType, official or not, that may include English, Korean, or alphanumeric characters. We reflect the classification discussed in Section 3.1 in the CategoryCode, UpdateInfo contains data on the creation and updating of other attributes, and the SpatialDepth for expressing display and hierarchical aspects.

The IndoorPOI_Location_Basic class inherits the IndoorPOI_Location abstract class. It has the unique ID, as an IndoorLocation_ID Type, position expressed a 3-dimensional

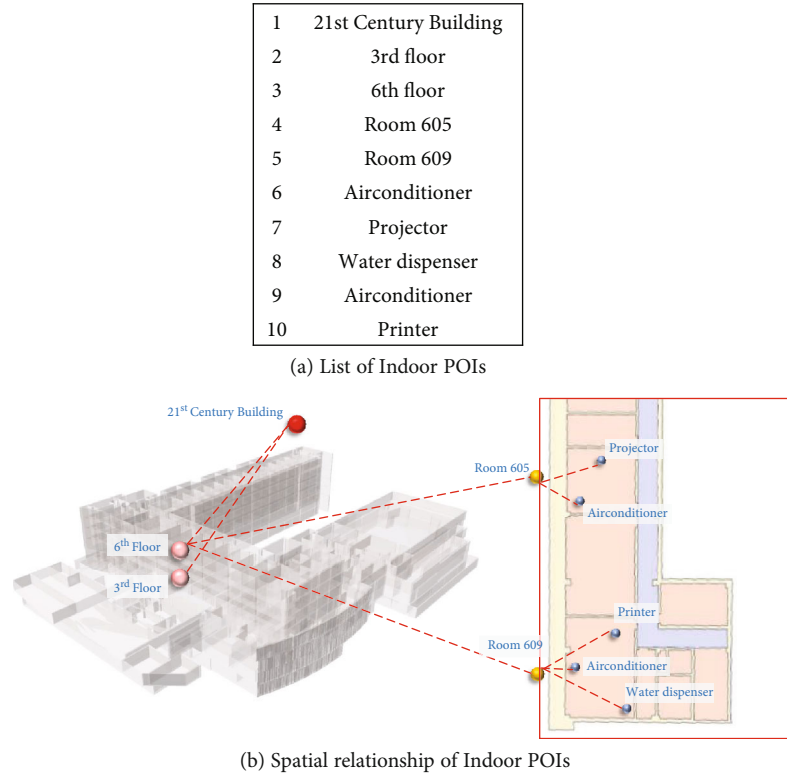


FIGURE 7: Sample Indoor POI data for Experiment.

IndoorLocationPointType, address as a LocationAddressType similar to the POI data model [2], accuracy information depending on location data collection method such as survey, grant of address position, and drawing, and the LinkInfo as a LinkInfoType for other linked location information.

IndoorPOI_Properties class has attributes of serialNumber, attributeCode, attributeValue, and UpdateInfo. The serialNumber is the serial number of the object, the attributeCode is a predefined attribute value according to the type, attributeValue may take any value in various types (anyType), and the UpdateInfo contains information on the update of the attribute.

The IndoorPOI_Alias class is the critical feature for alias management in Indoor POI, which helps significantly to manage the data efficiently. This class includes information on the alias name, as well as create and delete, which corresponds to the alias creation and deletion dates from the alias database, respectively. The IndoorPOI_Basic class is for how the POI is displayed visually to the user, according to user specifications. It contains attributes on level of detail (LOD) in level, supporting multiline or single-line descriptions in descriptionForMultiLine and descriptionForSingleLine. Finally, the IndoorPOI_Authority class expresses the author of the Indoor POI, which is responsible for the information in the feature, indicating the attribute author for the name of the author and LinkInfo for the corresponding agency or organization where the author is affiliated.

The association class SpatialTemporalHistory described in the preceding section represents the time series management for both feature-based and location-based management in the data model, having association relationships with the IndoorPOI_Basic and IndoorPOI_Location_Basic class. This

class contains attributes dateStart and dateEnd to describe the period and the respective linked InPOI_ID and InLocationID. This period denotes the validity of the existence of the Indoor POI in a particular location, whether it is a feature-based or location-based approach. It also neither refers to the actual creation and deletion dates of the POI in the dataset nor the existence or removal of the real-world feature or space it represents.

We incorporate these measures to the Indoor POI data model, resulting in the spatial-temporal Indoor POI data model shown in Figure 6. We add the SpatialTemporalHistory association class and the multiplicity of the association relationships between IndoorPOI_Basic and IndoorPOI_Location_Basic classes. This approach presents a more compact approach to the one applied to the spatial-temporal POI data model presented in [2], which used four more classes—separate ones for the location and history information for each of the two time series management methods.

4. Experimental Implementation

The described Indoor POI data model describes spatial relationships between Indoor POI entities having a hierarchical structure and considers time series management in mind. To demonstrate this, we conduct an experiment considering a use case for the management of facilities located in the interior of a building. We do this for sample Indoor POI objects through spatial hierarchy and historical attribute query in this section using a sample set of 10 Indoor POI objects listed in Figure 7(a). For simplicity, these points represent selected locations in the 21st Century Building of the University of

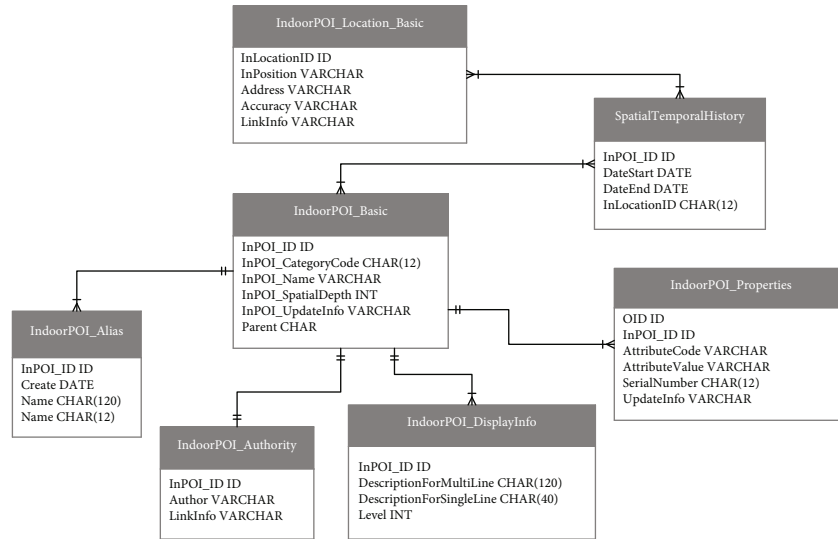
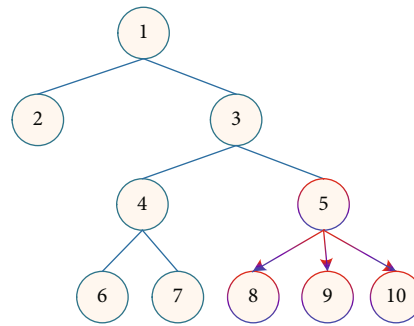


FIGURE 8: Database schema for the experimental implementation.



(a) Illustration of Indoor POI hierarchy

```

SELECT e1."parent", e2."InPOI_Name", e1."InPOI_Name", e1."InPOI_ID"
FROM "IndoorPOL_Basic" as e1
INNER JOIN "IndoorPOL_Basic" as e2
ON e1."parent" = e2."InPOI_ID"
WHERE e2."InPOI_Name" = 'Room 609'
    
```

parent	InPOI_Name	InPOI_Name	InPOI_ID
character(12)	character varying	character varying	character(12)
1	0000000000005	Room 609	Water Dispenser 0000000000008
2	0000000000005	Room 609	Airconditioner 0000000000009
3	0000000000005	Room 609	Printer 0000000000010

(b) Spatial hierarchy Indoor POI query

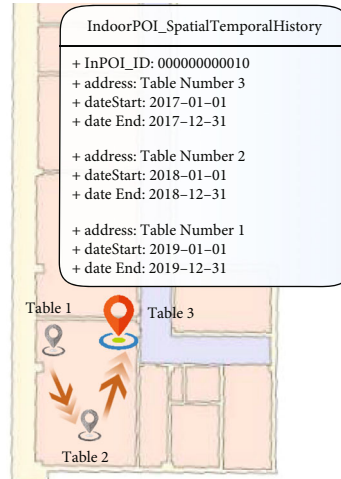
FIGURE 9: Experimental implementation of indoor POI spatial hierarchy.

Seoul campus. At the topmost spatial depth, an Indoor POI representing the whole building exists, followed by an Indoor POI for the 3rd- and 6th-floor levels, respectively, of the whole building. Within the 6th-floor level, in the next spatial depth are Indoor POIs representing two rooms on that floor, and the final spatial depth contains objects each contained respective rooms, as illustrated in Figure 7(b).

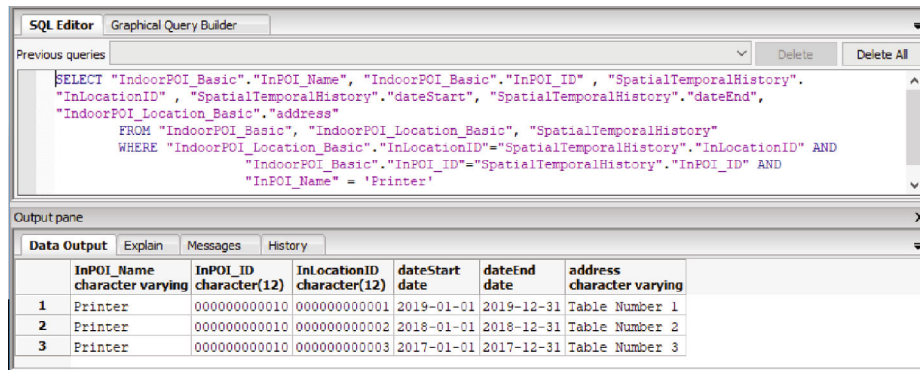
To evaluate the potential benefits of the data model, we show to demonstrate the key characteristics of Indoor POI by constructing a relational database and implementing SQL-based queries. For simplicity, we mapped each concrete class as a single table in the database schema to more clearly

see how each UML class works. This mapping is also an ideal strategy since class hierarchies in the model are shallow. Since the classes specify IDs explicitly, we were able to use these as keys in order to map the respective relationships directly. We entered the sample data as features in PostgreSQL, a free and open-source relational database management system, through the devised database schema shown in Figure 8, based on the UML model Figure 6.

For instance, the facility manager would like to know which facilities are present inside a room. To do this, we attempt to search Indoor POI that exhibits the self-aggregation relationship for expressing spatial hierarchy. Figure 9(a) illustrates



(a) History of Indoor POI location



(b) Feature-based spatial-temporal Indoor POI query

FIGURE 10: Experimental implementation of feature-based spatial-temporal POI.

the hierarchical relationships among Indoor POIs based on information illustrated in Figure 7. We perform an SQL query on the IndoorPOI_Basic class to identify child nodes lying within a node (in this case, a parent), as shown in Figure 9(b). The result of the query for Indoor POI for “Room 609” as a parent node shows a list of all Indoor POIs inside that entity, i.e., its child nodes, namely, “Water Dispenser”, “Air Conditioner”, and “Printer”.

On the other hand, time series management measures of the proposed data model enable search of historical attributes of both features and locations having this information, say the facility manager wants to identify locations where a facility presently and previously exists, so to demonstrate a feature-based history such as a facility being moved in different locations on different times, as shown in Figure 10(a). Figure 10(b) shows the result of the location history of the “Printer” Indoor POI from the SQL query on the Spatial-TemporalHistory attribute, enumerating the various locations that it has existed at and each corresponding period.

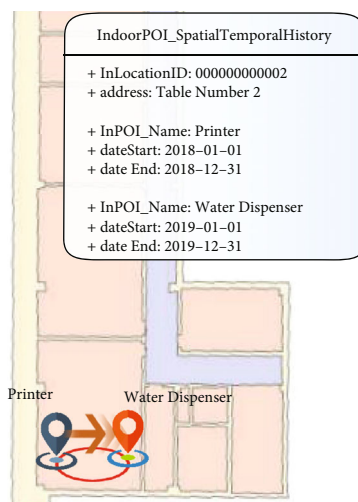
Similarly, historical attribute search also enables management of the POI history of a particular location, if the facility manager wants to know which locations a facility has been used and transferred to across time, requiring the data model to handle location-based management, such as in Figure 11(a). Figure 11(b) shows the result of the POI history of a location

names “Table Number 2”, having two Indoor POIs existing in history, namely, “Printer” and “Water Dispenser”, as well as their corresponding validity periods. Experimental results from this section show that the data model can express spatial and temporal information management through hierarchical and historical queries, respectively.

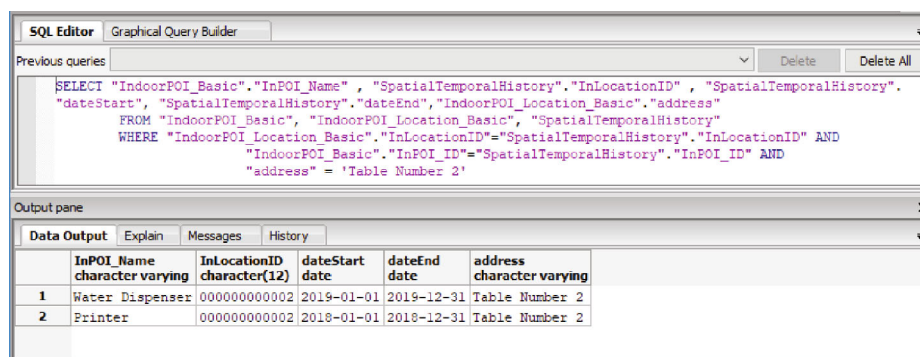
5. Conclusions and Future Studies

POI is an essential element in providing LBS across a wide array of application domains. While there have been numerous efforts and studies regarding its expansion, utilization, and standardization, there is still a limited outlook on how this concept extends towards the indoor environment. Considering important characteristics of Indoor POI in several aspects, in contrast to conventional POIs used in outdoor space, there is a need to specify a data model to assure data quality, provide means of validation, and enable analysis.

This paper proposes an Indoor POI data model considering various spatial aspects and time series management. Based on the three roles of POI, searching, sharing, and labeling, we improved upon a previous generic POI data model, to formulate the spatial-temporal Indoor POI data model with components for temporal information management. This data model enables alias management, expression of spatial hierarchy,



(a) Feature history on a location



(b) Location-based spatial-temporal Indoor POI query

FIGURE 11: Experimental implementation of location-based spatial-temporal POI.

display across various levels, uniform categorization, unique type identification, and management of historical information for efficiency, user-friendly and directed display, and cross-platform and cross-application sharing. This data model enables the extension of creating POIs to POI Level 2—for unnamed facilities such as CCTVs, restrooms, or other features or spaces that exist indoors. In addition, this paper has demonstrated the data model's support for hierarchical and historical information management through querying.

There are some limitations of this paper that the authors would like to address in future studies. First, this data model presents only the data model of the Indoor POIs themselves. Its implementation in more LBS-specific cases and applications, such as user location-based queries or pathfinding, would necessitate methodologies for integration with other datasets, for instance, IndoorGML in cases of use cases in indoor navigation platforms. Furthermore, avenues for visualization, such as with omnidirectional images, which have been demonstrated in the literature to provide effective and efficient visualization indoors, may be explored. We demonstrated that the use of the data model using a sample dataset, as no extensive yet compatible Indoor POI dataset, is obtainable. Should it be made available, investigations on the storage efficiency and more detailed comparisons with the traditional POI model are possible. This small size of

the sample dataset, although successfully demonstrating the features of the proposed model, enabled us to perform a simple approach in designing the database schema. Hence, the MDA (model-driven architecture) approach may be suitable for more massive datasets aimed for more sophisticated applications. Finally, since Indoor POIs deal with the second level of POIs, extending into the third level, i.e., intangible assets or historical levels, should be incorporated.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by a grant (20NSIP-B135746-04) from the National Spatial Information Research Program (NSIP) funded by Ministry of Land, Infrastructure and Transport of Korean government.

References

- [1] O. Ozdikis, F. Orhan, and F. Danismaz, "Ontology-based recommendation for points of interest retrieved from multiple data sources," in *Proceedings of the International Workshop on Semantic Web Information Management - SWIM '11*, pp. 1–6, Athens, Greece, June 2011.
- [2] J. Park, H. Y. Kang, and J. Lee, "A spatial-temporal POI data model for implementing location-based services," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 34, no. 6, pp. 609–618, 2016.
- [3] A. Heikkinen, A. Okkonen, A. Karhu, and T. Koskela, "A distributed POI data model based on the entity-component approach," in *2014 IEEE Symposium on Computers and Communications (ISCC)*, Funchal, Portugal, June 2014.
- [4] M. Trapp, L. Schneider, N. Holz, and J. Döllner, "Strategies for visualizing points-of-interest of 3D virtual environments on mobile devices," in *Proceedings of the International Symposium on Location Based Services & TeleCartography*, pp. 1–14, Potsdam, Germany, 2009, http://www.hpi.uni-potsdam.de/fileadmin/hpi/FG_Doellner/publications/2009/TSHD09/MobilePOI.pdf.
- [5] M. Kim and J. Lee, "Developing a method to generate IndoorGML data from the omni-directional image," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-2/W4, no. 2W4, pp. 17–19, 2015.
- [6] K. Chen, Y. K. Lai, and S. M. Hu, "3D indoor scene modeling from RGB-D data: a survey," *Computational Visual Media*, vol. 1, no. 4, pp. 267–278, 2015.
- [7] OGC (Open Geospatial Consortium), "OGC Glossary of Terms," <http://www.opengeospatial.org/ogc/glossary/> (accessed May 15, 2019).
- [8] H. Jung and J. Lee, "Development of an omnidirectional-image-based data model through extending the IndoorGML concept to an indoor patrol service," *Journal of Sensors*, vol. 2017, Article ID 5379106, 14 pages, 2017.
- [9] OGC (Open Geospatial Consortium), "IndoorGML v.1.0.3," 2018, <http://docs.opengeospatial.org/is/14-005r4/14-005r4.html>.
- [10] H. Shin, Y. Chon, Y. Kim, and H. Cha, "A participatory service platform for indoor location-based services," *IEEE Pervasive Computing*, vol. 14, no. 1, pp. 62–69, 2015.
- [11] R. H. Jan and Y. R. Lee, "An indoor geolocation system for wireless LANs," in *2003 International Conference on Parallel Processing Workshops, 2003. Proceedings*, pp. 29–34, Kaohsiung, Taiwan, October 2003.
- [12] M. Werner, M. Kessel, and C. Marouane, "Indoor positioning using smartphone camera," in *2011 International Conference on Indoor Positioning and Indoor Navigation*, pp. 1–6, Guimarães, Portugal, September 2011.
- [13] S. Gao and S. Prasad, "Employing spatial analysis in indoor positioning and tracking using Wi-Fi access points," in *ISA '16: Proceedings of the Eighth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*, pp. 27–34, Burlingame, CA, USA, October 2016.
- [14] J. Lee, M. Seo, J. Kim, S. Hwang, T. Kim, and K. S. Kim, "Management of subdivided dynamic indoor environments by autonomous scanning system10.1109/AIKE.2018.00051," in *2018 IEEE First International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pp. 224–227, Laguna Hills, CA, USA, September 2018.
- [15] A. Jamali, A. Abdul Rahman, P. Boguslawski, P. Kumar, and C. M. Gold, "An automated 3D modeling of topological indoor navigation network," *GeoJournal*, vol. 82, no. 1, pp. 157–170, 2017.
- [16] S. Zlatanova, G. Sithole, M. Nakagawa, and Q. Zhu, "Problems in indoor mapping and modelling," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 40, no. 4W4, pp. 63–68, 2013.
- [17] M. Werner, "Selection and ordering of points-of-interest in large-scale indoor navigation systems," in *2011 IEEE 35th Annual Computer Software and Applications Conference*, pp. 504–509, Munich, Germany, July 2011.
- [18] D. Zeinalipour-Yazti and C. Laoudias, "The anatomy of the anyplace indoor navigation service," *SIGSPATIAL Special*, vol. 9, no. 2, pp. 3–10, 2017.
- [19] Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha, "Indoor localization using FM signals," *IEEE Transactions on Mobile Computing*, vol. 12, no. 8, pp. 1502–1517, 2013.
- [20] T. Spangenberg, "Standardization, modeling and implementation of points of interest – a touristic perspective," *International Journal of u- and e- Service, Science and Technology*, vol. 6, no. 6, pp. 59–70, 2013.
- [21] ISO, "ISO 19112:2003 (Geographic information: spatial referencing by geographic identifiers)," International Organization for Standardization, 2003, <http://www.iso.org>.
- [22] ISO, "ISO 19155:2017 (Geographic information: geographic information – place identifier (PI) architecture)," International Organization for Standardization, 2018, <http://www.iso.org>.
- [23] ESRI, *Points of Interest and Categories*, 2019, September 2019, <https://pro.arcgis.com/en/pro-app/help/data/indoors/points-of-interest-and-categories.htm>.
- [24] J. Park and J. Lee, "Establishing required LOD and positioning accuracy for indoor spatial information applications in public administrative works," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 35, no. 2, pp. 103–112, 2017.
- [25] Z. Zheng, Y. Chen, S. Chen, L. Sun, and D. Chen, "Location-aware POI recommendation for indoor space by exploiting WiFi logs," *Mobile Information Systems*, vol. 2017, Article ID 9601404, 16 pages, 2017.
- [26] K. Kim and K. Lee, "Handling points of interest (POIs) on a mobile web map service linked to indoor geospatial objects: a case study," *ISPRS International Journal of Geo-Information*, vol. 7, no. 6, p. 216, 2018.
- [27] V. De Graaff, R. A. De By, and M. De Keulen, "Automated semantic trajectory annotation with indoor point-of-interest visits in urban areas," in *Proceedings of the 31st Annual ACM Symposium on Applied Computing - SAC '16*, pp. 552–559, Pisa, Italy, April 2016.
- [28] R. Martin-Brualla, Y. He, B. C. Russell, and S. M. Seitz, "The 3D jigsaw puzzle: mapping large indoor spaces," in *Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*, vol. 8691, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., pp. 1–16, Springer, Cham, 2014.
- [29] B. Al Delail, L. Weruaga, and M. J. Zemerly, "CAViAR: context aware visual indoor augmented reality for a University Campus," in *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, pp. 286–290, Macau, China, December 2012.

- [30] M. Ruta, F. Scioscia, S. Ieva, D. De Filippis, and E. Di Sciascio, "Indoor/outdoor mobile navigation via knowledge-based POI discovery in augmented reality," in *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pp. 26–30, Singapore, Singapore, December 2015.
- [31] P. Pombinho, A. P. Afonso, and M. B. Carmo, "Point of interest awareness using indoor positioning with a mobile phone," in *Proceedings of the 1st International Conference on Pervasive and Embedded Computing and Communication Systems - Volume 1: PECCS*, pp. 5–14, Vilamoura, Algarve, Portugal, 2011.
- [32] M. Stroila, A. Yalcin, J. Mays, and N. Alwar, "Route visualization in indoor panoramic imagery with open area maps," in *2012 IEEE International Conference on Multimedia and Expo Workshops*, pp. 499–504, Melbourne, VIC, Australia, July 2012.
- [33] B. Grefßmann, H. Klimek, and V. Turau, "Towards ubiquitous indoor location based services and indoor navigation," in *2010 7th Workshop on Positioning, Navigation and Communication*, pp. 107–112, Dresden, Germany, March 2010.
- [34] G. Shen, Z. Chen, P. Zhang, T. Moscibroda, and Y. Zhang, "Walkie-Markie: indoor pathway mapping made easy," in *Proceedings of the 10th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2013*, pp. 85–98, Lombard, IL, USA, April 2013.
- [35] T. Edlinger and E. von Puttkamer, "Exploration of an indoor-environment by an autonomous mobile robot," *IEEE/RSJ/GI International Conference on Intelligent Robots and Systems*, vol. 2, pp. 1278–1284, 1994.
- [36] J. B. Hayet, F. Lerasle, and M. Devy, "A visual landmark framework for indoor mobile robot navigation," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 4, pp. 3942–3947, 2002.
- [37] H. Zhang and C. Ye, "An indoor wayfinding system based on geometric features aided graph SLAM for the visually impaired," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 9, pp. 1592–1604, 2017.
- [38] S. A. Cheraghi, V. Namboodiri, and L. Walker, "GuideBeacon: beacon-based indoor wayfinding for the blind, visually impaired, and disoriented," in *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 121–130, Kona, HI, USA, March 2017.
- [39] G. E. Legge, P. J. Beckmann, B. S. Tjan et al., "Indoor navigation by people with visual impairment using a digital sign system," *PLoS One*, vol. 8, no. 10, article e76783, 2013.
- [40] H. Wu, A. Marshall, and W. Yu, "Path planning and following algorithms in an indoor navigation model for visually impaired," in *Second International Conference on Internet Monitoring and Protection (ICIMP 2007)*, p. 38, San Jose, CA, USA, July 2007.
- [41] L. C. De Oliveira, A. O. Andrade, E. C. De Oliveira, A. Soares, A. Cardoso, and E. Lamounier, "Indoor navigation with mobile augmented reality and beacon technology for wheelchair users," in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 37–40, Orlando, FL, USA, February 2017.
- [42] A. R. Claridades, I. Park, and J. Lee, "Integrating IndoorGML and indoor POI data for navigation applications in indoor space," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 37, no. 5, pp. 359–366, 2019.
- [43] P. Viaene, A. Vanclooster, K. Ooms, and P. De Maeyer, "Thinking aloud in search of landmark characteristics in an indoor environment," in *2014 Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS)*, pp. 103–110, Corpus Christ, TX, USA, November 2014.
- [44] J. Frankenstein, S. Brüssow, F. Ruzzoli, and C. Hölscher, "The language of landmarks: the role of background knowledge in indoor wayfinding," *Cognitive Processing*, vol. 13, no. S1, pp. 165–170, 2012.
- [45] Y. R. Esteves, J. B. Concejero, and A. V. Jiménez, "Indoor localization of the points of interest using RO-SLAM," in *Proceedings of the 6th International Conference on Data Communication Networking - Volume 1: DCNET*, pp. 35–42, Colmar, Alsace, France, 2015.
- [46] F. T. Alaoui, V. Renaudin, and D. Betaille, "Points of interest detection for map-aided PDR in combined outdoor-indoor spaces," in *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8, Sapporo, Japan, September 2017.
- [47] P. Zhuang, D. Wang, and Y. Shang, "SMART: simultaneous indoor localization and map construction using smartphones," in *Proceedings of the International Joint Conference on Neural Networks*, pp. 1–8, Barcelona, Spain, July 2010.
- [48] O. Willems, *Exploring a pure landmark-based approach for indoor localisation*, Delft University of Technology, 2017.
- [49] B. Al Delail, L. Weruaga, M. J. Zemerly, and J. W. P. Ng, "Indoor localization and navigation using smartphones augmented reality and inertial tracking," in *Proceedings of the IEEE International Conference on Electronics, Circuits, and Systems*, pp. 929–932, Abu Dhabi, United Arab Emirates, December 2013.
- [50] S. Donkers, H. Ledoux, J. Zhao, and J. Stoter, "Automatic conversion of IFC datasets to geometrically and semantically correct CityGML LOD3 buildings," *Transactions in GIS*, vol. 20, no. 4, pp. 547–569, 2016.
- [51] S. Kardinal Jusuf, B. Mousseau, G. Godfroid, and V. Soh Jin Hui, "Integrated modeling of CityGML and IFC for city/neighborhood development for urban microclimates analysis," *Energy Procedia*, vol. 122, pp. 145–150, 2017.
- [52] M.-S. Tryfona, *Bidirectional enrichment of CityGML and multi-view stereo mesh models*, Delft University of Technology, 2017.
- [53] B. Willenborg, M. Piltz, and T. H. Kolbe, "Integration of semantic 3D city models and 3D mesh models for accuracy improvements of solar potential analyses," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, no. 4/W10, pp. 223–230, 2018.
- [54] H. Pundt, "Semantically enriched POI as ontological foundation for web-based and mobile spatial applications," in *Universal Ontology of Geographic Space: Semantic Enrichment for Spatial Data*, vol. 2014, pp. 186–206, IGI Global, 2012.
- [55] J. Choi and J. Lee, "Developing an alias management method based on word similarity measurement for POI application," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 37, no. 2, pp. 81–89, 2019.
- [56] G. Sithole and S. Zlatanova, "Position, location, place and area: an indoor perspective," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III–4, pp. 89–96, 2016.
- [57] H. Jung and J. Lee, "Indoor subsampling to implement IndoorGML for indoor navigation," *ISPRS - International*

Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 40, no. 2W4, pp. 25–27, 2015.

- [58] S. Zlatanova, L. Liu, and G. Sithole, “A conceptual framework of space subdivision for indoor navigation,” in *ISA '13: Proceedings of the Fifth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*, pp. 37–41, Orlando, FL, USA, November 2013.
- [59] M. Kruminaitė and S. Zlatanova, “Indoor space subdivision for indoor navigation,” in *ISA '14: Proceedings of the Sixth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*, pp. 25–31, Dallas/Fort Worth, TX, USA, November 2014.
- [60] WWWC (W3C), *Building Topology Ontology*, 2020, <https://w3c-lbd-cg.github.io/bot/>.
- [61] M. H. Rasmussen, M. Lefrançois, G. F. Schneider, and P. Pauwels, “BOT: the building topology ontology of the W3C linked building data group,” *Semantic Web Journal*, 2019, <http://www.semantic-web-journal.net/content/bot-building-topology-ontology-w3c-linked-building-data-group-0>.
- [62] H. Lu, C. Guo, B. Yang, and C. S. Jensen, “Finding frequently visited indoor POIs using symbolic indoor tracking data,” in *19th International Conference on Extending Database Technology (EDBT)*, pp. 449–460, Bordeaux, France, March 2016.
- [63] K. Georgiou, T. Constambeys, C. Laoudias, L. Petrou, G. Chatzimilioudis, and D. Zeinalipour-Yazti, “Anyplace: a crowdsourced indoor information service,” *Proceedings - IEEE International Conference on Mobile Data Management*, vol. 1, pp. 291–294, 2015.

Review Article

Priority Analysis of Remote Sensing and Geospatial Information Techniques to Water-Related Disaster Damage Reduction for Inter-Korean Cooperation

Sunmin Lee,^{1,2} Sung-Hwan Park,² Moun-Jin Lee ,¹ and Taejung Song³

¹Korea Environment Institute (KEI), Republic of Korea

²University of Seoul, Republic of Korea

³Korea Information Technology Valley, Republic of Korea

Correspondence should be addressed to Moun-Jin Lee; leemj@kei.re.kr

Received 26 March 2020; Revised 23 May 2020; Accepted 2 July 2020; Published 17 July 2020

Academic Editor: Lei Zhang

Copyright © 2020 Sunmin Lee et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The social and economic harm to North Korea caused by water-related disasters is increasing with the increase in the disasters worldwide. Despite the improvement of inter-Korean relations in recent years, the issue of water-related disasters, which can directly affect the lives of people, has not been discussed. With consideration of inter-Korean relations, a government-wide technical plan should be established to reduce the damage caused by water-related disasters. Therefore, the purpose of this study was to identify remote sensing and GIS techniques that could be useful in reducing the damage caused by water-related disasters while considering inter-Korean relations and the disasters that occur in North Korea. To this end, based on the definitions of disasters in South and North Korea, water-related disasters that occurred during a 17-year period from 2001 to 2017 in North Korea were first summarized and reclassified into six types: typhoons, downpours, floods, landslides, heavy snowfalls, and droughts. In addition, remote sensing- and GIS-based techniques in South Korea that could be applied to water-related disasters in North Korea were investigated and reclassified according to applicability to the six disaster types. The results showed that remote sensing and other monitoring techniques using spatial information, GIS-based database construction, and integrated water-related disaster management have high priorities. Especially, the use of radar images, such as C band images, has proven essential. Moreover, case studies were analyzed within remote sensing- and GIS-based techniques that could be applicable to the water-related disasters that occur frequently in North Korea. Water disaster satellites with high-resolution C band synthetic aperture radar are scheduled to be launched by South Korea. These results provide basic data to support techniques and establish countermeasures to reduce the damage from water-related disasters in North Korea in the medium to long term.

1. Introduction

After the first Inter-Korean Summit that took place on April 27, 2018, additional summits were held to discuss inter-Korean issues. These summits mainly discussed economic and cultural exchanges. However, natural disasters are important because North and South Korea are geographically close and can affect each other. Furthermore, for the first time, North Korea–United States summits were held in Singapore on June 12, 2018, and February 27–28, 2019, and issues related to North Korea were discussed internationally. During discussions at the summit, the United States and

North Korea agreed to resolve a variety of issues, including complete denuclearization, a peace regime, normalization of United States–North Korea relations, and the repatriation of a corpse from the Korean War. However, as with the inter-Korean summit, no issues pertaining to North Korea's safety from natural disasters have been discussed, despite the severe current situation with regard to natural disasters in North Korea.

Water-related disasters are increasing in North Korea as unusual climate problems emerge worldwide. Correspondingly, disaster safety problems are receiving more attention [1]. According to the 2018 Global Risk Index, North Korea

has the 41st highest risk index among 191 countries [2]. In comparison, South Korea ranked relatively low at 166th. For geographical reasons, the occurrence of disasters at the inter-Korean border directly affects the entire Korean Peninsula, especially South Korea, because the water systems flow into South Korea. For example, in September 2009, six South Koreans were killed by the discharge of the Hwanggang Dam upstream on the Imjin River. Therefore, to reduce the damage caused by water-related disasters in North Korea, it is necessary to develop technical measures at the national level.

In this study, the frequency of water-related disasters in North Korea to identify applicable techniques was analyzed to ensure safety through inter-Korean cooperation. This plan reflects recent changes in North Korea's diplomatic relations in terms of humanitarian issues, which may foster responses to water-related disasters. Based on statistics regarding water-related disasters in North Korea, this study investigated the applicability of science and technology employed in South Korea, which has similar geographic characteristics, and established priorities for techniques that could support inter-Korean cooperation. To this end, the trend of water-related disasters in North Korea is first summarized and analyzed. Then, forms of science and technology used to respond to water-related disasters in South Korea are surveyed, and a ranking of the types of science and technology that would be applicable to North Korea is determined. Finally, case studies of remote sensing data of water-related disasters are analyzed. As a result of these processes, science and technology could be employed to reduce the damage caused by the disasters.

Due to the secretive nature of North Korea, data on the damage caused by disasters are difficult to obtain. Therefore, this study summarized data on the occurrence of water-related disasters in North Korea based on information released by the international community and data published by the North Korean government during a 17-year period from 2001 to 2017. The water-related disasters were categorized into six types based on the definitions of disasters in North and South Korea and the current status in North Korea. Through this process, water-related disaster-related techniques used in South Korea were investigated and classified into whether they could reduce damage from disasters occurring in North Korea. South Korean techniques for water-related disaster damage reduction could be applied to adjacent areas of North Korea. Against this backdrop, the use of remote sensing technique is essential. In addition, the priorities among South Korean techniques were evaluated by considering the water-related disasters that occur in North Korea; focus group interviews were conducted to assess the technical aspects and impacts of disasters. Through this effort, the techniques most applicable to the problems resulting from water-related disasters all over North Korea were evaluated. Based on the evaluation results, the priorities of various techniques for application were calculated in North Korea through quadrant analysis. Finally, an additional analysis of application of these techniques was conducted for water-related disasters, which continue to cause significant damage in North Korea.

As the incidence of water-related disasters has increased, monitoring techniques using various types of data have been developed around the world, including in Korea. In particular, remote sensing technique is essential in low-access areas such as North Korea as it makes periodic monitoring of large areas possible. As shown in previous research, various types of disasters are currently being monitored using satellite images. In addition, various spatial analyses are performed based on monitoring data, allowing for the establishment of integrated water-related disaster management. Especially for water-related disasters, the use of SAR satellite imagery with clear water reflectivity is known to be very effective [3, 4].

This study sought to clarify which South Korean disaster damage reduction techniques using remote sensing and spatial information could be used for reducing damage in North Korea. The study first summarized the current situation of water-related disasters in North Korea through a review of the literature, websites, and related reports [5]. The status of water-related disasters in North Korea was aggregated by year and cause. Then, damage reduction techniques for water-related disasters used in South Korea were investigated and their applicability to North Korea was evaluated. Third, the methodologies and data (remote sensing and GIS data) used in South Korea were organized and analyzed. Fourth, the use of C band synthetic aperture radar satellites by South Korea, which can directly monitor North Korean water-related disasters in the future, was presented. Finally, the study summarized the approaches to reduce the damage in North Korea. The detailed workflow is shown in Figure 1.

2. Data and Methodology

2.1. Investigation of Natural Disasters Occurring in North Korea. To identify applicable techniques based on the occurrence of disasters in North Korea, North Korea's natural disaster occurrence data were collected. Due to the isolationist policy of the North Korean government, comprehensive published statistical data on natural disasters in North Korea are difficult to gather. Therefore, this study was based mainly on newspapers and press releases, which represent the official media of North Korea. In addition, data on North Korea published by disaster-related international organizations were reviewed to improve the credibility of the dataset. An interim report published by the International Federation of Red Cross and Red Crescent Societies (IFRC) provided information about North Korea's response to and recovery from natural disasters. In addition, data published in South Korea were compared and analyzed to minimize missing data. This process led to a summary of data on the frequency of natural disasters in North Korea as previously investigated [5].

By comparing data from various sources, systematic investigation and summarization of the status of disasters in North Korea were conducted [5]. Due to insufficient data on recent occurrences, the survey period was set as January 2001 to December 2017 (Table 1). Based on data from a previous study of disasters in North Korea [6], additional investigations and reviews were conducted and the results reclassified using the classification system presented in this

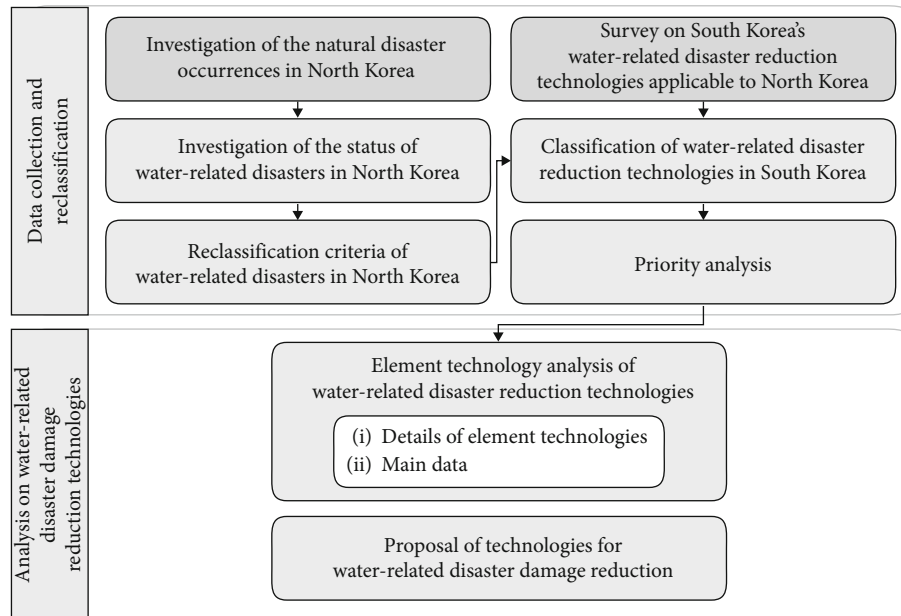


FIGURE 1: Detailed workflow for this study.

TABLE 1: Water-related disaster occurrence rates in North Korea between 2001 and 2017 [5].

	Typhoon	Downpour	Flood	Landslide	Heavy snowfall	Drought	Total of water-related disasters
2001	1	2	1	2	2	2	10
2002	5	3	1	1	0	0	10
2003	0	1	0	1	0	0	2
2004	1	5	2	1	0	0	9
2005	2	4	1	1	0	0	8
2006	1	3	2	2	0	0	8
2007	2	2	1	0	0	0	5
2008	2	1	1	0	0	0	4
2009	0	1	0	1	0	0	2
2010	3	4	4	0	0	0	11
2011	4	5	2	1	2	1	15
2012	3	5	5	2	0	1	16
2013	0	1	1	1	0	0	3
2014	0	0	0	0	0	1	1
2015	1	2	2	0	0	1	6
2016	1	2	1	2	0	0	6
2017	0	0	0	3	0	1	4
Total	26	41	24	18	4	7	120

study. North Korea is a very secretive country; as a result, it is difficult to acquire quantitative data on North Korea. Our results were obtained by reviewing previous research on natural disasters [5, 6], focusing on water-related disasters.

According to the integrated data [5, 6], 114 earthquakes occurred in North Korea, and more than one third of all natural disaster occurrences were caused by earthquakes. However, the magnitudes of the earthquakes were relatively low compared to the number of occurrences, so the damage was not significant. The earthquakes that occurred, most of

which were artificially induced, generally ranged in magnitude from 2.0 to 4.0. Natural disasters related to water are important among the results in [5, 6] because although earthquakes are reported frequently, the degree of damage is not directly proportional [5]. Other natural disasters, such as heat waves, are less frequent, which indicates that the most damaging natural disasters are water-related. According to Figure 2, where the number of natural disasters is compared, it can be seen that the occurrence of water-related disasters are relatively higher than other natural disasters. In addition

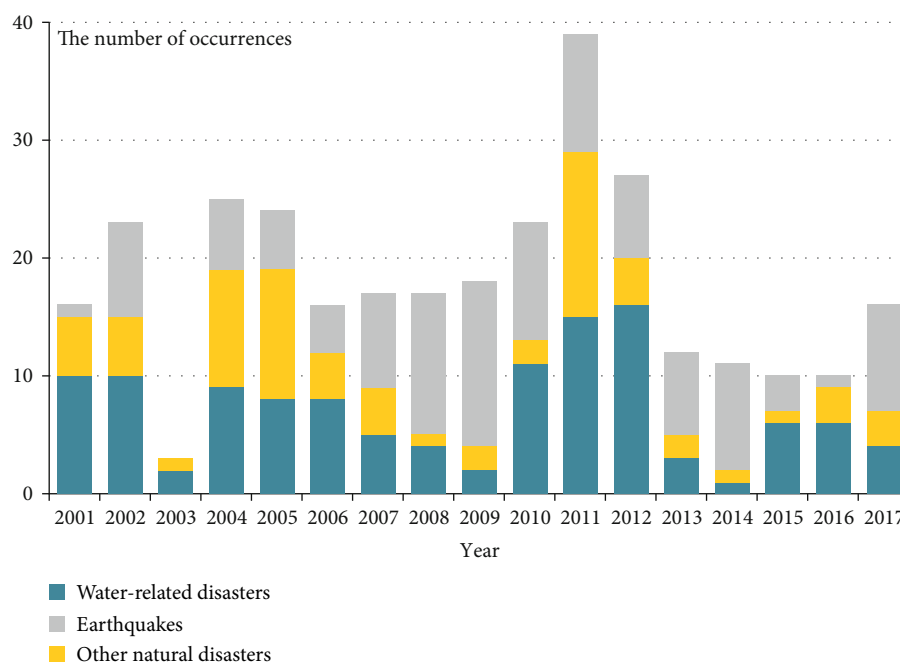


FIGURE 2: The number of natural disasters of water-related disasters, landslides, and other natural disasters.

to the frequency of occurrence, the actual damage from water-related disasters is higher than others [6, 7]; other disasters include cold wave, heat wave, and sandy dust.

Therefore, this study analyzed disasters that have a great impact on people due to the actual damage caused [6, 8]. This study classified water-related disasters into six categories according to the classification systems of South and North Korea: typhoons, downpours, floods, landslides, heavy snowfalls, and droughts.

Among the types of water-related disaster classified in this study, the disasters, including typhoons, strong winds, downpours, and floods, were relatively frequent and also resulted in large losses of life and property. The cumulative frequency of downpour in North Korea for the 17 years was highest after an earthquake. Thirteen downpours, about 30 percent of the total, were occurred together due to typhoons [6]. The occurrence of downpours was associated with floods, landslides, and tsunamis, which caused greater and wider damage. According to a report by a UN coordinator currently residing in North Korea, 231 people died and 212,000 were adversely affected by downpours in North Pyongan Province and South Pyongan Province in 2012 [9]. In 2013, the following year, 189 people died and 800,000 people were affected in the area. In 2014, the region suffered 18 months of drought. In 2015, Typhoon Goni caused severe damage in South Hwanghae Province, North and South Hamgyong Province, and the city of Naseon. The typhoon left 22,000 victims of dead, injured, and displaced in Naseon. In a typical case of downpour in North Hamgyong Province, which occurred between August 29 and September 2, 2016, 138 people died and 680,000 people were displaced, as reported by the Chosun Central Broadcasting Corporation (Figure 3(a)).

North Korea, which is mostly mountainous and has high average altitude, experiences cold weather and heavy snow damage during the winter season between December and February. Heavy snowfall in Gangwon-do caused heavy damage and economic loss in February 2011. Droughts, on the other hand, occurred mostly between March and June and throughout all regions of North Korea. The number of droughts was higher in Hwanghae Province than in other regions. A drought lasting from March to August, 2015 was among the worst (Figure 3(b)). In 2017, a drought occurred in the granary areas of Hwanghae and Pyongan Provinces, and a state of emergency was declared in June. According to the report by the UN coordinator, in recent years, North Korea has suffered from water-related disasters every year [9].

2.2. Investigation and Classification of Techniques for Water-Related Disasters in South Korea. This section examined the current state of science and technology related to water-related disasters in South Korea to suggest an appropriate direction for North Korea's disaster response technique. The National Science and Technology Information Service (NTIS), a national science and technology knowledge information portal operated by South Korea's Ministry of Science and Technology, was used to collect information on projects pertaining to water-related disaster reduction techniques in Korea [12]. NTIS provides a range of information about the projects, including content, research equipment, and researchers. Based on the NTIS service, projects in science and technology undertaken through the last decade (2008–2017) were collected to serve as basic data for analysis.

To investigate science and technology projects related to water-related disasters, research and development projects were identified using related keywords and were then



FIGURE 3: Damage from water-related disasters in North Korea: (a) flood of Hwanghae Province in 2015 [10] and (b) flood of North Hamgyong Province in 2016 [11].

classified into the six types (typhoon, downpour, flood, landslide, heavy snowfall, and drought) determined previously. To reduce the damage in North Korea, South Korean techniques using remote sensing and spatial information were searched for and the resulting reduction in damage from water-related disasters was examined. Finally, 22 projects classified by the type of disaster were collected. The projects were examined with the aim of identifying a maximum of ten of each technique. These were then categorized according to the six types. By analyzing the technical elements of each technique, the importance of using radar satellite data for water-related disasters was also analyzed.

3. Results

3.1. Element Technique Analysis of Water-Related Disaster Reduction Techniques. To analyze the overall priority of water-related disaster-reduction techniques, the technique was analyzed based on the investigation and classification results of techniques in South Korea. Total 22 research and development technologies were determined to be reduction techniques for water-related disasters. Then, the elements of each technique for water-related disasters were derived based on the results. The basic data essential to each technique are also listed.

The final result showed high priorities in three categories: water-related disaster monitoring using spatial information technique, research infrastructure and database construction based on a geographic information system data (GIS), and complex disaster integrated management. Especially, research database construction based on water-related disaster monitoring technique in the form of spatial information for the whole of North Korea for integrated disaster management can be used as a fundamental technique for reducing damage.

Representative techniques are summarized in Table 2. Sixteen projects, 84.21% of the total, applied spatial information-related techniques. Synthetic aperture radar (SAR) remote sensing was also related to 12 techniques

(63.16%), and optical and meteorological methods were related to 3 and 10 techniques, respectively, for a total of 68.42%. In particular, the C band, such as from RADARSAT-1, RADARSAT-2, and Sentinel-1, was used for the main applications of SAR remote sensing for water-related disasters.

In addition to the high use of radar satellite images for water-related disasters, such as the use of C band data, a technique using additional data elements was also represented. Various national thematic maps such as soil, land use, land cover, and forest maps, which are based on satellite images, have been used. These fundamental data could be also constructed for North Korea based on satellite images. Due to the occurrence of disasters in North Korea, the utilization of a technique that employs climate change scenarios to predict rainfall pattern variability in the future is important because the damage caused by typhoons or downpours is high [13, 14].

3.2. Case Studies of the Application of Satellite Images to Water-Related Disasters. Water-related disasters have been increasing all over the world, but especially in North Korea, where the damage caused by the disasters such as floods and typhoons is severe; the disasters are widespread in inaccessible areas and are the most destructive. As a result of research, the importance of accurate and continuous monitoring using remote sensing satellite imagery is increasing with the research on water-related disasters. By comparing before and after data of various water-related disasters or by tracking the progress of disasters, research aims to obtain a variety of information for selecting indicators, e.g., topography and vegetation characteristics, and effective response to disasters. This study organized disaster-related projects in South Korea and analyzed the data used and the techniques developed in each project that might be applicable for water-related disaster damage reduction in North Korea. Due to the secretive nature of North Korea, it is difficult to obtain actual data; however, data were derived by reviewing the available techniques to identify potential inter-Korean cooperative measures.

TABLE 2: Results from priority analysis of techniques required in North Korea.

Water-related disaster class of project	Project	Element technique	Main data
Flood	Development of system for vulnerability analysis of natural disaster	(1) SAR remote sensing (2) GIS	(1)RADARSAT-1, RADARSAT-2 (2) Land use and cover (3) DEM (4) Forest map
Flood	Development of system for flood of stream and risk analysis	(1) GIS (2) SAR remote sensing	(1) DEM (2) Land use and cover (3) Location of flood
Flood	Development of flash flood damage prediction method in mountainous areas	(1) SAR remote sensing (2) GIS	(1) RADARSAT-1, RADARSAT-2 (2) Land use and cover (3) DEM (4) Forest map
Flood	Pilot development and operation of urban flood forecasting project based radar-rainfall -	(1) SAR remote sensing (2) Optic remote sensing	(1) Sentinel-1 (2) COSMO-SkyMed (3) WorldView-2, WorldView-3
Flood	Development of the evaluation technology for complex causes of inundation vulnerability and the response plans in coastal urban areas for adaptation to climate change	(1) GIS	(1) Land use and cover (2) Climatic change scenario (RCP 4.8, 8.5)
Flood	Development of flood forecasting technology for catching golden time	(1) GIS	(1) Climatic change scenario (RCP 4.8, 8.5)
Flood	Establishment of city flood response system by region	(1) SAR remote sensing (2) GIS	(1) RADARSAT-2 (2) Sentinel-1 (3) Land use and cover (4) DEM etc.
Flood	Urban flood analysis and flood safety improvement technology	(1) SAR remote sensing (2) GIS	(1) RADARSAT-2 (2) Sentinel-1 (3) Land use and cover
Flood	Development of hydrological model for decision making to flood response	(1) SAR remote sensing (2) GIS	(1) RADARSAT-2 (2) Sentinel-1 (3) Land use and cover
Flood	Intelligent flood prediction and warning system	(1) SAR remote sensing (2) GIS	(1) RADARSAT-2 (2) Sentinel-1 (3) COSMO-SkyMed (4) Land use and cover
Typhoon	Implementation of typhoon related disaster information database using remote sensing data	(1) SAR remote sensing	(1) RADARSAT-1, RADARSAT-2 (2) Sentinel-1
Typhoon	Developing detection/prediction technique for tropical cyclone formation based on satellite and numerical model	(1) Meteorological satellite remote sensing	(1) Chollian 1, Chollian 2 (2) Meteorological data
Typhoon	A research for typhoon track prediction using end-to-end deep learning technique	(1) Meteorological satellite remote sensing	(1) Chollian 1, Chollian 2
Typhoon	Developing typhoon prediction system by multi-model ensemble technique	(1) Meteorological satellite remote sensing (2) SAR remote sensing	(1) Chollian 1, Chollian 2 (2) RADARSAT-1, RADARSAT-2 (3) Sentinel-1 (4) COSMO-SkyMed
Typhoon	Prediction of medium- and long-range tropical cyclone activity using the statistical-dynamical forecast system	(1) Meteorological satellite remote sensing	(1) Chollian 1, Chollian 2 (2) Meteorological data

TABLE 2: Continued.

Water-related disaster class of project	Project	Element technique	Main data
Typhoon	Study on methods to track northwestern Pacific typhoons paths using seismic array data	(1) Meteorological satellite remote sensing	(1) Chollian 1, Chollian 2 (1) Meteorological data
Typhoon	Disastrous weather/climate analysis and prediction	(1) GIS	(2) Climatic change scenario (RCP 4.5 and 8.5)
Downpour	Improvement of understanding and prediction capability for high-impact heavy precipitation systems over the Korean peninsula	(1) Meteorological satellite remote sensing (2) SAR remote sensing	(1) Chollian 1, Chollian 2 (2) RADARSAT-1, RADARSAT-2
Downpour	Development of a prediction system for heavy rainfalls caused by cloud clusters over the Korean peninsula	(1) GIS (2) SAR remote sensing	(1) Meteorological data (2) Sentinel-1 (3) COSMO-SkyMed
Drought	Development of the software module in response to drought events based on satellite imagery in Korea	(1) Optic remote sensing	(1) Landsat 5, Landsat 6, Landsat 7
Drought	Development of a drought monitoring system through integrated modeling of multi-sensor data	(1) Meteorological satellite remote sensing	(1) Chollian 1, RADARSAT-2 (2) Meteorological data
Heavy snowfall	Development of a prompt action technique using remote sensing	(1) Optic remote sensing	(1) Landsat 5, Landsat 6, Landsat 7

Various satellite images were used to analyze cases of flood disaster-related hazards (downpour, flood, typhoon, drought, landslide, etc.). Representatively, there was one case of massive flood mapping using optical satellite images [15]. Landsat satellite images, which are medium- to low-resolution satellite images, were used to evaluate the flood surface [16], and methodologies for mapping the inundation site were compared [17]. On the other hand, high-resolution satellite images have been used to monitor dam reservoirs [18] and to analyze spatial changes in floods over time [19]. Precise mapping of flooding in plain basin boundaries was studied [20]. In relation to typhoons, studies have been carried out to establish early warning systems for sea hurricanes [21] and information on sea-level environments [22].

Unlike optical satellite images, SAR images can be used to monitor floods irrespective of weather conditions and during both day and night. Accordingly, SAR is advantageous when estimating urban and agricultural damage from river flooding; and flood mapping techniques using SAR images have been studied continuously [23, 24]. In particular, studies to observe floods with the data obtained immediately after the flood have been actively conducted [25, 26]. A flood risk mapping study using 16 years of rainfall and historical satellite data was also carried out employing the advantages of satellite images, which can be monitored periodically [27]. In addition, flood depth was estimated using high-resolution SAR images [28], and filter comparison analysis was performed to interpret radar images for flood mapping [29].

In recent years, research has been conducted to solve additional environmental problems (landslides, slope stability, etc.) by applying GIS and machine learning technique to the monitoring of information constructed through satellite images. Typically, hydrological modeling for flood forecasting based on GIS [30], spatial modeling using GIS for

flooding in urban areas [31], and flood risk assessment in hazardous areas [32] have been performed. Recently, machine learning has been used to analyze the risk of floods [33], and multiple risks for complex disasters have been mapped [34]. In addition, in one case, disasters were evaluated by applying deep learning to satellite imagery [35]. There are also cases of employing various GIS-based techniques that use remote sensing data for vulnerability analysis of water-related disasters [36–38]. Satellite imagery is also being used actively to assess technological developments related to landslides, which have high frequency and cause large damage in North Korea [39, 40]. Various studies have been conducted to map landslides, including studies on soil moisture [41] for landslide risk assessment [42, 43].

Despite the low priority in the results from priority analysis of this study, satellite images are also used in studies related to drought and heavy snow, which need to be expanded in the long term. Satellite images are especially used for drought monitoring [44]. In addition, in some cases, the risk of drought has been assessed by geospatial methods [45]; for example, the impact of land destruction due to drought has been assessed in Brazil [46]. Examples of estimating snow depth due to snowfall based on MODIS images [47] and surface modeling [48] have also been reported.

The use of satellite images for water-related disasters and global disaster monitoring in the Korean Peninsula and other inaccessible areas will help prevent the occurrence of and mitigate the damage caused by the disasters [15, 49–51]. In this regard, South Korea is planning to launch the Compact Advanced Satellite 500-5 (CAS500-5) to reduce the damage caused by water-related disasters occurring in South and North Korea through inter-Korean cooperation. In particular, water disaster satellites with high-resolution C band SAR, which will be directly applicable to water-related

disasters, are scheduled to be launched by the Ministry of Environment and K-water. The C band radar satellite images will continuously monitor the disasters in South and North Korea, and integrated disaster management on the Korean Peninsula will finally be performed using a database constructed from the accumulated satellite images.

4. Discussion and Conclusions

There is a need for technological measures to reduce the damage caused by disasters at the levels of international relations and government development in North Korea. In particular, it is necessary to reinforce the basic capacity of North Korea's disaster response. However, North Korea's closed attitude allows only one-time assistance rather than fundamental technological improvements. Due to social and economic conditions in North Korea, the quality of life of North Korean citizens is lower than that in neighboring countries, and needs that affect daily life are therefore more important. Thus, it is necessary to lay the foundation for joint research and development with North Korea. Therefore, this study is aimed at providing a scientific support system for responding to water-related disasters. In particular, techniques related to water-related disasters, such as downpours and typhoons, which cause serious damage are shown as a high priority. In the case of earthquakes, on the other hand, although the frequency of occurrence is high, the importance is low because losses from earthquakes are not great.

Research already conducted and techniques already employed in South Korea are similar to those needed in North Korea [15, 52, 53]. The research and development technologies which have the highest priority among the investigated techniques should be developed first because they can provide accurate water-related disaster information for North Korea over the long term. Finally, based on the disasters investigated in this study that are currently occurring in North Korea, the techniques to be applied to North Korea can be classified into three categories in terms of their priority for development against water-related disasters: techniques for disaster monitoring using remote sensing and spatial information, those for geoinformation system and GIS-based database construction, and those for integrated water-related disaster management.

In conclusion, an integrated response system for water-related disasters on the Korean Peninsula could be established by comprehensively managing information pertaining to water-related disasters occurring in North Korea. The application of spatial analysis is essential for reducing damage caused by the disasters in spatially adjacent areas. Therefore, the construction of a spatial database, which is not currently available for North Korea, and the infrastructure for basic research are fundamental to all future disaster management efforts. Research and development technologies of low priority must also be developed continuously over the mid to long term.

The characteristics of North Korea make it difficult to manage spatial data for water-related disaster techniques. Therefore, it is necessary to establish a basic infrastructure for research based on GIS to integrate data from various

types of water-related disasters. Recent water-related disasters have tended to evolve from a single disaster into a more complex form, so an integrated disaster response system must be established through monitoring and cumulative analysis of the spatial distribution of disasters through continuous remote sensing data. Furthermore, information on water-related disasters should be integrated and managed systematically through joint development and international cooperation. The results of priority analyses of technological developments can enhance the application of scientific and technological products to practical situations. In addition, it will be possible to lay the groundwork for coping with water-related disasters and for reducing damage through the systematic sharing of the technique.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research was conducted at the Korea Environment Institute (KEI) with support from the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07041203) and support from the Korea Environment Industry & Technology Institute (KEITI) from 'The Application technology and system of satellite image radar in the environmental' and funded by the Korea Ministry of Environment (MOE) (2019002650001).

References

- [1] D. Alexander, *Natural Disasters*, Routledge, 2017.
- [2] T. Andrew, V. Luca, M. F. Montserrat, and D. Brian, *Inform Global Risk Index Results 2018*, Inform, Ed., Inform, EU, 2017.
- [3] S. Hakdaoui and A. Emran, "Extraction of water information based on SAR RADAR and Optical image processing: case of flood disaster in Southern Morocco," in *Geospatial Technology*, pp. 15–29, Springer, 2020.
- [4] N. Kussul, A. Shelestov, and S. Skakun, "Flood monitoring from SAR data," in *Use of Satellite and In-Situ Data to Improve Sustainability*, pp. 19–29, Springer, 2011.
- [5] S. Lee, T. Song, and M.-j. Lee, "A study on the inter-Korean cooperation for natural disaster damage reduction using spatial information," *Korean Journal of Remote Sensing*, vol. 35, no. 1, pp. 163–177, 2019.
- [6] T. G. Kang, J. Hoon, J. S. Lee et al., "Data collection for natural disaster management in the Democratic People's Republic of Korea and cooperation between the two Koreas," Korea Environment Institute, Sejong, 2016.
- [7] M.-J. Lee, S. J. Myung, T. G. Kang, T. H. Kim, S. Lee, and N.-W. Jo, "Disaster safety R & D project planning research for inter-Korean exchange and cooperation," Korea Environment Institute, Sejong, 2018.
- [8] S. J. Myung, H. J. Hong, H. I. Choi, and J. C. Jung, "Estimation of flood vulnerable areas in North Korea and collaboration strategies between South Korea and North Korea," Korea Environment Institute, Sejong, 2008.

- [9] OCHA, *Needs and Priorities DPR Korea*, H. P. Cycle, Ed., The United Nations Office for the Coordination of Humanitarian Affairs, 2016.
- [10] TRUST ORG, *North Korea – food aid reaches 140,000 hit by floods says WFP*, FloodList, 2016.
- [11] UN DPRK, *North Hamgyong floods 2016*, N. H. F. 2016, 2016.
- [12] NTIS, *National science & technology information service*, Ministry of Science and ICT, 2019.
- [13] S. Kim, Y. Tachikawa, and K. Takara, *Recent flood disasters and progress of disaster management system in Korea*, 2007.
- [14] S.-Y. Park, B.-J. Kim, and S.-H. J. J. O. K. S. O. H. M. Ahn, “Characteristics of natural disaster in North Korea,” *Journal of the Korean Society of Hazard Mitigation*, vol. 10, no. 3, pp. 21–29, 2010.
- [15] Y. Piao, H.-S. Lee, K.-T. Kim, and K.-S. Lee, “Methodology to apply low spatial resolution optical satellite images for large-scale flood mapping,” *Korean Journal of Remote Sensing*, vol. 34, no. 5, pp. 787–799, 2018.
- [16] A. Maxim, R. Adrian, G. Lucian, M. Gabriel, and R. Bogdan, “Assessing flooded surface area using Landsat satellite data on the Siret river downstream of the Lower Danube,” *Annals of the University Dunarea de Jos of Galati: Fascicle II, Mathematics, Physics, Theoretical Mechanics*, vol. 41, 2018.
- [17] D. Munasinghe, S. Cohen, Y. F. Huang, Y. P. Tsang, J. Zhang, and Z. Fang, “Intercomparison of satellite remote sensing-based flood inundation mapping techniques,” *Journal of the American Water Resources Association*, vol. 54, no. 4, pp. 834–846, 2018.
- [18] S. Yoon, S. Lee, K. Park, S. Jang, and J. Rhee, “Development of a storage level and capacity monitoring and forecasting techniques in Yongdam Dam Basin using high resolution satellite image,” *Korean Journal of Remote Sensing*, vol. 34, no. 6_1, pp. 1041–1053, 2018.
- [19] N. M. de Musso, D. Capolongo, A. Refice, F. P. Lovergine, A. D’Addabbo, and L. Pennetta, “Spatial evolution of the December 2013 Metaponto plain (Basilicata, Italy) flood event using multi-source and high-resolution remotely sensed data,” *Journal of Maps*, vol. 14, no. 2, pp. 219–229, 2018.
- [20] M. Mishra, V. Dugesar, K. N. Prudhviraju, S. B. Patel, and K. Mohan, “Precision mapping of boundaries of flood plain river basins using high-resolution satellite imagery: a case study of the Varuna river basin in Uttar Pradesh, India,” *Journal of Earth System Science*, vol. 128, no. 4, p. 105, 2019.
- [21] Y. Yu, H. Shi, and Y. Wang, “Early warning method for sea typhoons using remote-sensing imagery based on improved support vector machines (SVMs),” *Journal of Coastal Research*, vol. 82, no. sp1, pp. 180–185, 2018.
- [22] D. Song, L. Guo, Z. Duan, and L. Xiang, *Impact of major typhoons on sea surface environment in the northwestern Pacific derived from satellite remote sensing*, 2018.
- [23] M. A. Clement, C. G. Kilsby, and P. Moore, “Multi-temporal synthetic aperture radar flood mapping using change detection,” *Journal of Flood Risk Management*, vol. 11, no. 2, pp. 152–168, 2018.
- [24] Y. T. Jung, S.-E. Park, C.-S. Baek, and D.-H. Kim, “Evaluation of polarimetric parameters for flood detection using PALSAR-2 quad-pol data,” *Korean Journal of Remote Sensing*, vol. 34, no. 1, pp. 117–126, 2018.
- [25] R. Hostache, M. Chini, L. Giustarini et al., “Near-real-time assimilation of SAR-derived flood maps for improving flood forecasts,” *Water Resources Research*, vol. 54, no. 8, pp. 5516–5535, 2018.
- [26] X. Huang, C. Wang, and Z. Li, “A near real-time flood-mapping approach by integrating social media and post-event satellite imagery,” *Annals of GIS*, vol. 24, no. 2, pp. 113–123, 2018.
- [27] N. Alahacoon, K. Matheswaran, P. Pani, and G. Amarnath, “A decadal historical satellite data and rainfall trend analysis (2001–2016) for flood hazard mapping in Sri Lanka,” *Remote Sensing*, vol. 10, no. 3, p. 448, 2018.
- [28] F. Cian, M. Marconcini, P. Ceccato, and C. Giupponi, “Flood depth estimation by means of high-resolution SAR images and lidar data,” *Natural Hazards and Earth System Sciences*, vol. 18, no. 11, pp. 3063–3084, 2018.
- [29] D. Kim, H.-S. Jung, and W. Baek, “Comparative analysis among radar image filters for flood mapping,” *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 34, no. 1, pp. 43–52, 2016.
- [30] S. Zafar, H. M. S. Azhar, and A. Tahir, “A GIS based hydrological model for river water level detection & flood prediction featuring morphological operations,” *Proceedings of International Conference on Artificial Life and Robotics*, vol. 23, pp. 191–195, 2018.
- [31] S. A. Mohamed, “Application of satellite image processing and GIS-spatial modeling for mapping urban areas prone to flash floods in Qena governorate, Egypt,” *Journal of African Earth Sciences*, vol. 158, p. 103507, 2019.
- [32] V. A. Chinaiwala, T. S. Dalia, D. P. Dave, J. R. Gaekwad, Y. D. Jariwala, and N. Soni, “Flood hazard assessment and identification of danger zones using GIS,” *METHODOLOGY*, vol. 4, no. 4, 2018.
- [33] M. Ma, C. Liu, G. Zhao et al., “Flash flood risk analysis based on machine learning techniques in the Yunnan Province, China,” *Remote Sensing*, vol. 11, no. 2, p. 170, 2019.
- [34] O. Rahmati, S. Yousefi, Z. Kalantari et al., “Multi-hazard exposure mapping using machine learning techniques: a case study from Iran,” *Remote Sensing*, vol. 11, no. 16, p. 1943, 2019.
- [35] L. Yang and G. Cervone, “Analysis of remote sensing imagery for disaster assessment using deep learning: a case study of flooding event,” *Soft Computing*, vol. 23, no. 24, pp. 13393–13408, 2019.
- [36] S. Lee, S. Lee, M.-J. Lee, and H.-S. Jung, “Spatial assessment of urban flood susceptibility using data mining and geographic information system (GIS) tools,” *Sustainability*, vol. 10, no. 3, p. 648, 2018.
- [37] J. Lim and K.-s. Lee, “Flood mapping using multi-source remotely sensed data and logistic regression in the heterogeneous mountainous regions in north korea,” *Remote Sensing*, vol. 10, no. 7, p. 1036, 2018.
- [38] S. Samanta, D. K. Pal, and B. Palsamanta, “Flood susceptibility analysis through remote sensing, GIS and frequency ratio model,” *Applied Water Science*, vol. 8, no. 2, p. 66, 2018.
- [39] Francioni, Calamita, Coggan et al., “A multi-disciplinary approach to the study of large rock avalanches combining remote sensing, GIS and field surveys: the case of the Scanno Landslide, Italy,” *Remote Sensing*, vol. 11, no. 13, p. 1570, 2019.
- [40] C. Zhao and Z. Lu, “Remote sensing of landslides—a review,” *Remote Sensing*, vol. 10, no. 2, p. 279, 2018.
- [41] L. Zhuo, Q. Dai, D. Han, N. Chen, B. Zhao, and M. Berti, “Evaluation of remotely sensed soil moisture for landslide hazard assessment,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 1, pp. 162–173, 2019.

- [42] F. Fiorucci, D. Giordan, M. Santangelo, F. Dutto, M. Rossi, and F. Guzzetti, "Criteria for the optimal selection of remote sensing optical images to map event landslides," *Natural Hazards and Earth System Sciences*, vol. 18, no. 1, pp. 405–417, 2018.
- [43] P. Lu, Y. Qin, Z. Li, A. C. Mondini, and N. Casagli, "Landslide mapping from multi-sensor data through improved change detection-based Markov random field," *Remote Sensing of Environment*, vol. 231, p. 111235, 2019.
- [44] B. R. Rushi, V. Mishra, W. L. Ellenburg, F. M. Qamer, A. S. Limaye, and D. Irwin, "Application of satellite remote sensing in drought monitoring in Bangladesh using a user-friendly tool," *AGU Fall Meeting Abstracts*, 2018.
- [45] M. K. Abuzar, M. Shafiq, S. A. Mahmood et al., "Drought risk assessment in the Khushab region of Pakistan using satellite remote sensing and geospatial methods," *International Journal of Economic Environmental Geology*, vol. 10, no. 1, pp. 48–56, 2019.
- [46] D. A. Mariano, C. A. C. d. Santos, B. D. Wardlow et al., "Use of remote sensing indicators to assess effects of drought and human-induced land degradation on ecosystem health in Northeastern Brazil," *Remote Sensing of Environment*, vol. 213, pp. 129–143, 2018.
- [47] D. Kim, H.-S. Jung, and J.-C. Kim, "Comparison of snow cover fraction functions to estimate snow depth of South Korea from MODIS imagery," *Korean Journal of Remote Sensing*, vol. 33, no. 4, pp. 401–410, 2017.
- [48] G. De Lannoy, A. Vanrykel, H. Lievens, E. Kim, and L. Brucker, "Snow estimation under a vegetation gradient using satellite remote sensing data and land surface modeling during SnowEx 2017," *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 6294–6297, 2018.
- [49] J. Kim, H. Ho, M.-J. Um, and Y. Kim, "Spatial and temporal variations in the water use efficiency and its drought signal on the Korean peninsula using MODIS-derived products," *Korean Journal of Remote Sensing*, vol. 34, no. 3, pp. 553–564, 2018.
- [50] J. Park, I. Jung, and K. Park, "Evolution of bias-corrected satellite rainfall estimation for drought monitoring system in South Korea," *Korean Journal of Remote Sensing*, vol. 34, no. 6_1, pp. 997–1007, 2018.
- [51] S.-J. Park and C.-W. Lee, "Simulation of the flood damage area of the Imjin River basin in the case of North Korea's Hwang-gang Dam discharge," *Korean Journal of Remote Sensing*, vol. 34, no. 6_1, pp. 1033–1039, 2018.
- [52] W.-K. Baek, S.-H. Park, N.-K. Jeong, S. Kwon, W.-J. Jin, and H.-S. Jung, "A study for the techniques and applications of NIR remote sensing based on static analyses of NIR-related papers," *Korean Journal of Remote Sensing*, vol. 33, no. 5_3, pp. 889–900, 2017.
- [53] S. Lee and M.-J. Lee, "Susceptibility mapping of Umyeonsan using logistic regression (LR) model and post-validation through field investigation," *Korean Journal of Remote Sensing*, vol. 33, no. 6_2, pp. 1047–1060, 2017.

Research Article

An Improved Full-Aperture ScanSAR Imaging Method Integrating the MIAA Based Aperture Interpolation

Jiaqi Ning,^{1,2} Dacheng Liu,¹ Kaiyu Liu,¹ Heng Zhang ,¹ and Yingjie Wang¹

¹Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China 100190

²University of Chinese Academy of Sciences, School of Electronic, Electrical and Communication Engineering, Beijing, China 100049

Correspondence should be addressed to Heng Zhang; caszhmail@163.com

Received 4 February 2020; Revised 25 May 2020; Accepted 8 June 2020; Published 17 July 2020

Academic Editor: Sang-Hoon Hong

Copyright © 2020 Jiaqi Ning et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the scanning synthetic aperture radar (ScanSAR) mode, the radar antenna sweeps through different range subswaths to image a wide swath. The full-aperture imaging algorithm for ScanSAR data has been widely used because it can be realized by exploiting the existing standard high-precision Stripmap SAR processor and does not require stitch processing in the azimuth. However, both the focused image and the interferogram achieved by full-aperture processing suffer from spikes. The spikes adversely affect the ScanSAR-related applications, such as target detection and interferometry. To effectively suppress the spikes, an improved algorithm based on the missing-data iterative adaptive approach (MIAA) is proposed in this manuscript. Besides, the proposed method can also improve the azimuth resolution of ScanSAR images. Simulation and experimental results demonstrate that this algorithm has better performance when processing ScanSAR data compared with existing methods.

1. Introduction

Scanning synthetic aperture radar (ScanSAR) is a kind of operation widely used mode in the modern SAR system [1, 2]. By periodically sweeping the antenna beam through different range subswaths, ScanSAR can obtain a wide swath at the expense of sacrificing azimuth resolution [3–6]. In each subswath, the received data are blocked into bursts of radar echoes in the azimuth direction. In general, burst-by-burst [7–9] and full-aperture [10, 11] approaches can be used to process ScanSAR data and these two kinds of approaches have different processing flows. Among them, the burst-by-burst approach is the standard approach, which is computationally efficient. However, the full-aperture approach has also been used a lot by the SAR community now because it can make use of existing standard Stripmap SAR processing programs [10, 12]. The ScanSAR image focused by the full-aperture approach has been successfully used for lots of applications [13–21]. Particularly, for the ScanSAR data of the ALOS-2 satellite, only the image processed by the full-aperture approach can be used for interferometric applications [13].

In the full-aperture approach, gaps between bursts are filled with zeros and then all the bursts of a subswath can be coherently processed as Stripmap data. However, the coherently processed multiple bursts result in strong interference modulations of the azimuth impulse response function, which are called “spikes” [5, 10, 11, 22, 23]. These spikes can contaminate ScanSAR image and influence the interpretation. Furthermore, spikes also exist in the phase of the processed ScanSAR images. Thus, the spikes also occur in the ScanSAR interferogram, introducing additional noise to interferogram and degrading the interferogram quality [5, 11, 22, 23]. Spike contamination of full-aperture ScanSAR image limits ScanSAR-related applications, such as target detection and interferometry. Therefore, the development of effective techniques to suppress the spikes is very useful. Although the spikes can be reduced by a low-pass filtering operation [10], the residual effect is enough to cause image visual disturbances [22], and the originally rough azimuth resolution will get worse by this approach [10]. As the azimuth resolution of the full-aperture result is controlled by the envelope of the spikes [10], the spike suppression can enhance the azimuth resolution at the same time.

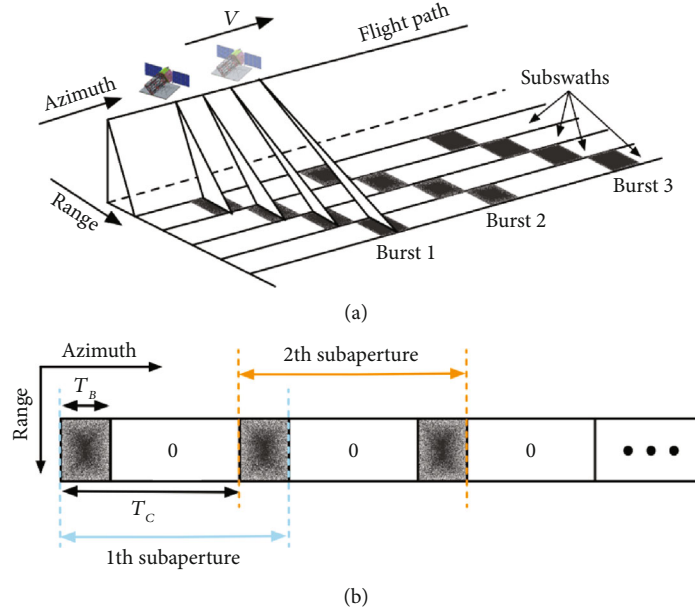


FIGURE 1: (a) The ScanSAR operation mode. (b) Acquired signal of a ScanSAR subswath.

The algorithms of suppressing the spikes while improving azimuth resolution are very beneficial to the ScanSAR-related applications.

In [22], an algorithm referred to as the linear prediction model aperture interpolation technique (LPM-AIT) was proposed to suppress spikes when the full-aperture imaging algorithm is used to process ScanSAR data. In the LPM-AIT, the gaps between bursts are recovered by the linear prediction model (LPM), and the coefficient of the LPM is estimated by the Burg algorithm [24]. A similar method is also used in the data infilling part of the algorithm in [8] to enhance the azimuth resolution when ScanSAR data imaged by a burst-by-burst approach. However, when using the LPM, the missing-data proportion should be no more than 50% [25], and the data recovery capability of the LPM decreases significantly as the proportion of missing data increases, which makes this algorithm only suitable for the ScanSAR system with no more than 2 subswaths. The number of subswaths in the ScanSAR mode is typically 2-5 [10], which means that the missing-data proportion will reach 50% to 80%. Thus, the spike suppression performance of the LPM-AIT is limited when the ScanSAR data are processed by the full-aperture imaging algorithm.

In this manuscript, an improved full-aperture ScanSAR imaging algorithm integrating the missing-data iterative adaptive approach (MIAA) based aperture interpolation technique (MIAA-AIT) is proposed. The remainder of this manuscript is organized as follows. The problem we studied is described in Section 2. Section 3 is a description of the proposed method. The results and discussions of the method are given in Section 4, and the conclusions are drawn in Section 5.

2. Problem Statement

ScanSAR obtains a wide-swath image by periodically sweeping the antenna beam through different range subswaths, as

shown in Figure 1(a). In each subswath, the received signal in the range direction is the same as those in the Stripmap mode. In the azimuth direction, however, the received signal is blocked into bursts and is equivalent to data truncated from the Stripmap mode data which causes incomplete signal phase history. Two adjacent bursts and zeros between them as a whole is called a subaperture, as shown in Figure 1(b).

With T_B as the burst duration and T_C as the burst cycle period, $N_B = T_B \times \text{PRF}$ is the number of samples per burst, and $N_C = T_C \times \text{PRF}$ is the number of samples in a burst cycle period, where PRF is the sampling frequency in the azimuth direction. Setting the number of subswaths in the ScanSAR system to N_S , then $T_C = N_S \times T_B$ [7, 26, 27]. The gaps between bursts can be treated as missing samples. The missing proportion γ can be expressed as (1)

$$\gamma = \frac{T_C - T_B}{T_C} = \frac{N_C - N_B}{N_C} = 1 - \frac{1}{N_S} \#. \quad (1)$$

In the following, we focus only on the azimuth direction of one subswath. For ScanSAR mode, a point-target azimuth signal before azimuth compression is

$$S_{\text{raw}}(\eta) = \sum_n A_0 \cdot \text{rect}\left(\frac{\eta - \eta_c^n}{T_B}\right) \exp\{j\pi K_a \eta^2\} \#, \quad (2)$$

where η is the azimuth time, A_0 is an arbitrary complex constant, K_a is the azimuth chirp rate, n is the burst number, η_c is the burst center, and the zero Doppler moment of the target η_a is set as zero. After padding zeros in the burst intervals, ScanSAR mode data can be coherently processed by standard Stripmap processors, referred to as the full-aperture approach. A point response of ScanSAR using the full-aperture imaging algorithm is equivalent to coherently adding the compression results of each burst.

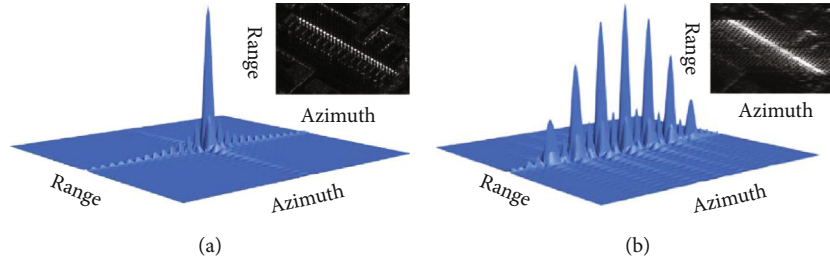


FIGURE 2: Schematic facade in two dimensions of a point target in Stripmap mode and ScanSAR mode. The upper right corner shows a real SAR scene. (a) Stripmap mode. (b) ScanSAR mode.

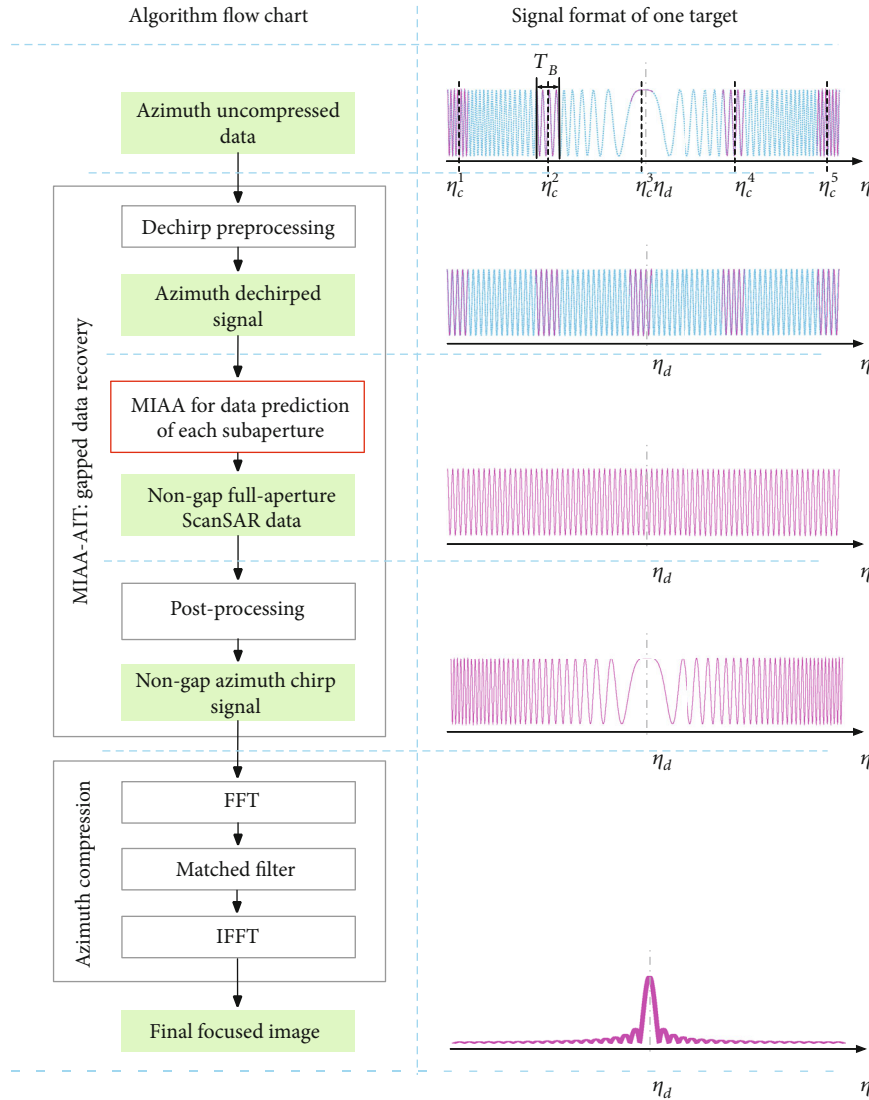


FIGURE 3: The diagram of processing flow chart and schematic signal format of one target in each step.

$$S(\eta) = \sum_n A_0 T_B \text{sinc}(K_a T_B \eta) \exp\{j\pi K_a \eta(\eta - 2\eta_c^n)\} \# \quad (3)$$

When several such pulses are added together, different phase slopes give spikes in the summation output. As shown in Figure 2(b), the main lobe and side lobe of the two-

dimensional focus result of one point target are split into several parts along the azimuth direction, which is called “spike” phenomenon. Compared to the point target of the Stripmap mode, shown in Figure 2(a), the focus result of the point target is strongly deteriorated due to the existence of spikes. The upper right corner of Figures 2(a) and 2(b) are

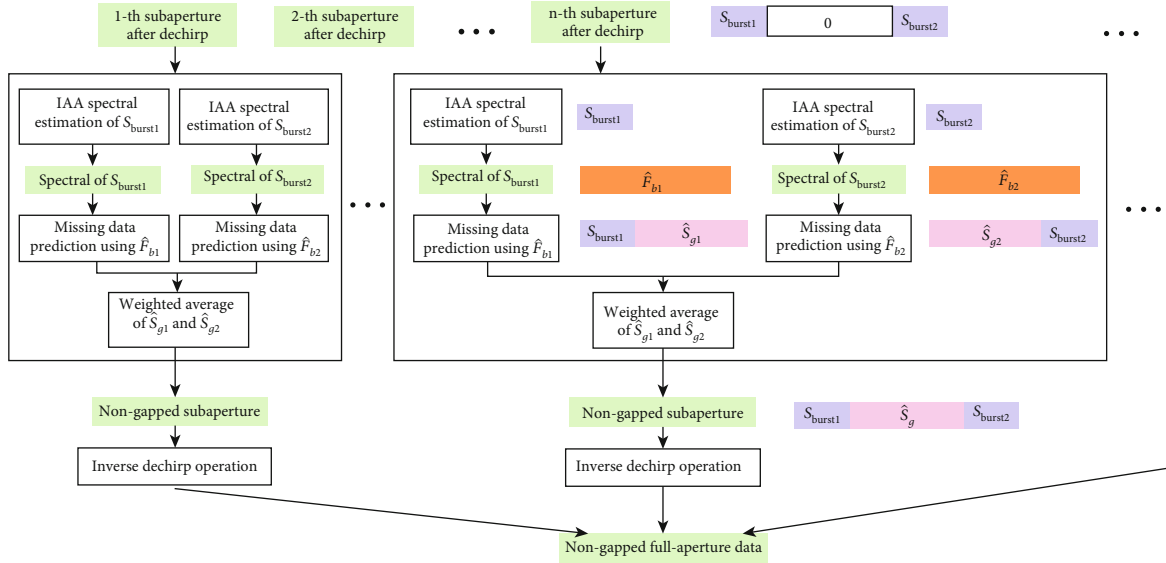


FIGURE 4: Processing flow of the MIAA-AIT.

the images of strip and ScanSAR mode in the same real SAR scene, respectively. We can see that the spikes seriously contaminate the ScanSAR image and influence the interpretation and applications [22].

Infilling the gaps between bursts with valid data is an effective strategy to suppress spikes and improve the azimuth resolution at the same time. Li et. al [22] used the LPM for data infilling between the burst. However, the number of samples that can be effectively predicted by the LPM is limited, and the missing-data proportion should be no more than 50% [25]. The number of subswaths in the ScanSAR mode is typically 2-5 [10], which means that the missing-data proportion will reach 50% to 80%. The performance of the LPM degrades significantly, and the effect of spike suppression is limited at such a high missing-data proportion. Therefore, a spike suppression approach that is more suitable for the ScanSAR data should be proposed.

3. Proposed MIAA-AIT Algorithm

After range compression and range cell migration correction (RCMC), we aim at recovering the gaps of the azimuth uncompressed ScanSAR data by the MIAA-AIT algorithm. Recovering missing data caused by incomplete phase history from available data is widely used in [24, 25]. To make the recovered result more reliable, the quadratic phase term is usually removed by the dechirp operation and retained the linear phase term [22]. Because the phase history in ScanSAR mode is limited to the short span of the image by the antenna beam width, performing dechirp operation on the entire data will cause a large number of targets located at the same distance bin to have the same frequency, which will cause target confusion [9]. Thus, the proposed MIAA-AIT is performed for each subaperture to obtain nongapped subswath data.

Figure 3 shows the diagram of the processing flow chart and schematic signal format of one target in each step. The green boxes represent signal formats at different steps. The

proposed MIAA-AIT includes three steps which are dechirp operation, MIAA for missing data prediction, and postprocessing. Among them, the red square is the key step of the manuscript, which will be explained in detail in the following. Finally, the azimuth compression is applied to the nongapped subswath data to acquire the final focused ScanSAR image.

3.1. Dechirp Preprocessing. The phase of each target is quadratic, as shown in (2), which is inconvenient for the application of MIAA. Dechirp preprocessing in the azimuth is performed by multiplying the conjugate of the azimuth chirp for each subaperture. After the dechirp processing, the target phase is linear with the azimuth, which is suitable for applying the MIAA.

3.2. Gap Recovery by the MIAA. The detailed data recovery process for each subaperture is shown in the black box of Figure 4. Let us take the n -th single subaperture as an example. Supposing that (4) is the azimuth signal of a subaperture after dechirp preprocessing,

$$\mathbf{S}_s = \left[\mathbf{S}_{\text{burst1}}[1 \times N_B] \mathbf{S}_{\text{gap}}[1 \times (N_c - N_B)] \mathbf{S}_{\text{burst2}}[1 \times N_B] \right]^T \#, \quad (4)$$

where $[\cdot]^T$ denotes the transpose, $\mathbf{S}_{\text{burst1}}$ and $\mathbf{S}_{\text{burst2}}$ are the samples of two bursts in a subaperture which are available samples, and \mathbf{S}_{gap} is the missing samples between two bursts that need to be estimated. The MIAA uses iterative adaptive approach (IAA) spectrum estimate [28] to retrieve the missing data. First, the MIAA uses the IAA to obtain an accurate spectral estimation from the given samples. Then, the missing-data recovery step is performed using the IAA spectrum estimates by a linear minimum mean-squared error (MMSE) estimator with a negligible additional computational burden [29]. In the implementation of the proposed MIAA-AIT, the missing samples are recovered by the weighted average of the estimated results by $\mathbf{S}_{\text{burst1}}$ and $\mathbf{S}_{\text{burst2}}$ for robustness and efficiency.

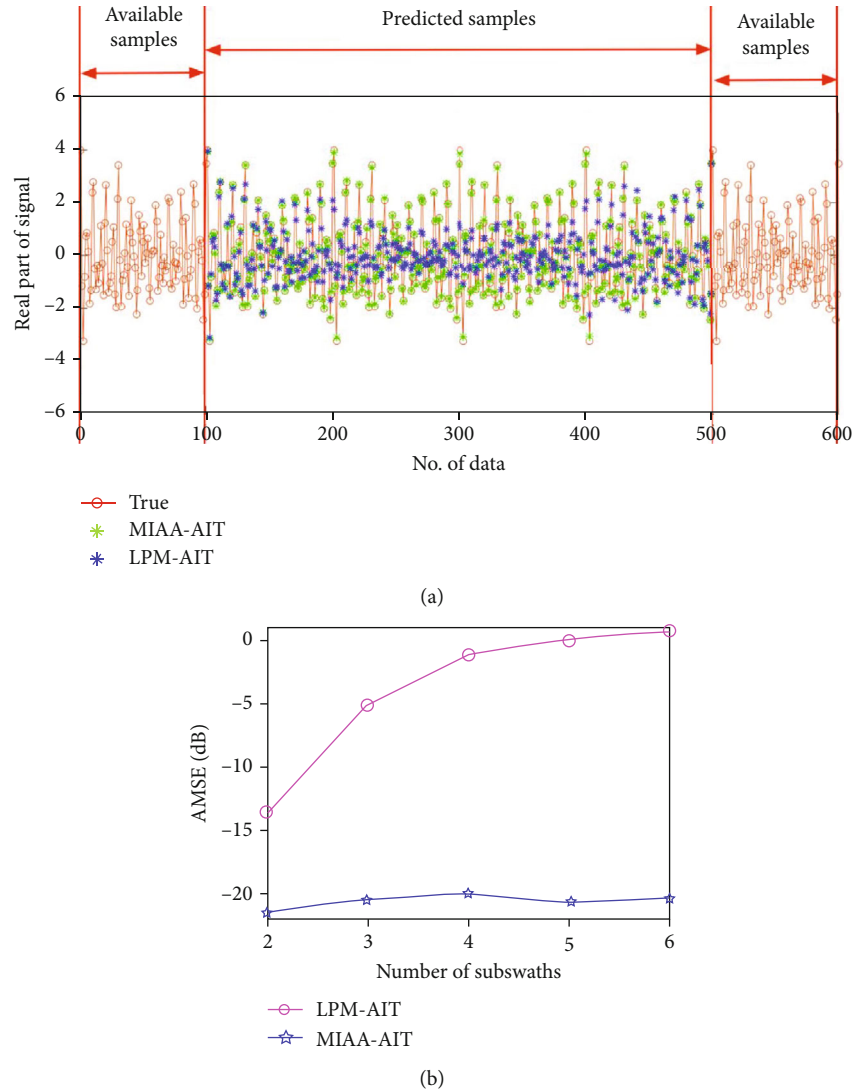


FIGURE 5: Comparison of the effectiveness of missing-data recovery. (a) Missing data recovery results for one random trial when the missing-data proportion is 80%. (b) AMSEs of the estimation results for different missing-data proportions.

Let us take the estimation by S_{burst1} as an example. Let $S_b = S_{\text{burst1}}$, $S_g = S_{\text{gap}}$ and $S = [S_b \ S_g]^T$. Let $\omega_k = 2\pi(k/K)$, $k = 0, 1, 2 \dots K-1$ be the frequency points in the frequency domain. K should be much larger than N_B [28, 29], and K is set to 8 times of N_B in our experiments. Let

$$\begin{aligned} \mathbf{a}_b(\omega_k) &= \left[1 \ e^{j\omega_k} \ \dots \ e^{j(N_B-1)\omega_k} \right]^T ; \mathbf{a}_g(\omega_k) \\ &= \left[e^{jN_B\omega_k} \ e^{j(N_B+1)\omega_k} \ \dots \ e^{j(N_C-1)\omega_k} \right]^T \# \end{aligned} \quad (5)$$

where $\mathbf{a}_b(\omega_k)$ and $\mathbf{a}_g(\omega_k)$ are the Fourier vectors corresponding to S_b and S_g at frequency ω_k .

First, the spectral of the available data S_b at the frequency ω_k is estimated by the IAA, referred to as $\hat{F}_{b1}(\omega_k)$. The IAA makes use of an adaptive weighting matrix to get the spectral of the available data in an iterative manner until convergence.

$$\hat{F}_{b1}(\omega_k) = \frac{\mathbf{a}_b^H(\omega_k) \mathbf{R}^{-1} S_b}{\mathbf{a}_b^H(\omega_k) \mathbf{R}^{-1} \mathbf{a}_b(\omega_k)} \# \quad (6)$$

where $(\cdot)^H$ denotes the conjugate transpose and \mathbf{R} is the covariance matrix of the given data, i.e.,

$$\mathbf{R} = \sum_{k=0}^{K-1} |F_{b1}(\omega_k)|^2 \mathbf{a}_b(\omega_k) \mathbf{a}_b^H(\omega_k) \# \quad (7)$$

The initialization of the IAA can be completed by setting \mathbf{R} to the identity matrix \mathbf{I} . Instead of computing \mathbf{R} , we can directly get the inverse of \mathbf{R} in each iteration. By taking advantage of the Toeplitz structure of the covariance matrix, the inversion of the covariance matrix can be converted to matrix multiplication operations with a much lower calculation amount [30, 31]. Typically, no more than 10 to 15 iterations are required for (6) reaching convergence [28, 29].

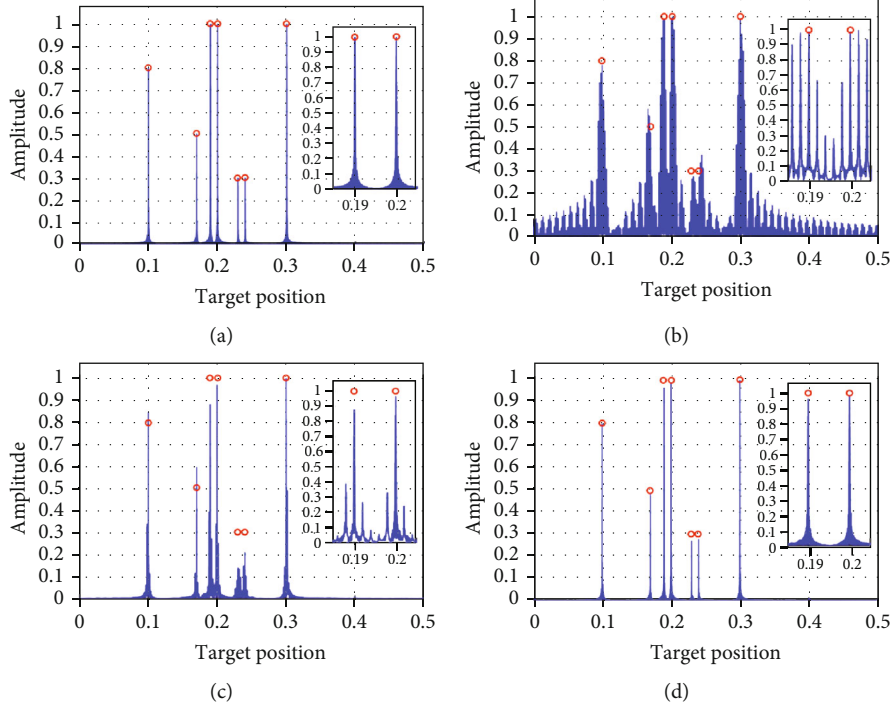


FIGURE 6: Comparison of target amplitude estimation results with (a) complete samples. (b) Available samples. (c) Available samples and LPM-AIT-estimated gap samples. (d) Available samples and MIAA-AIT-estimated gap samples.

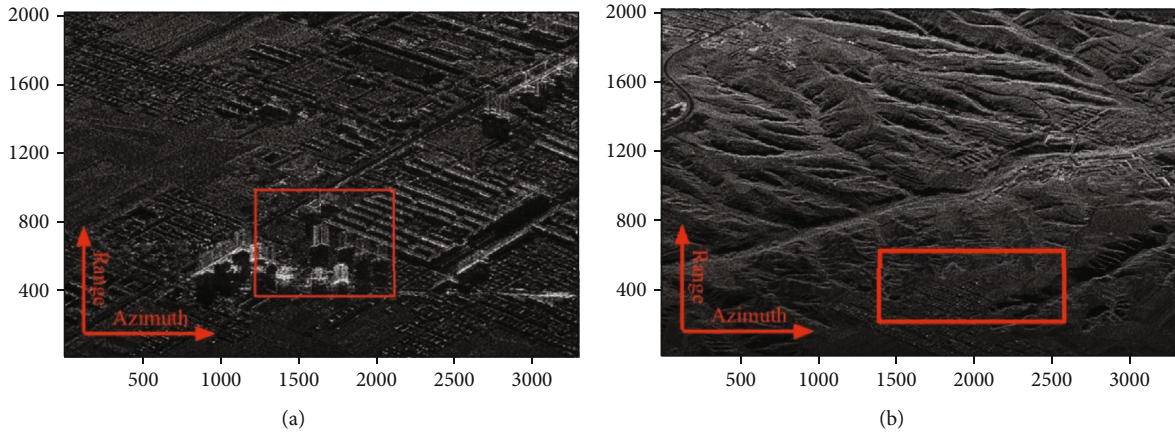


FIGURE 7: Two representative scenes imaged by the proposed algorithm. (a) Urban areas. (b) Rural areas.

Once the estimated spectrum $\hat{F}_{b1}(\omega_k)$ has been computed, the missing samples in the time domain can be estimated by an MMSE estimator as shown in (8), which is the second step of the MIAA [29].

$$\hat{\mathcal{S}}_{g1} = \sum_{k=0}^{K-1} [|F_{b1}(\omega_k)|^2 \mathbf{a}_b^H(\omega_k) \mathbf{R}^{-1} \mathbf{S}_b] \mathbf{a}_g(\omega_k) \# \quad (8)$$

Note that $\hat{F}_{b1}(\omega_k)$ can be obtained from (6), and $\mathbf{a}_b^H(\omega_k) \mathbf{R}^{-1} \mathbf{S}_b$ is the numerator on the right side of (6), which suggests that the computation cost of the missing data estimation after the IAA is very low. The whole missing samples estimation processing is parameter free. Similarly, the processing of obtaining $\hat{\mathcal{S}}_{g2}$ with \mathbf{S}_{burst2} is similar to obtaining

$\hat{\mathcal{S}}_{g1}$ with \mathbf{S}_{burst1} in the same way. Then, the missing samples are recovered by the weighted average of the $\hat{\mathcal{S}}_{g1}$ and $\hat{\mathcal{S}}_{g2}$ as shown in Figure 4.

3.3. Postprocessing. The inverse operation of dechirp processing in step 3.1 is applied to recover the subaperture signal format corresponding to the standard ScanSAR mode. Figure 4 shows how to use the proposed MIAA-AIT to recover the gaps between the bursts. After all the gaps are recovered, subapertures are stitched together along the azimuth direction to obtain nongapped full-aperture data. Finally, azimuth compression is completed for nongapped full-aperture data, and the enhanced ScanSAR image can be obtained as shown in Figure 3.

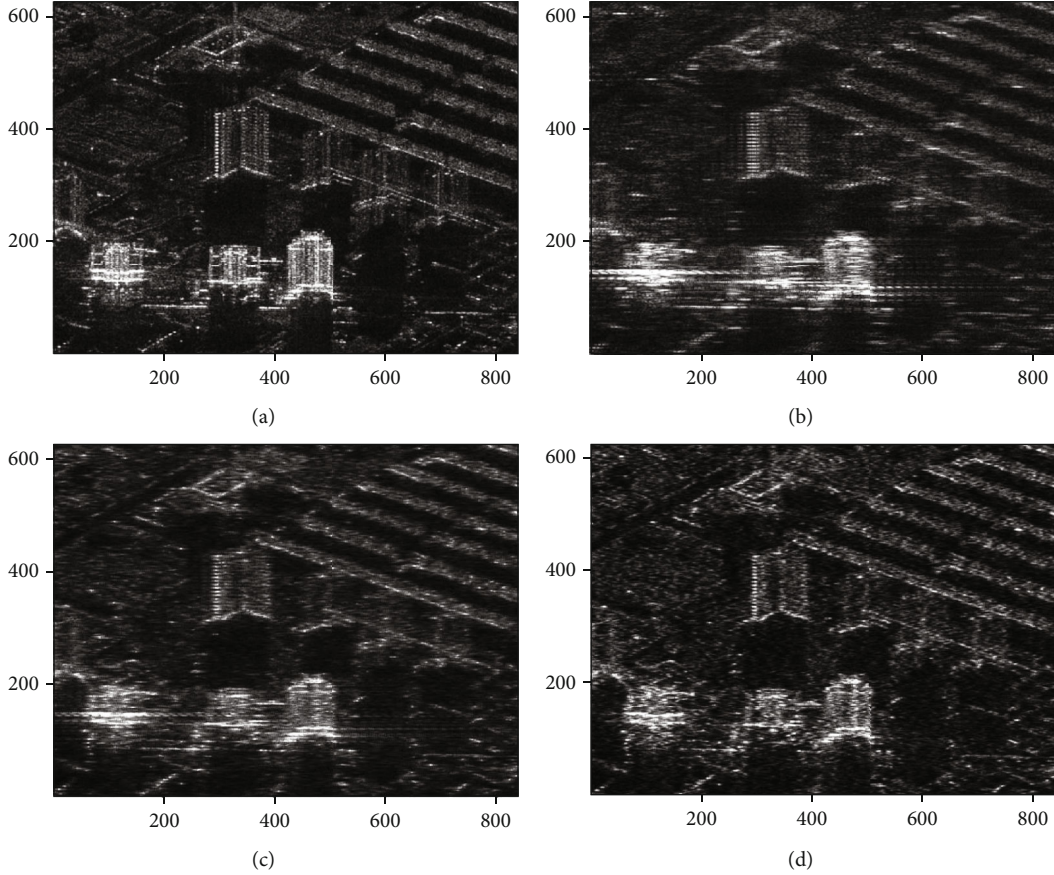


FIGURE 8: Magnified area shown in Figure 7(a). (a) Stripmap mode. (b) ScanSAR data imaged by the full-aperture approach. (c) ScanSAR data imaged by the LPM-AIT. (d) ScanSAR data imaged by the MIAA-AIT.

4. Result

4.1. Simulation Results. Here, one-dimensional (1-D) simulation results are presented to demonstrate the performance of the proposed MIAA-AIT in two aspects: missing data recovery and spike suppression.

4.1.1. Accuracy of Missing Sample Estimation. The signal we consider consists of seven complex-valued sinusoidal components located at normalized frequencies $f_1 = 0.1$, $f_2 = 0.17$, $f_3 = 0.19$, $f_4 = 0.2$, $f_5 = 0.23$, $f_6 = 0.24$, and $f_7 = 0.3$ with complex amplitudes $A_1 = 0.8$, $A_2 = 0.5$, $A_3 = A_4 = 1$, $A_5 = A_6 = 0.3$, and $A_7 = 1$. The phases of this seven targets are set as $\alpha_1 = 0.5$, $\alpha_2 = 0.3$, $\alpha_3 = 0.8$, $\alpha_4 = 0.5$, $\alpha_5 = 0.6$, $\alpha_6 = 0$, and $\alpha_7 = 0.8$. The signal is corrupted by Gaussian white noise with a zero mean and a standard deviation of 0.1. We first illustrate the missing-data recovery ability of the proposed algorithm for the case of an 80% missing-data proportion, which represents five subswaths. The burst length N_B is 100 in the simulations, and therefore, the number of missing data points between two bursts in a subaperture is 400. The recovery result of the 400 missing samples for one random trial among 100 Monte Carlo trials is presented in Figure 5(a). It can be concluded from Figure 5(a) that all the missing samples can be estimated with high accuracy by the proposed MIAA-AIT algorithm. With the LPM-AIT, however, only

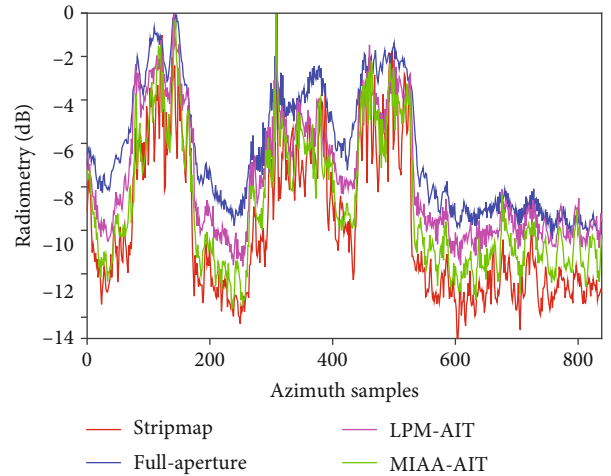


FIGURE 9: Radiometry comparison result of MIAA-AIT, LPM-AIT, full-aperture, and Stripmap over Figure 8.

missing samples that are close to the available data can be estimated with a relatively high accuracy, while the estimates in the middle exhibit a large deviation.

Next, we fix the burst length $N_B = 100$ and change the missing-data proportion from 50% to 83.3%, which represents 2 to 6 subswaths. The standard deviation of the

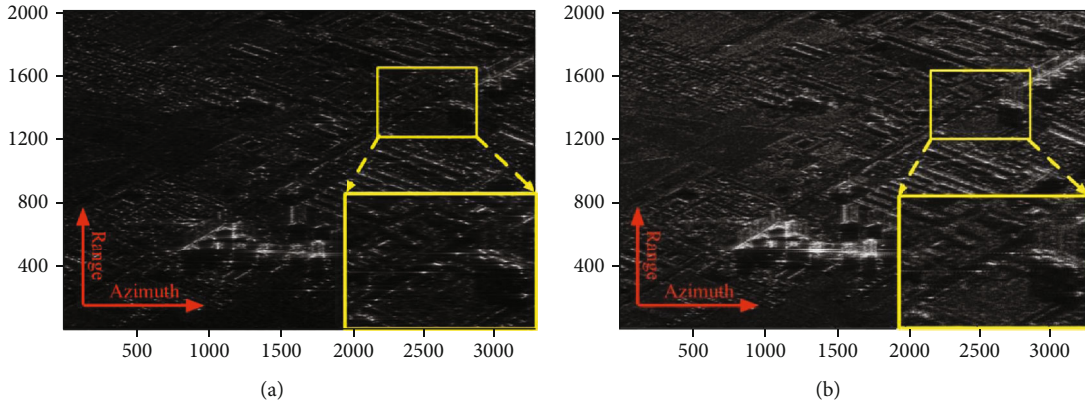


FIGURE 10: (a) Imaging result using only the data recovered by the LPM-AIT. (b) Imaging result using only the data recovered by the proposed algorithm.

Gaussian white noise is fixed at 0.1. For each case, 100 Monte Carlo trials are performed. The average mean squared error (AMSE) of the MIAA-AIT and the LPM-AIT for different missing-data proportions are shown in Figure 5(b). It can be concluded that the effect of the proposed algorithm is better than the LPM-AIT as the missing-data proportion increases. And the effect of the LPM-AIT deteriorates significantly when the missing-data proportion exceeds 50%. The average AMSE of MIAA-AIT is -20.5185 dB. The average AMSE of LPM-AIT is -7.6171 dB, which is 12.9014 dB higher than that of MIAA-AIT. These results demonstrate that the MIAA-AIT can predict missing data more accurately than the LPM-AIT. As a result, the proposed algorithm is more practical in the ScanSAR mode.

4.1.2. Performance in Spike Suppression. The effect of suppressing spikes caused by coherently processed multiple bursts is evaluated in this section. In this simulation, ten subapertures are used and the signal and noise settings are the same as that in part 4.1.1. The focused amplitude results of the complete samples, the available samples, the available samples with LPM-AIT-estimated samples, and the available samples with MIAA-AIT-estimated samples are shown in Figures 6(a)–6(d), respectively. The result of processing with gapped samples shows obvious spikes, as shown in Figure 6(b). In comparison, the LPM-AIT has some effect in alleviating the spikes, however, spikes still exist as shown in Figure 6(c). In Figure 6(d), spikes are almost completely suppressed, and the amplitudes are more accurate. Taking the targets amplitude estimation results with complete samples (Figure 6(a)) as reference, we calculate the difference between the other three estimation results and the reference, respectively. The sum of the difference between the amplitude estimation result from available samples (Figure 6(b)) and the reference is 72.898. The sum of the difference between the amplitude estimation result from LPM-AIT-estimated samples (Figure 6(c)) and the reference is 14.027. The sum of the difference between the amplitude estimation result from MIAA-estimated samples (Figure 6(d)) with the reference is 1.472. It can be seen from the numerical comparison that the amplitude estimated results by MIAA-AIT are

more similar to the reference, which means the MIAA-AIT is more effective than the LPM-AIT. It is obvious that the MIAA-AIT has a better performance than the LPM-AIT in spike suppression.

4.2. Experimental Results. To demonstrate the performance of the proposed algorithm in spike suppression, experimental results on real SAR data are provided. As described in Section 2 and Figure 2(b), in the range direction, the received raw signal of ScanSAR are the same as those in the Stripmap mode; in the azimuth direction, the raw ScanSAR data are equivalent to the periodically truncated Stripmap mode data. Therefore, obtaining raw data of the ScanSAR mode is equivalent to introducing periodic gaps into raw stripmap data. In addition, to verify the effectiveness of the proposed algorithm in the real SAR scene, we need to utilize the imaging result of the Stripmap mode as a reference. Thus, in our experiment, we obtain ScanSAR mode raw data by introducing periodic gaps into raw stripmap data.

The raw data were collected with an experimental interferometric airborne SAR system developed by the Institute of Electronics, Chinese Academy of Science. The system operates at the C-band (5.4 GHz); the PRF is 750 Hz; the average flight speed is 68 m/s, and the size of the antenna is 0.9 m. First, the spike suppression performance of the proposed algorithm is demonstrated. Then, interferograms acquired with the MIAA-AIT are given.

4.2.1. The Results of Spike Suppression. In the experiment, the ScanSAR system has five subswaths with a missing-data proportion of 80%, which is generally used in the ScanSAR system. The length of the synthetic aperture time is about 5000 samples, the length of one burst is set as 300 samples, and the length of the burst cycle is 1500 samples. Two representative scenes, shown in Figure 7, are selected to demonstrate the advantages of the proposed algorithm. Both two scenes are corresponding to 11 bursts with an area of 1485 meters in azimuth direction and 1100 meters in range direction. The pixel spacing in the azimuth is about one-fifth of the pixel spacing in range; thus, the images are multilooked 5 times in azimuth when displaying. The overall ScanSAR images are well focused, and no obvious spike phenomena exist.

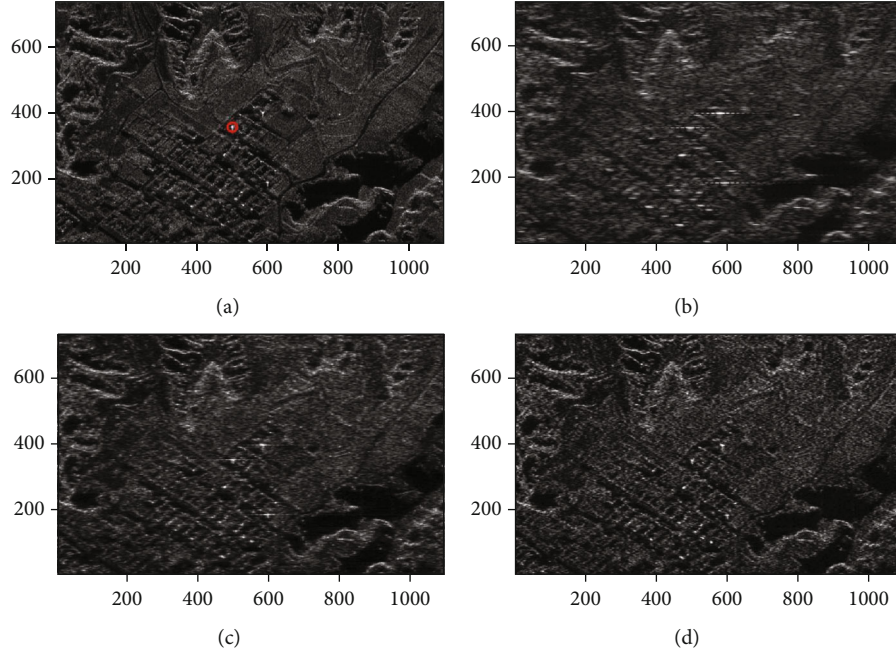


FIGURE 11: Magnified area shown in Figure 7(b). (a) Stripmap mode. (b) ScanSAR data imaged by a full-aperture approach. (c) ScanSAR data imaged by the LPM-AIT. (d) ScanSAR data imaged by the MIAA-AIT.

The area shown in Figure 7(a) is primarily composed of man-made structures, representing an urban area. The local scene highlighted in Figure 7(a) by a red rectangle is magnified for a comparison with the other algorithms as shown in Figure 8. Based on a visual comparison of Figures 8(b)–8(d), the LPM-AIT can reduce the spikes to some extent compared to the full-aperture approach. The proposed algorithm has a better performance in spike suppression. The outlines of the artificial buildings are clearer in Figure 8(d) because the spikes are suppressed more thoroughly by the proposed algorithm.

To illustrate the advantage of the proposed method in radiometry performance, we integrate pixels of the amplitude image of Figures 8(a)–8(d) across range direction as shown in Figure 9. In the area of buildings where most of the scatterers are bright scatterers, the performance of both MIAA-AIT and LPM-AIT is close to that of the Stripmap and the performance of the MIAA-AIT is slightly better than the LPM-AIT. This shows that both MIAA-AIT and LPM-AIT have a good processing effect on bright scatterers, but MIAA-AIT is better than LPM-AIT. In the area near the buildings, where few strong scatterers represent, the difference in radiometry performance is relatively large. The radiometry performance in these areas reflects the spike suppression performance of the algorithms. The lower the value, the better the spike suppression performance. Overall, the value of MIAA-AIT is lower than that of the LPM-AIT and closer to the value of the Stripmap. Thus, the MIAA-AIT has better suppression performance than the LPM-AIT. This further demonstrates the image quality improvement of the proposed algorithm.

We want to verify whether this improvement comes from the improvement of the quality of the recovered missing data by the proposed algorithm compared to the LPM-AIT. By

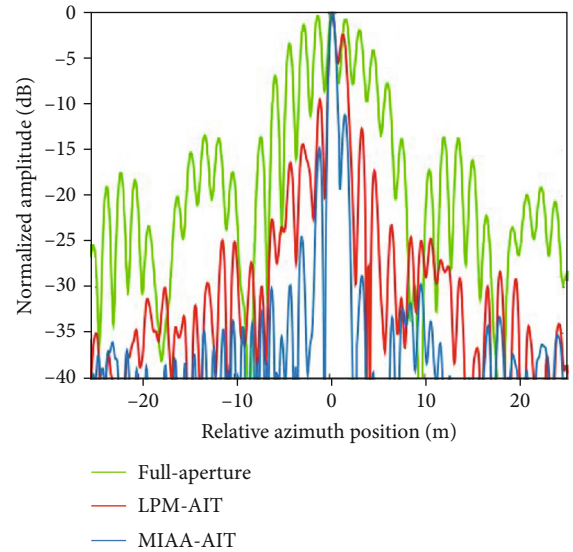


FIGURE 12: Azimuth profiles of the selected point-like target in Figure 10(a).

TABLE 1: IC comparison of two representative scenes (Figures 8 and 10).

	IC value of Figure 8	IC value of Figure 10
Full-aperture	2.2061	1.0281
LPM-AIT	2.8176	1.2505
MIAA-AIT	3.1959	1.5731

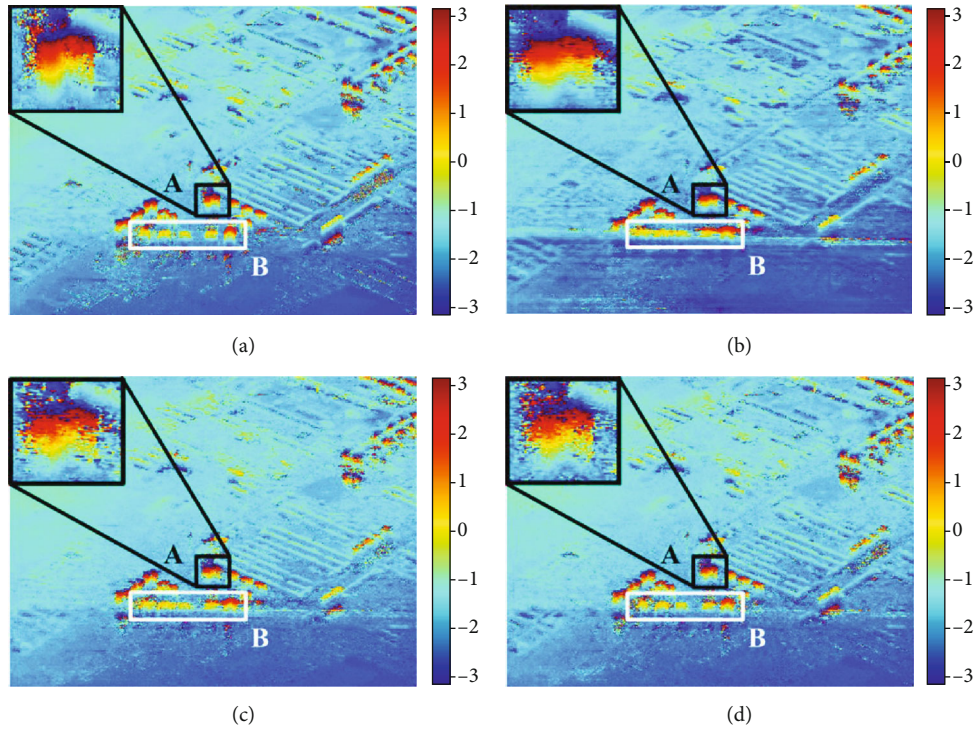


FIGURE 13: Interferograms of Scene in Figure 7(a). (a) Stripmap mode. (b) ScanSAR data processed by full-aperture approach. (c) ScanSAR data processed by the IPM-AIT. (d) ScanSAR data processed by the MIAA-AIT.

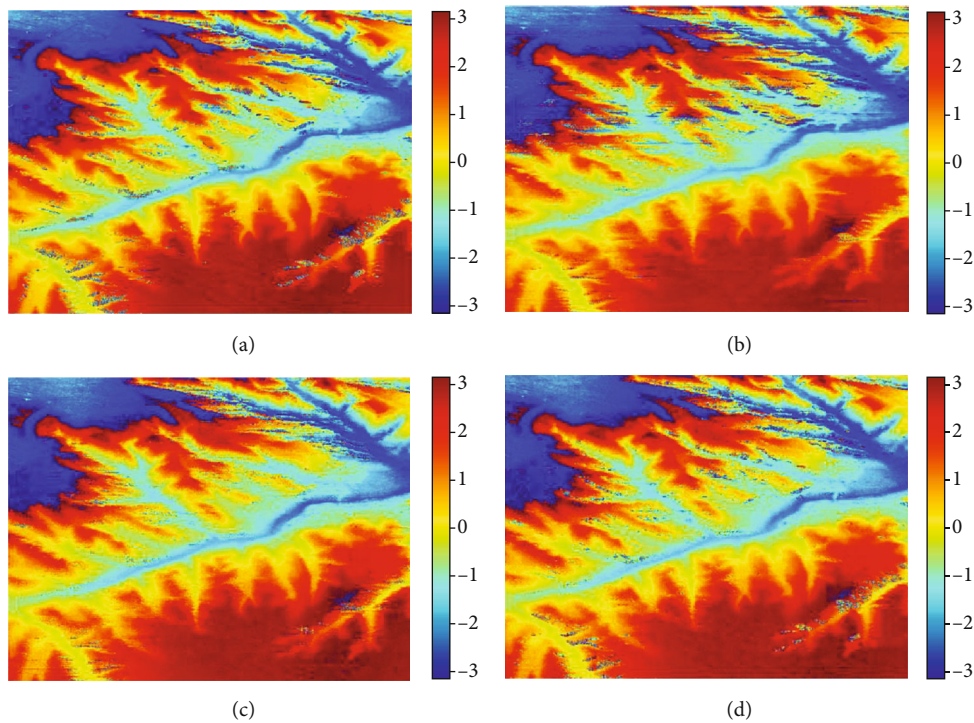


FIGURE 14: Interferograms of Scene in Figure 7(b). (a) Stripmap mode. (b) ScanSAR data processed by a full-aperture approach. (c) ScanSAR data processed by the LPM-AIT. (d) ScanSAR data processed by the MIAA-AIT.

simulation experiments, as shown in Figure 5, we have shown that the missing data recovered by the proposed algorithm has better quality than the missing data recovered by the

LPM-AIT. Due to the complexity of real SAR data, it is meaningless to directly compare the recovered data. Therefore, we can verify the effectiveness of the recovered data by

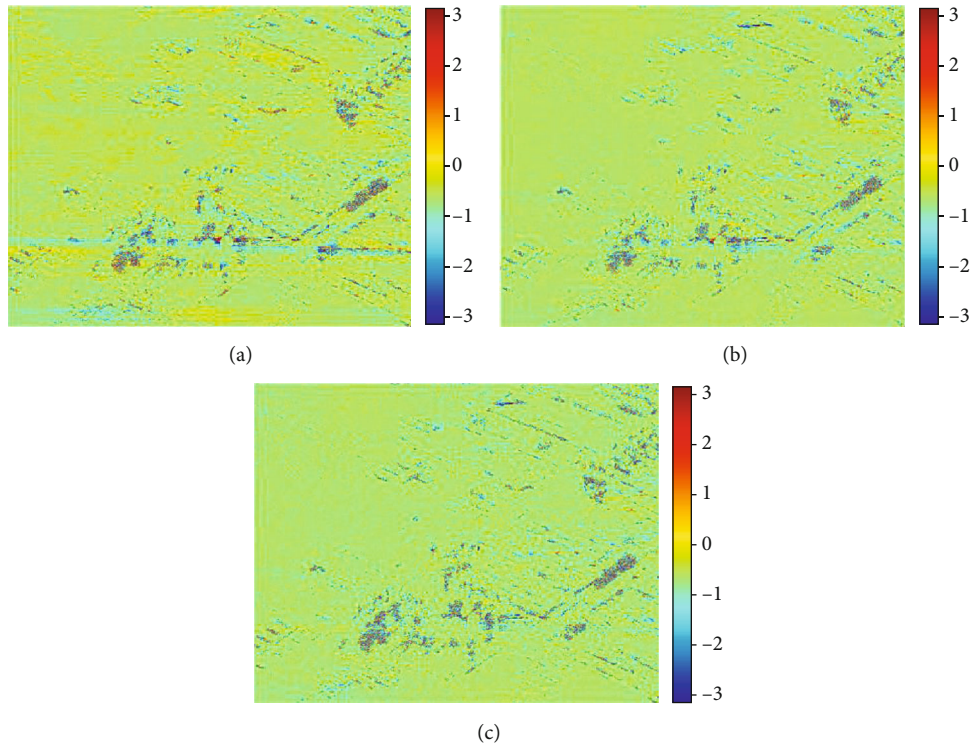


FIGURE 15: Difference maps of interferograms in Figure 13 processed by full-aperture, LPM-AIT, and MIAA-AIT, against Stripmap mode. (a) full-aperture against Stripmap; (b) LPM-AIT against Stripmap; (c) MIAA-AIT against Stripmap.

comparing the imaging results using only the recovered data. To compare the validity of the recovered data, we replace the available data (the burst) with zero and complete the imaging process using only the predicted data by LPM-AIT and MIAA-AIT method, respectively. The imaging results are given in Figure 10. Obviously, the MIAA-AIT retains more details of the scene. The local scene marked by the yellow rectangle is enlarged and displayed in the lower right corner. We can see that in Figure 10(b) many relatively weak targets such as flat ground and low-rise houses are clearer than that in Figure 10(a). It shows that the signals of these relatively weak targets are not effectively recovered by the LPM-AIT, but can be better recovered by the proposed algorithm. More useful information can be remained in the data predicted by the proposed algorithm, so the final image processed by the proposed algorithm is better than that processed by the LPM-AIT.

In addition to an urban area, the results of the proposed algorithm applied to a rural area are presented in Figure 7(b). The results of the local scene marked by the red rectangle in Figure 7(b) are shown in Figure 11. The proposed algorithm also has a better performance in suppressing spikes in rural areas. Compared with Figure 11(b) and 11(c), the shadows are more properly defined, and the contours of the mountains are more distinguishable in Figure 11(d). The spikes neighboring the strong scatterers are almost all eliminated in Figure 11(d). Azimuth profiles of the point-like target marked by a red circle in Figure 11(a) are plotted in Figure 12, where a green line, a red line, and a blue line represent the azimuth profiles using the full-aperture algo-

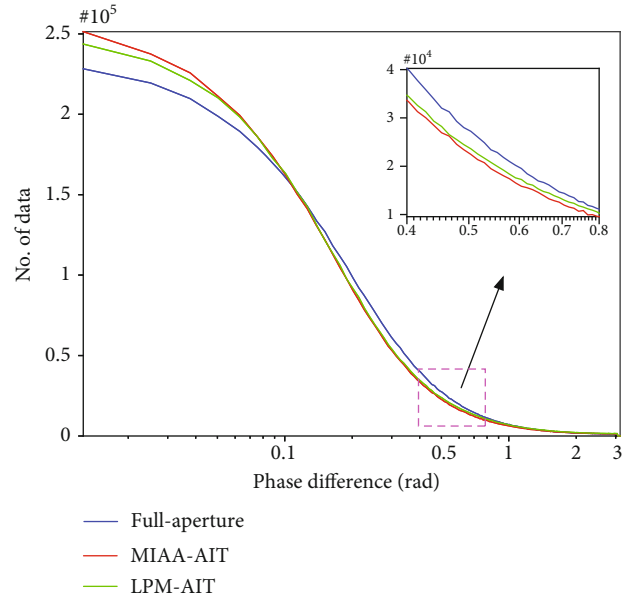


FIGURE 16: Statistical curve of the absolute value of the phase difference shown in Figure 15.

rithm, the LPM-AIT, and the MIAA-AIT, respectively. It can be seen that spikes are more thoroughly eliminated by the proposed algorithm, and the azimuth resolution of the target significantly increased compared with the result processed by the full-aperture approach.

TABLE 2: Statistical results of difference maps of interferograms processed by full-aperture, LPM-AIT, and MIAA-AIT, against Stripmap mode. (a) Interferograms in Figure 12. (b) Interferograms in Figure 13.

(a)			
	Phase diff. < 0.1 rad	0.1 rad ≤ phase diff. ≤ 0.5 rad	Phase diff. > 0.5 rad
Full-aperture	49.49%	35.84%	14.67%
LPM-AIT	52.37%	33.29%	14.34%
MIAA-AIT	55.08%	31.83%	13.09%

(b)			
	Phase diff. < 0.1 rad	0.1 rad ≤ phase diff. ≤ 0.5 rad	Phase diff. > 0.5 rad
Full-aperture	43.44%	41.56%	15.00%
LPM-AIT	44.83%	42.41%	12.76%
MIAA-AIT	45.96%	44.24%	9.80%

Image contrast (IC) value can be used to evaluate the quality of the SAR image [32]. Here, we use the standard deviation amplitude contrast function in reference [32] to calculate the IC value of each image as shown in (9).

$$IC = \frac{\sqrt{E\{(A - E\{A\})^2\}}}{E\{A\}} \#, \quad (9)$$

where A is the amplitude of the image, $E\{\bullet\}$ means averaging operation, thus $E\{A\}$ is the mean of the image amplitude. Table 1 shows the IC value comparison of the imaging results by three different methods on the urban scene shown in Figure 8, and the nonurban scene shown in Figure 11. The IC value of the image processed by the proposed algorithm is higher than that processed by the LPM-AIT in both two scenes, which means the image quality obtained by the proposed method is better. This further illustrates the effectiveness of the processed algorithm.

4.2.2. The Results of Interferometry Application. Section 4.2.1 has thoroughly demonstrated the spike suppression ability of the proposed algorithm over the ScanSAR images. Interferometric application can also benefit from the remarkable spike suppression performance to improve the ScanSAR interferogram quality. In this section, interferograms obtained using the proposed algorithm are presented to show the ability of the method in interferometry application. Interferograms of the two representative scenes in Figure 7 processed using the Stripmap mode, the full-aperture approach, the LPM-AIT, and the proposed approach are shown in Figures 13 and 14.

Interferograms are contaminated due to phase information in the spikes when they are processed with only the gapped data, as shown in Figure 13(b). Sites A and B in Figure 13 are buildings. The shape of the building at site A processed by the MIAA-AIT is narrower than that processed by the LPM-AIT and more similar to that processed by the Stripmap mode. In site B, although it has been improved compared to Figure 13(b), the outlines of the buildings proc-

essed by the LPM-AIT is blurred due to the phase information in spikes. The outlines of the buildings processed by the MIAA-AIT are clearer than the LPM-AIT and more similar to that processed by the Stripmap mode. Figure 15 shows three-phase difference maps of interferograms in Figure 13 processed by full-aperture, LPM-AIT, and MIAA-AIT, against Stripmap mode. It can be seen that on both sides of buildings, the phase difference of MIAA-AIT is much smaller than the LPM-AIT. Figure 16 shows the statistical curve of the absolute value of the phase difference. To show the results more clearly, the phase difference is displayed in logarithmic format. And the display range is only to π , because the differences are almost within π . Table 2(a) shows the numerical statistics results. It can be seen that the point number of MIAA-AIT with very small phase difference are larger than that of the LPM-AIT and full-aperture. And the point number of MIAA-AIT with large phase difference is smaller than that of the full-aperture and LPM-AIT. Therefore, we can conclude the MIAA-AIT has better spike performance in interferometric application in urban areas.

In the rural scene, the interferogram processed by the proposed algorithm also has good performance in spike suppression, as shown in Figure 14. The interferogram processed by the full-aperture approach exists spikes, introducing additional noise to the interferogram. It can be seen that the spike phenomenon in the interferogram processed by the MIAA-AIT is less than that of LPM-AIT, but it is not so obvious. The reason is that there are fewer bright points in the rural area than in the urban area. Figure 17 shows three-phase difference maps of interferograms in Figure 14 processed by full-aperture, LPM-AIT, and MIAA-AIT, against Stripmap mode. The corresponding statistical results are shown in Figure 18. Same as Figure 16, the phase difference is displayed in logarithmic format and the display range is only to π . Table 2(b) shows the numerical statistics. In the range where the phase difference is close to 0, the point number of MIAA-AIT is much higher than the number of full-aperture and LPM-AIT. And the point number of the MIAA-AIT with large phase difference is smaller than that of full-aperture and LPM-AIT. Therefore, it can be concluded that the effect

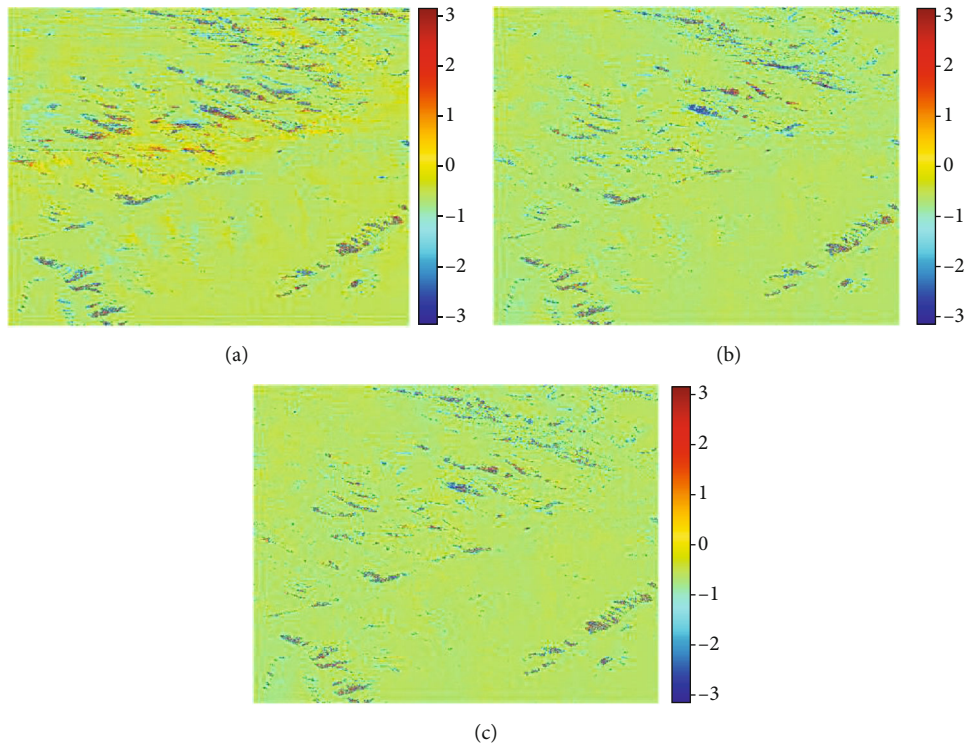


FIGURE 17: Difference maps of interferograms in Figure 14 processed by full-aperture, LPM-AIT, and MIAA-AIT, against Stripmap mode. (a) Full-aperture against Stripmap; (b) LPM-AIT against Stripmap; (c) MIAA-AIT against Stripmap.

of MIAA-AIT in rural area is still better than that of LPM-AIT, but compared with urban areas, the advantages are not so prominent.

Based on the above analysis, the proposed algorithm can be used in interferometry application to improve interferograms in the ScanSAR mode.

5. Conclusions

This manuscript studies the spike contamination problems of ScanSAR mode data. When ScanSAR data are processed using the full-aperture approach, images are contaminated by spikes, and additional noise is introduced to the interferograms. Spike contamination of full-aperture ScanSAR image limits ScanSAR-related applications, such as target detection and interferometry. To solve this problem, an improved algorithm named the MIAA-AIT is proposed in this manuscript. In the MIAA-AIT, the gaps between bursts are recovered by the MIAA, which is a nonparametric missing-data recovery methodology. Compared with the existing LPM based algorithm, the proposed algorithm has a better performance on spike suppression and can improve the azimuth resolution of the full-aperture ScanSAR images. Experiment results show that more useful information can be remained in the data predicted by the proposed algorithm, therefore the final image processed by the proposed algorithm is better than that processed by the existing LPM based algorithm. The proposed MIAA-AIT algorithm has better applicability than the LPM-AIT for spike suppression and azimuth resolution enhancement when processing

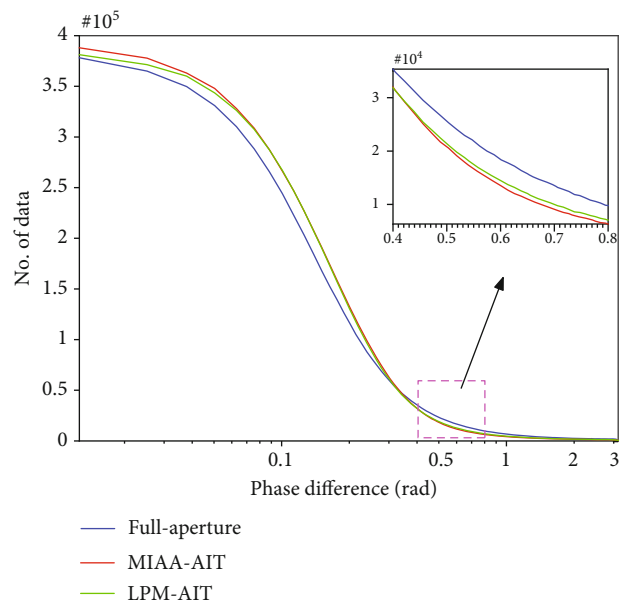


FIGURE 18: Statistical curve of the absolute value of the phase difference shown in Figure 16.

ScanSAR data by the full-aperture imaging algorithm. The proposed algorithm can also be used in interferometry to improve the quality of interferograms.

The results of this study show that by infilling the gaps between bursts with valid samples, the ScanSAR images processed by full-aperture approach can be significantly

improved. The spikes can be suppressed effectively, and the azimuth resolution can be improved at the same time. This is very beneficial to ScanSAR-related applications. Therefore, it is worth further study.

Data Availability

The raw SAR data used to support the findings of this study have not been made available because the data used in this article is owned by the Department of Space Microwave Remote Sensing System, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive comments on an earlier version of this paper. This research was funded by the National Key R&D Program of China grant number 2017YFB0502700.

References

- [1] Y. Deng, F. Zhao, and R. Wang, "Brief analysis on the development and application of spaceborne SAR," *Journal of Radars*, vol. 17, pp. 1–10, 2012.
- [2] Y. Deng and R. Wang, "Exploration of advanced bistatic SAR experiments," *Journal of Radars*, vol. 3, pp. 1–9, 2014.
- [3] K. Tomiyasu, "Conceptual performance of a satellite borne, wide swath synthetic aperture radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-19, no. 2, pp. 108–116, 1981.
- [4] R. K. Moore, J. P. Claassen, and Y. H. Lin, "Scanning spaceborne synthetic aperture radar with integrated radiometer," *IEEE Transactions on Aerospace and Electronic Systems*, vol. -AES-17, no. 3, pp. 410–421, 1981.
- [5] I. G. Cumming, Y. Guo, and F. Wong, *Analysis and precision processing of RADARSAT ScanSAR data*, Proc. Geomatics in the Era of RADARSAT; Publishing House, Ottawa, Canada, 1997.
- [6] A. M. Guarnieri and C. Prati, "ScanSAR focusing and interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 4, pp. 1029–1038, 1996.
- [7] J. Holzner and R. Bamler, "Burst-mode and ScanSAR interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 1917–1934, 2002.
- [8] J. Ning, R. Wang, Y. Deng, N. Li, H. Song, and W. Fei, "AET-IAA: a novel approach to enhance azimuth resolution for ScanSAR interferometry," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 12, pp. 5674–5685, 2017.
- [9] A. Moreira, J. Mittermayer, and R. Scheiber, "Extended chirp scaling algorithm for air- and spaceborne SAR data processing in stripmap and ScanSAR imaging modes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 5, pp. 1123–1136, 1996.
- [10] R. Bamler and M. Eineder, "ScanSAR processing using standard high precision SAR algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 1, pp. 212–218, 1996.
- [11] C. Liang, Q. Zeng, and J. Jiao, "An assessment of ScanSAR interferometric processing using full-aperture approach," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 9, pp. 1559–1563, 2014.
- [12] I. G. Cumming and F. H. Wong, *Digital Processing of Synthetic Aperture Radar Data: Algorithms and Implementation*, Artech House, Norwood, MA, USA, 2005.
- [13] C. Liang, Z. Liu, E. J. Fielding, and R. Bürgmann, "InSAR time series analysis of L-Band wide-swath SAR data acquired by ALOS-2," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4492–4506, 2018.
- [14] N. Cao, H. Lee, H. C. Jung, and H. Yu, "Estimation of water level changes of large-scale Amazon wetlands using ALOS-2 ScanSAR differential interferometry," *Remote Sensing*, vol. 10, no. 6, pp. 966–981, 2018.
- [15] G. Gomba, F. R. González, and F. De Zan, "Ionospheric phase screen compensation for the Sentinel-1 TOPS and ALOS-2 ScanSAR modes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 223–235, 2017.
- [16] A. Pepe, A. B. Ortiz, P. R. Lundgren, P. A. Rosen, and R. Lanari, "The stripmap-ScanSAR SBAS approach to fill gaps in Stripmap deformation Time Series with ScanSAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 12, pp. 4788–4804, 2011.
- [17] R. Natsuaki, H. Nagai, T. Motohka et al., "SAR interferometry using ALOS-2 PALSAR-2 data for the Mw 7.8 Gorkha, Nepal earthquake," *Earth, Planets and Space*, vol. 68, no. 1, pp. 15–27, 2016.
- [18] J. Horstmann, W. Koch, S. Lehner, and R. Tonboe, "Ocean winds from RADARSAT-1 ScanSAR," *Canadian Journal of Remote Sensing*, vol. 28, no. 3, pp. 524–533, 2014.
- [19] M. Santoro, C. Beer, O. Cartus et al., "Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements," *Remote Sensing of Environment*, vol. 115, no. 2, pp. 490–507, 2011.
- [20] A. S. Arnesena, T. S. F. Silva, L. L. Hess et al., "Monitoring flood extent in the lower Amazon River floodplain using ALOS/PALSAR ScanSAR images," *Remote Sensing of Environment*, vol. 130, pp. 51–61, 2013.
- [21] B. Rabus, M. Eineder, A. Roth, and R. Bamler, "The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 57, no. 4, pp. 241–262, 2003.
- [22] N. Li, R. Wang, Y. Deng et al., "Improved full-aperture ScanSAR imaging algorithm based on aperture interpolation," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 5, pp. 1101–1105, 2015.
- [23] I. G. Cumming, Y. Guo, and F. Wong, "A comparison of phase-preserving algorithms for burst-mode SAR data processing," in *GARSS'97. 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing - A Scientific Vision for Sustainable Development*, pp. 731–733, Singapore, 1997.
- [24] T. G. Moore, B. W. Zuerndorfer, and E. C. Burt, "Enhanced imagery using spectral-estimation-based techniques," *Lincoln Laboratory Journal*, vol. 10, pp. 171–186, 1997.

- [25] J. Salzman, D. Akamine, R. Lefevre, and J. C. Kirk, "Interrupted synthetic aperture radar (SAR)," *IEEE Aerospace and Electronic Systems Magazine*, vol. 17, no. 5, pp. 33–39, 2002.
- [26] C. Liang and E. J. Fielding, "Interferometric processing of ScanSAR data using stripmap processor: new insights from coregistration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 4343–4354, 2016.
- [27] R. Natsuaki, T. Motohka, M. Shimada, and S. Suzuki, "Burst misalignment evaluation for ALOS-2 PALSAR-2 ScanSAR-ScanSAR interferometry," *Remote Sensing*, vol. 9, no. 3, pp. 216–224, 2017.
- [28] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source localization and sensing: a nonparametric iterative adaptive approach based on weighted least squares," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, pp. 425–443, 2010.
- [29] P. Stoica, J. Li, and J. Ling, "Missing data recovery via a nonparametric iterative adaptive approach," *IEEE Signal Processing Letters*, vol. 16, no. 4, pp. 241–244, 2009.
- [30] M. Xue, L. Xu, and J. Li, "IAA spectral estimation: fast implementation using the Gohberg-Semencul factorization," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3251–3261, 2011.
- [31] G.-O. Glentis and A. Jakobsson, "Efficient implementation of iterative adaptive approach spectral estimation techniques," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4154–4167, 2011.
- [32] F. Berizzi and G. Corsini, "Autofocusing of inverse synthetic aperture radar images using contrast optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 3, pp. 1185–1191, 1996.

Research Article

RS and GIS Supported Urban LULC and UHI Change Simulation and Assessment

Pei Liu,^{1,2} Shoujun Jia,³ Ruimei Han ,² Yuanping Liu,⁴ Xiaofeng Lu,^{1,2} and Hanwei Zhang¹

¹Key Laboratory of Spatio-temporal Information and Ecological Restoration of Mines (MNR), Henan Polytechnic University, Jiaozuo, Henan 454003, China

²School of Surveying and Mapping Land Information Engineering, Henan Polytechnic University, Jiaozuo, 454003 Henan, China

³College of Surveying and Geo-informatics, Tongji University, Shanghai 200092, China

⁴Collaborative Innovation Center of Aerospace Remote Sensing Information Processing and Application of Hebei Province, China

Correspondence should be addressed to Ruimei Han; hrm@hpu.edu.cn

Received 22 February 2020; Revised 26 May 2020; Accepted 15 June 2020; Published 7 July 2020

Academic Editor: Sang-Hoon Hong

Copyright © 2020 Pei Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rapid urbanization has become a major urban sustainability concern due to environmental impacts, such as the development of urban heat island (UHI) and the reduction of urban security states. To date, most research on urban sustainability development has focused on dynamic change monitoring or UHI state characterization, while there is little literature on UHI change analysis. In addition, there has been little research on the impact of land use and land cover changes (LULCCs) on UHI, especially simulates future trends of LULCCs, UHI change, and dynamic relationship of LULCCs and UHI. The purpose of this research is to design a remote sensing-based framework that investigates and analyzes how the LULCCs in the process of urbanization affected thermal environment. In order to assess and predict the impact of LULCCs on urban heat environment, multitemporal remotely sensed data from 1986 to 2016 were selected as source data, and Geographic Information System (GIS) methods such as the CA-Markov model were employed to construct the proposed framework. The results showed that (1) there has been a substantial strength of urban expansion during the 40-year study period, (2) the farthest distance urban center of gravity moves from north-northeast (NEE) to west-southwest (WSW) direction, (3) the dominate temperature was middle level, sub-high level, and high level in the research area, (4) there was a higher changing frequency and range from east to west, and (5) there was a significant negative correlation between land surface temperature and vegetation and significant positive correlation between temperature and human settlement.

1. Introduction

Land use and land cover changes (LULCCs), as one of the most significant processes related to earth ecological environment problems and social progress issues [1–5], related to global and regional changes [6, 7], have largely affected earth biochemical cycle [8, 9], sustainable use of resources [10], biodiversity [11], and urban planning and policymaking [12]. As a specific type of LULCCs, urban sprawl which plays an important role in urban intelligent growth and modernization is a sign of development and progress of human civilization and urbanization [13]. By prediction, more than

60% of the human population will live in cities by 2030 [14] which results in increasing replacement of natural landscape by the human-made landscape and will cause the temperature in the urban area to be higher than that in the suburban surrounding or rural area [15], which is also known as UHI. UHI phenomenon will not only accelerate urban environmental temperature and air pollution but also significantly increase energy consumption, increase urban temperature, and reduce quality of life. UHI is realized as a significant factor leading to global warming, heat-related mortality, and nonforecast climate change. A comprehensive study of the influencing factors of the UHI effect is critical for

formulating reasonable urban planning policies and mitigating the effects of UHI [16–18].

The urban sustainability development which can comprehensively consider the ecological environment and human environment at the natural, economic, and social levels in the process of urban growth, and reflect the integrity of human settlement environment and health status of the ecosystem as a whole, is regarded as the basis for regional urban environmental system governance and prevention policy formulation [19, 20]. The urban expansion was found to have a significant impact on local temperatures, in Chapman et al.'s review article which found that in some cases by up to 5 degrees [16]. UHI is closely associated with urban structure and will further increase by urban sprawl [21]. In the past decades, most researchers examined LULCCs and UHI in isolation, with few considering their combined effect.

Furthermore, existing satellite-based studies have typically evaluated and assessed the status of UHI at any given time but have limitations for studying the dynamic progress of LULCCs and UHI. Most research focused on evaluating the current or past status of LULCCs or UHI, while there has been little attempt to simulate or predict future change even though this information is crucial to inform effective sustainable urban development policy. Change simulation can provide valuable information for future prediction, as well as can indicate anthropogenic impact and identify degradation and deforestation which is useful for urban development planning. Various dynamic prediction models, including empirical-statistical models, optimization models, agent-based models, and Cellular Automata-Markov (CA-Markov) models, have been used for LULCC simulation [22], while they are seldom used for UHI change trend prediction.

In this research, we proposed a strategy to assess and analyze the impact of LULCCs on UHIs, as well as to simulate, predict, and explore the relationship and interaction between LULCCs and thermal environment trend in the future. For this purpose, we employed multitemporal remotely sensed data captured in 1986, 1996, 2006, and 2016, GIS spatial analysis methods, and CA-Markov trend simulation model. We test the method on Zhengzhou city, one of the fastest-growing metropolitan cities. Outputs are expected to contribute to urban planning, urban security assessment, and sustainable development in urban environments.

The rest part of the paper is organized as follows: a brief introduction about the research area, preliminary work, and preprocessing of datasets is given in Section 2. The proposed research framework based on remote sensing (RS) is drawn and applied to the research area in Section 3. The results are shown in Section 4; analysis and discussion are given in Section 5. Finally, the conclusions are drawn in Section 6.

2. Study Area and Datasets

The study site is located in Zhengzhou city (Figure 1), the capital of Henan Province in the central part of the P.R. China, with a total area of 7507 km² as well as a population of 9 878 000 inhabitants. Zhengzhou is one of the National Central Cities in China and serves as the political, economic, technological, and educational center of the province, as well

as a major transportation hub in China (<http://en.wikipedia.org/wiki/Zhengzhou>). The annual average temperature of the city is 14.5°C, and the general terrain trend is tilt from southwest to northeast. The study area is undergoing the accelerating of Chinese agglomeration, economic development, and urban expansion.

The primary dataset in this study is multitemporal Landsat remotely sensed data, with data covering the period from 1986 to 2016, which were acquired from USGS (<https://earthexplorer.usgs.gov/>). As a crucial instrument aboard Landsat satellites to collect imagery, the spatial resolution of TM/ETM+ is 30 m for the VIR-NIR band and 60 m/120 m for the TIR band. Landsat data have been widely used for LULCCs or UHI monitoring, while few of the research attempted to track and combine the long-term dynamic of LULCCs and UHI for environment security assessment. The second major dataset used is GlobelLand30, which we utilized as a reference source (<http://www.globallandcover.com>). The vector data, demographic data, and simulated data in the future are also collected for future stats analysis. A more detailed description about datasets, including captured date, sensor types, and resolution, can be found in Table 1. The data availability statement: all datasets used in this paper including the primary remotely sensed data, boundary vector data, and processed results can be obtained from hyperlink: https://pan.baidu.com/s/1d9GwkEpwryWYgHU_78qV5g, with password: c8rp. Researchers who are interested in this topic can download the data from the above hyperlink or contact the corresponding author to obtain source data to conduct secondary analysis.

The selected remotely sensed data were preprocessed to overcome geometric and atmospheric conditions, distortion, and errors through atmospheric and radiometric correction. To reduce the geometric distortion and radiometric difference, the selected remotely sensed data were preprocessed through geometric and radiometric correction. Specifically, following the previous works in [23], the radiometric calibration procedure is finished using the commercial software (ENVI). In addition, the FLAASH module of ENVI software was used to complete the atmospheric correction.

3. Methodology

The methodology employed in this research includes three main stages: (1) LULC and UST maps depicted from 1986 to 2026 with the help of SVM, MWA algorithms, and CA-Markov model; (2) analyzing and discussing of spatial distribution and temporal change of LULC and UST with respect to expansion intensity, buffer zone analysis, human settlement transfer, etc.; (3) quantitative and qualitative evaluation relationship between LULC and UST, more details can be found in Figure 2.

3.1. Land Use and Land Cover Classification. Land use and land cover (LULC) maps were retrieved from four-time nodes remotely sensed dataset over the metropolitan city of Zhengzhou. In this image classification processing, one of the most state-of-art machine learning algorithm, support vector machine (SVM), was selected to classify research area into five different categories (agricultural land, vegetation,

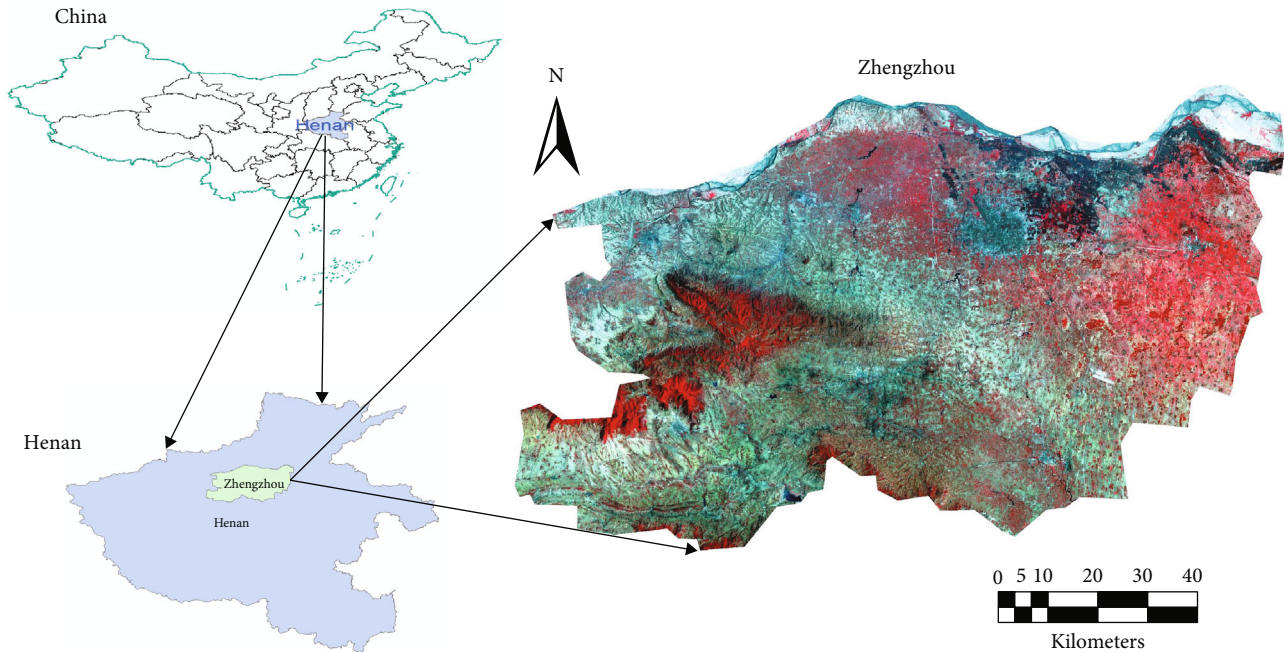


FIGURE 1: Location of Zhengzhou, Henan Province, China. Cropped image (1996, false-color composite Landsat TM image courtesy of the U.S. Geological Survey, <https://usgs.gov>) shows the full extent of the study area.

TABLE 1: Selected RS and GIS datasets.

Data type	Date	Satellite/sensor	Band	Resolution/scale/person	Source
RS data	1986.08.12	Landsat 5	1, 2, 3, 4, 5, 6, 7	30 m/120 m	https://www.usgs.gov
RS data	1996.04.15	Landsat 5	1, 2, 3, 4, 5, 6, 7	30 m/120 m	https://www.usgs.gov
RS data	2006.04.09	Landsat 5	1, 2, 3, 4, 5, 6, 7	30 m/120 m	https://www.usgs.gov
RS data	2016.04.23	Landsat 8	1, 2, 3, 4, 5, 7, 10, 11	30 m/100 m	https://www.usgs.gov
GlobeLand30	2010	HJ, BJ, Landsat	Classification map	30 m	http://www.globallandcover.com
Vector data	2018	GIS shape file	1	1:4 000 000	https://www.gadm.org

waterbody, bare land, and building land) according to the Classification Criteria for Land Use Status/GB-T21010-2015 and GlobeLand30 standard products [24]. The process for LULC classification using the SVM classifier includes training samples selection, SVM kernel determination, feature vectors inputs, and SVM classifier application. About 100 training samples of each class are selected by an expert in the field of RS; the most robust radial basis function (RBF) was selected as the SVM classifier:

$$K(x_i, x_j) = \exp(-\gamma|x_i - x_j|^2), \quad \gamma > 0, \quad (1)$$

where γ is the bias term in kernel function for polynomial and sigmoid kernels, and Landsat spectral bands except thermal infrared are selected as inputs.

Classification accuracy will affect subsequent change analysis; in this research, overall accuracy (OA) and kappa coefficient (KC) were chosen as the evaluation criterion. OA and KC are a nonparametric test which can reflect the consistency of labelled value and predicted value [12]. The mathematical model of OA and KC can be expressed as

$$OA = \frac{\sum_{i=1}^q n_{ii}}{n} * 100, \quad (2)$$

$$Kappa = \frac{n \sum_{i=1}^q n_{ii} - \sum_{i=1}^q n_{i+} n_{+i}}{n^2 - \sum_{i=1}^q n_{i+} n_{+i}}, \quad (3)$$

where q is the number of classes, n represents the total number of considered pixel, n_{ii} is the diagonal element of the confusion matrix, n_{i+} represents the marginal sum of the rows in the confusion matrix, and n_{+i} represents the marginal sum of the columns in the confusion matrix.

3.2. Retrieval of Urban Surface Temperature. By literature review, there are three algorithms, including single-channel algorithm, monowindow algorithms, and radiative transfer equation algorithm, which are widely used for urban surface temperature retrieval from a single-band Landsat thermal band [25]. In this research, Qin et al.'s monowindow algorithm (MWA) was selected for retrieval urban surface temperature from Landsat TM/ETM+ image. The mathematical model of MWA can be expressed as

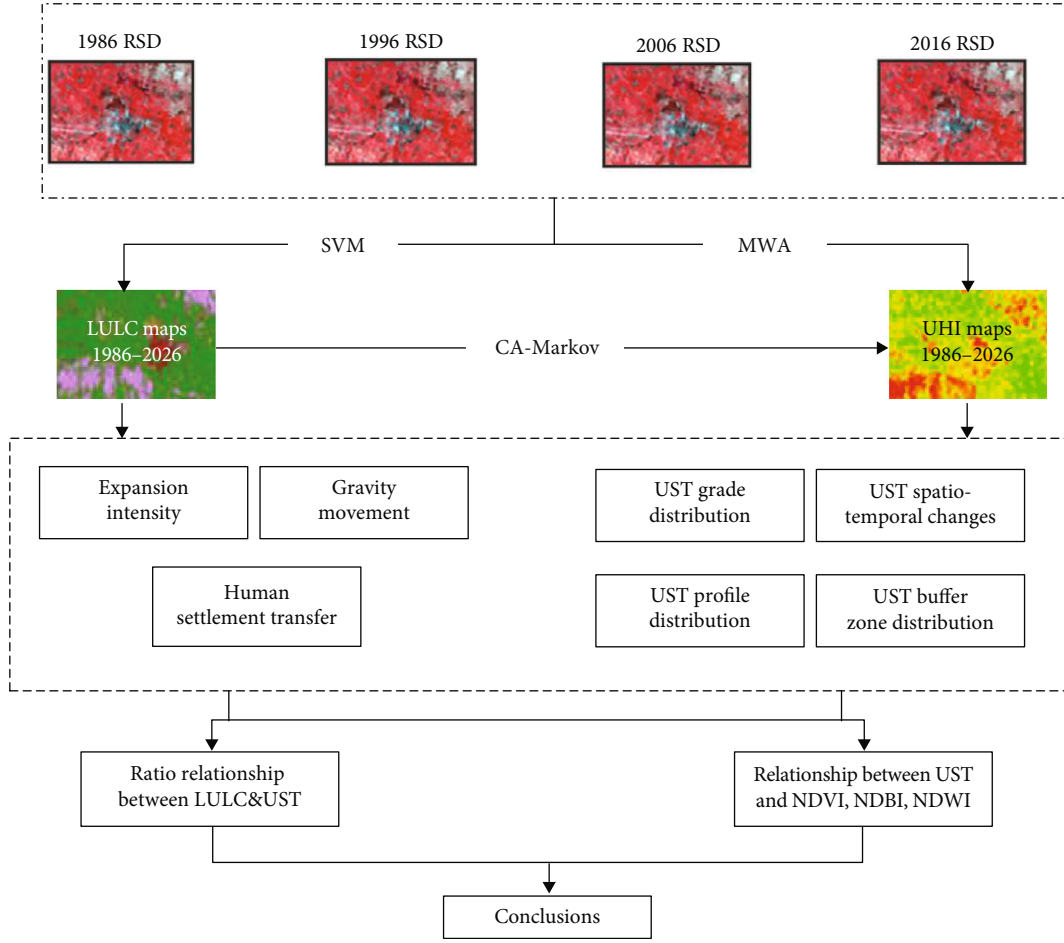


FIGURE 2: Flow chart.

$$T_s = \frac{a_6(1 - C_6 - D_6) + [b_6(1 - C_6 - D_6) + C_6 + D_6]T_{\text{sensor}} - D_6T_a}{C_6}, \quad (4)$$

$$C_6 = \varepsilon\tau_6, \quad (5)$$

$$D_6 = (1 - \tau_6)[1 + (1 - \varepsilon)\tau_6], \quad (6)$$

$$\text{UHI} = 16.0110 + 0.92621T_0, \quad (7)$$

where UHI is the urban surface temperature; T_{sensor} is the brightness temperature, in a normal case, constant $a_6 = -67.355351$ and constant $b_6 = 0.458606$; ε is the ground surface emissivity; τ_6 is the atmospheric transmittance; T_a is the effective mean atmospheric temperature; and T_0 is the near-surface air temperature. While like many other urban heat island analysis studies [15, 26], we chose a method to obtain TOA spectral radiance and focus on the dynamic evolution of urban surface temperature in spatial and temporal. The standardized method used in this research is as follows:

$$N_i = \frac{\text{UHI}_i - \text{UHI}_{\min}}{\text{UHI}_{\max} - \text{UHI}_{\min}}, \quad (8)$$

where N_i is normalization temperature value of the i th

pixel position; UHI_i means the urban heat temperature value of the i th pixel position; UHI_{\max} and UHI_{\min} represent the maximum and minimum value of UHI before normalization.

Considering the actual situation of the research area, the statistical mean-standard deviation method is selected to divide the thermal field into different levels. According to the statistics of temperature, the average and standard deviation of UHI are used as the demarcation point of the division interval. The mean values of average and standard deviation of UHI maps obtained in 1986, 1996, 2006, and 2016, where $T = 0.512$, $T_s = 0.162$ were taken as the demarcation point of the thermal interval. The divided interval results are shown in Table 2.

3.3. CA-Markov Model. The CA model [27] has a capacity for spatial and temporal change simulation which can be defined as

$$S_{i+1} = f(S_i, \mathbb{N}), \quad (9)$$

where f is the local transition rule of the cell, S is a set of cellular states, N is the cellular field, and $t + 1$ and t represent the start and end time. The Markov model [28] that has been widely used for trend simulation at various scenarios can be expressed as

TABLE 2: Levels of divided thermal interval.

Levels	Criteria	Interval
Low	$T \leq T - 2T_s$	$T \leq 0.188$
Sub-low	$T - 2T_s < T \leq T - T_s$	$0.188 < T \leq 0.350$
Medium	$T - T_s < T \leq T + T_s$	$0.350 < T \leq 0.674$
Sub-high	$T + T_s < T \leq T + 2T_s$	$0.674 < T \leq 0.836$
High	$T + 2T_s < T$	$0.836 < T$

$$S_{t+1} = P_{ij} + S_t, \quad (10)$$

where S_{t+1} and S_t are statuses at time $t + 1$ and t , respectively,

$$P_{ij} = \begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{n1} & \cdots & P_{nn} \end{bmatrix}, \quad (0 \leq P_{ij} \leq 1, \sum_{i=1}^n P_{ij} = 1, 2, \dots, n);$$

P_{ij} is the transition probability matrix in a state. The CA-Markov model is a combination of the CA model and the Markov model, both of which are dynamic models with discrete states [29]. And when combined with the CA model, the CA-Markov model can overcome the limitation that the Markov model failed to catch the spatial distribution in the future and can be used to simulate spatial-temporal changes. In this research, LULC and UST maps in 2006 and 2016 were selected as the study years to calculate the transition area matrix using the Markov model. And a standard 5 by 5 contiguity filter was selected for the CA model which means the condition of the future pixel is not only decided by information from the previous state but also considered by corresponding surrounding pixels.

3.4. Urban Expansion Intensity. The rate of urban expansion (RUE) and the intensity of urban expansion (IUE) are the most common methods for describing urban expansion. The RUE describes the annual average area change of the built-up area during the research period, while IUE refers to the proportion of urban land use expansion area of a space unit in the total urban area during the research period. The RUE and IUE only study the quantitative change of urban built-up area at the beginning and end of a certain period, while ignoring the dynamic process of urban growth. In this paper, based on urban growth intensity, the concept of urban expansion intensity index (UEII) was proposed [30], which describes the degree of differentiation of urban expansion in different directions and denotes the growth of the urban areas of a spatial unit as a percentage of the total area of the land unit in the study period [31] and can be used to study the urban growth process more reasonably and accurately. The mathematical model of UEII is as

$$UEII = \frac{[(UA_{t+n} - UA_t)/T_A]}{n} * 100, \quad (11)$$

where UEII is the urban expansion intensity index of the spatial unit during periods between t and $t + n$; UA_{t+n} and UA_t are the urban area in the spatial unit at time t and $t + n$, respectively; T_A is the total land area; n is the research period.

The UEII can be divided into different levels according to relevant literature [31] and shown in Table 3.

3.5. Urban Gravity Center Movement. The gravity center model reflects the direction of movement and distance to the center of gravity over time [32, 33]. The city's center of gravity reflects the geometric equilibrium of urban space to a certain extent, in which spatial position will be constantly moving during the growth of the urban built-up area. The movement direction of the gravity center reflects which direction of urban growth. To make further analysis of the city's spatial expansion, the urban's gravity center is introduced in this research. The mathematic model of the gravity center can be described as

$$X_t = \frac{\sum_{i=1}^n (a_{ii} + x_i)}{\sum_{i=1}^n a_{ii}}, \quad (12)$$

$$Y_t = \frac{\sum_{i=1}^n (a_{ii} + y_i)}{\sum_{i=1}^n a_{ii}}, \quad (13)$$

where X_t and Y_t are the horizontal and vertical coordinates of the gravity center of the t_{th} year, a_{ii} means the area of i_{th} polygon of the t_{th} year, and x_i and y_i mean the horizontal and vertical gravity center coordinates of the i_{th} polygon. The mathematical equation of gravity center distance can be expressed as follows:

$$D = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2}, \quad (14)$$

where D is the distance of the urban gravity center shifted from the beginning to ending time point; X_i and Y_i represent the horizontal and vertical urban center gravity coordinates, respectively, in the initial moment of research, while X_j and Y_j represent corresponding coordinates in the ending moment.

4. Results

4.1. LULC Maps and Urban Expansion Intensity. There are a total of 15538, 13829, 15991, and 13937 pixels that were selected for training in remotely sensed data captured on 12 August 1986, 15 April 1996, 9 April 2006, and 23 April 2016, respectively. The number of selected test pixels (ground truth points) is 5643 points. The classification maps from 1986 to 2016 are shown in Figure 3; overall accuracy and kappa coefficient evaluation results are shown in Table 4. On the basis of classification maps of 2006 and 2016, the LULC map of 2026 predicted with the CA-Markov model is shown in Figure 4.

According to the above-mentioned urban expansion intensity model, the statistical results of building area classification and simulation in 1986, 1996, 2006, 2016, and 2026 (shown in Figures 3 and 4), and urban growth intensity graded criteria, the level of urban growth intensity is obtained (Table 5).

TABLE 3: Levels of urban expansion intensity index.

Levels	Slow speed	Low speed	Moderate speed	Fast speed	High speed
Range	(0, 0.28]	(0.28, 0.59]	(0.59, 1.05]	(1.05, 1.92]	(1.92, +∞)

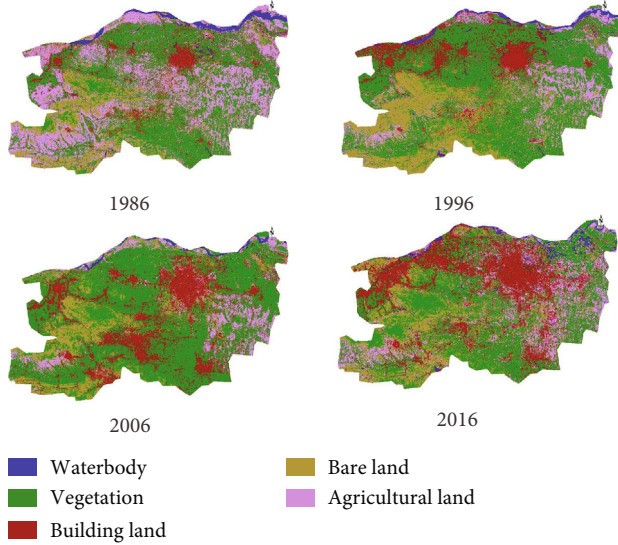


FIGURE 3: LULC classification maps over Zhengzhou city from 1986 to 2016.

TABLE 4: Classification accuracy.

	1986	1996	2006	2016
OA	79.26%	84.01%	82.86%	85.13%
Kappa	0.78	0.82	0.81	0.83

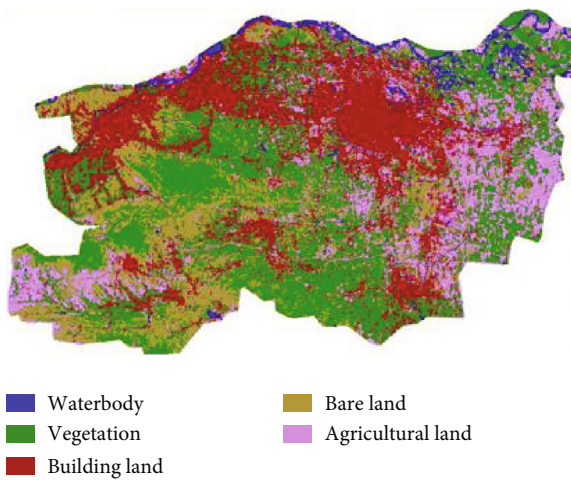


FIGURE 4: Predicted LULC map of Zhengzhou city in 2026.

4.2. *Urban Gravity Center Changes.* Based on the gravity center model, the gravity center and changing footprint of Zhengzhou’s urban built-up area from 1986 to 2026 were calculated. The gravity centers and the trends of gravity center development are shown in Table 6.

The results showed that the direction and distance of the city’s center of gravity movement are quite different in a different period. The center of gravity movement of each period during the study period has the following characteristics: (1) The city center of Zhengzhou moved 9.65 km to the northwest from 1986 to 1996. The dominant driving forces was the rapid development of urban construction in Xingyang, a country-level city now becomes as a district of the metropolis of Zhengzhou, from 1986 to 1996. The development of Jinshui District of Zhengzhou was earlier and faster than other urban areas, which led to the gravity center movement to Jinshui and Xingyang district. (2) The city center moved 15.69 km to the southeast from 1996 to 2006, mainly because with the acceleration of urbanization, the Zhengzhou station was expanded; the Xinzheng international airport, the city around expressway, and some large-scale construction project were gradually put into service, which expanded the surface area of man-made land type in the southeast direction and causing the shift of gravity obviously. (3) The city center moved 7.87 km to the north from 2006 to 2016, mainly due to the large-scale construction of Zhengzhou High-tech Industrial Zone and the operation of Zhengzhou East High-Speed Rail Station operated in 2012, which resulted in a rapid increase in the man-made surface in the north. Meanwhile, the development of old district Erqi city of Zhengzhou has been saturated. Therefore, the development speed of Erqi District was lower than that of Jinshui District during 2006 to 2016, which made the gravity center move to the north in the past 10 years. (4) According to the prediction result of LULC using the CA-Markov model, the gravity center will move 1.14 km westward from 2016 to 2026. With the continuous urbanization process, the urbanization of Zhengzhou has basically become saturated and the expansion speed has slowed down.

4.3. *Spatial Orientation of Urban Expansion.* The vector data of human settlement of the adjacent periods between 1986 and 1996, 1996 and 2006, 2006 and 2016, and 2016 and 2026 were overlapped and carried out urban intensity growth analysis in NNE, NEE, SEE, SSE, SSW, SWW, NWW, and NNW directions (Figure 5 and Table 7).

According to the above results, in general, the strongest urban growth intensity was in the NEE and SWW directions from 1986 to 2026, which is the dominant direction of urban growth in the study area. Combined with the geographical location of the administrative area of the study area, the strongest urban growth intensity is concentrated in Jinshui District, Erqi District, and Xinmi, mainly because the urbanization

TABLE 5: Levels of urban expansion intensity from 1986 to 2026.

Research period	1986-1996	1996-2006	2006-2016	2016-2026
UEII	0.665	1.235	1.182	0.916
Growth levels	Medium	Rapid	Rapid	Medium

TABLE 6: Gravity center from 1986 to 2026 in the WGS84 system.

Year	Horizontal coordinate	Vertical coordinate	Offset distance (km)
1986	728213.496	3842411.355	
1996	720616.338	3848368.364	9.65
2006	727688.303	3834364.776	15.69
2016	726145.678	3842084.841	7.87
2026	725018.082	3842235.005	1.14

process is accelerating with the rapid development of Zhengzhou city. As the core area of Zhengzhou city, Jinshui District and Erqi District have become the center of urbanization and have continued to expand to the outside. Xinmi is located in the central part of Zhengzhou and adjacent to the Erqi District, which is affected by the urbanization of the Erqi District. Taking advantage of its superior development environment and conditions, the urban growth intensity of Xinmi has increased significantly.

Temporal dimension studies showed that the urban growth intensity in NEE and NWW orientation was relatively strong from 1986 to 1996, which is mainly due to the rapid development of Erqi District and Xingyang District. Urban growth was mainly in the direction of NEE, SWW, and SSW from 1996 to 2006, and during this period, the urban growth pattern was consistent, and the growth of other cities was relatively slow, which made the city's center of gravity move to the southeast. From 2006 to 2016, the urban growth intensity in the NEE, NWW, and SWW directions was relatively strong. During this period, there was vigorously built activities in transportation facilities and high-tech industries in Zhengzhou city, which further strengthened the land development efforts and increased the intensity of urban growth. In the meantime, the land use development status of prefecture-level cities such as Gongyi and Xinmi was growing rapidly because of an unsaturated state. According to the prediction of LULC results, there will be a similar trend of urban growth between the period of 2016 to 2026 and the period of 2006 to 2016. The urbanization construction of Zhengzhou city has reached a new stage, and urban growth will be relatively stable and mature at that time.

4.4. UHI Results and Statistical Features. The results of UHI from 1986 to 2016 are retrieved using WMA, and the UHI degree of 2026 is simulated using the CA-Markov model. UHI distribution maps and UHI degrees from 1986 to 2026 are shown in Figure 6 and Table 8.

The spatial distribution maps of UHI (Figure 6) in the study area showed that the low temperature and sub-low temperature zone were mainly located in the Yellow River basin in the northeast of the study area. And the high-

temperature zone is mostly located in the urban built-up zone, north of the central part. Based on the statistical results of Figure 7 and Table 8, the rank of different temperature areas has characteristics as that the dominant areas by proportion are medium temperature and sub-high temperature level, followed by high temperature, sub-low temperature, and low-temperature areas.

4.5. Spatial-Temporal Changes of Urban Thermal Environment.

According to the statistical results of the UHI degree, the dynamic curve of the UHI level is obtained (Figure 8). The changes and proportions of UHI in the study area between 1986 to 1996, 1996 to 2006, and 2006 to 2016 were calculated (shown in Tables 9–12). The methodologies used for LULCC analysis were adopted to analyze spatial-temporal changes of the urban thermal environment (shown in Figures 7–9).

The dynamic curve of the UHI trend result (Figure 8) showed that the area of the low-temperature zone and sub-low temperature zone is shrunken during the period from 1986 to 2026. The dominated UHI zones during this period are medium, sub-high, and high-temperature zone. The area of the low-temperature zone has a downward trend, and the area of the sub-low temperature zone experiences a slight increase and then a continuous decline. The change trend of the sub-high temperature zone is opposite to that of the sub-low temperature zone. The middle-temperature zone and the high-temperature zone are both in a wave-like rising trend, while the trend of medium temperature zone is opposite to that of the high-temperature zone.

Combined with the distribution map of the urban thermal environment in the study area for the past 40 years, a long-term UHI comparison was carried out (Figure 9). Between 1986 and 1996, the thermal environment change process of Zhengzhou city was complicated, with a large change range, fast change dynamic. The area of sub-low temperature and medium temperature zone increased; the area of low-temperature, sub-high temperature, and high-temperature zone decreased. The dynamic of the medium temperature zone changed the most, that is, more than 10%. Between 1996 and 2006, the thermal environment of Zhengzhou has a tendency to change to a high-temperature area. The area of medium and low temperature decreased, the area of sub-high temperature increased, and the high-temperature area changed the most, that is, more than 85%. While the change range of other degrade is relatively small, in particular, the area of the low-temperature zone is almost unchanged. From 2006 to 2016, the trend of the thermal environment was similar to that between 1986 and 1996. The area occupied by extreme temperature level decreased; the area occupied by intermediate temperature level increased. And the urban thermal environment had a tendency to concentrate to medium and high-temperature levels. According to the prediction result, between 2016 and 2026, the overall thermal environment of the city has a small change and is in a state of dynamic equilibrium. It is predicted that by 2026, urban surface temperature will further evolve into high-temperature areas, and the problem of the UHI effect will become more prominent.

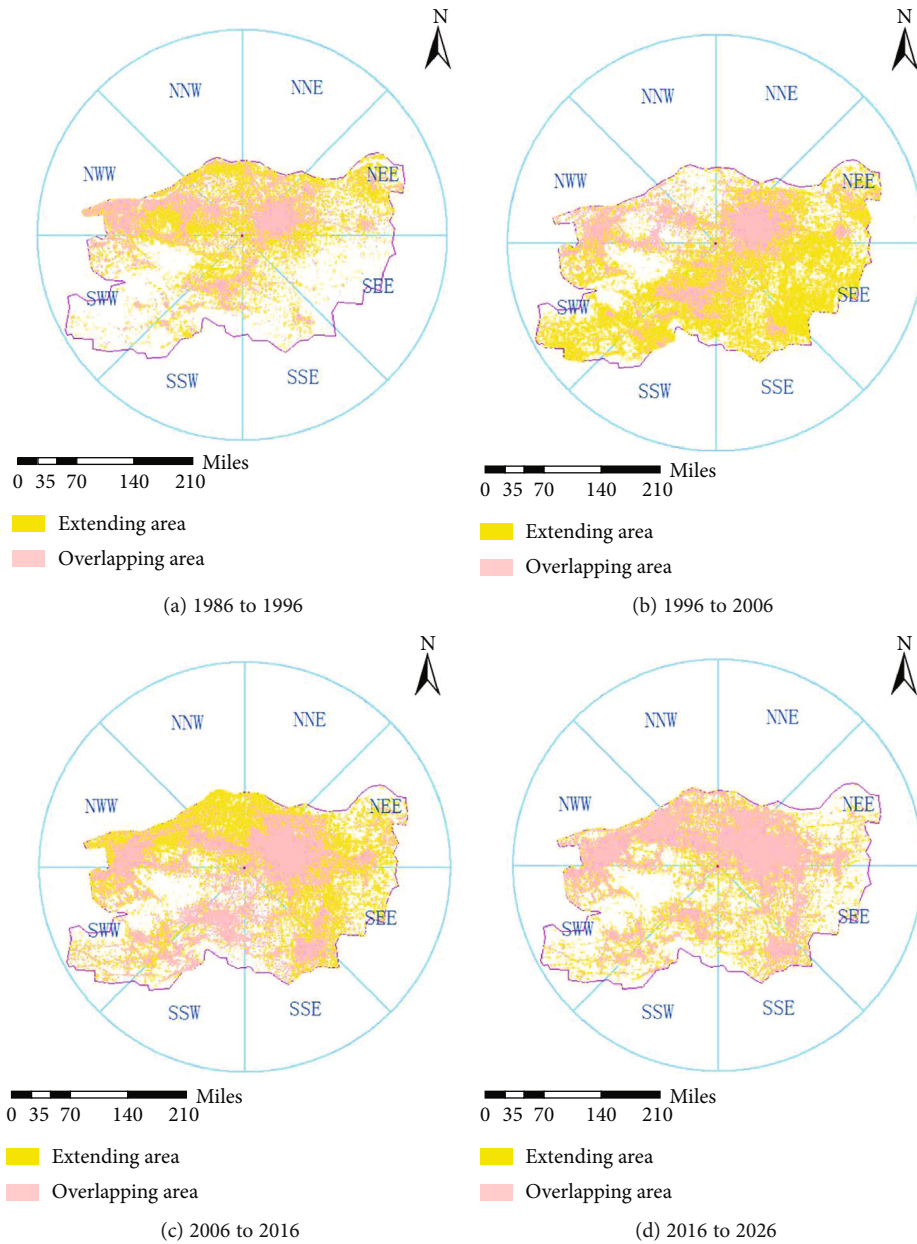


FIGURE 5: The results of urban expansion.

TABLE 7: Urban growth intensity in various azimuth from 1986 to 2026.

Azimuth	1986-1996	1996-2006	2006-2016	2016-2026
NNE	0.404	0.351	0.726	0.476
NEE	3.267	4.770	4.610	3.093
SEE	0.234	1.569	1.731	1.120
SSE	0.072	0.813	0.274	0.385
SSW	0.328	2.641	0.257	0.786
SWW	0.887	4.187	2.129	2.694
NWW	2.264	0.593	2.540	1.376
NNW	0.380	0.079	0.964	0.573

4.6. Spatial Character of UHI. In this study, spatial characters of UHI are analyzed using profile distribution and buffer zone methods. Profile analysis can intuitively describe the spatial distribution pattern and overall evolution rule of the urban thermal environment. Erqi Memorial Tower which located in the core area of Zhengzhou city is selected as the center of profile, and the section line was constructed in east-west and north-south directions. The east-west section line is based on the Jianshe road and passes through important buildings such as Hongsen Building, Air Defense Academy, Zhengzhou West Bus Station, Jianshe West Road, Zhengzhou Vocational and Technical College, Zhengzhou Confucian Temple, and Zhengzhou No. 96 Middle School. The north-south direction section line is roughly along the Erqi Road, passing important places such as Zhengzhou

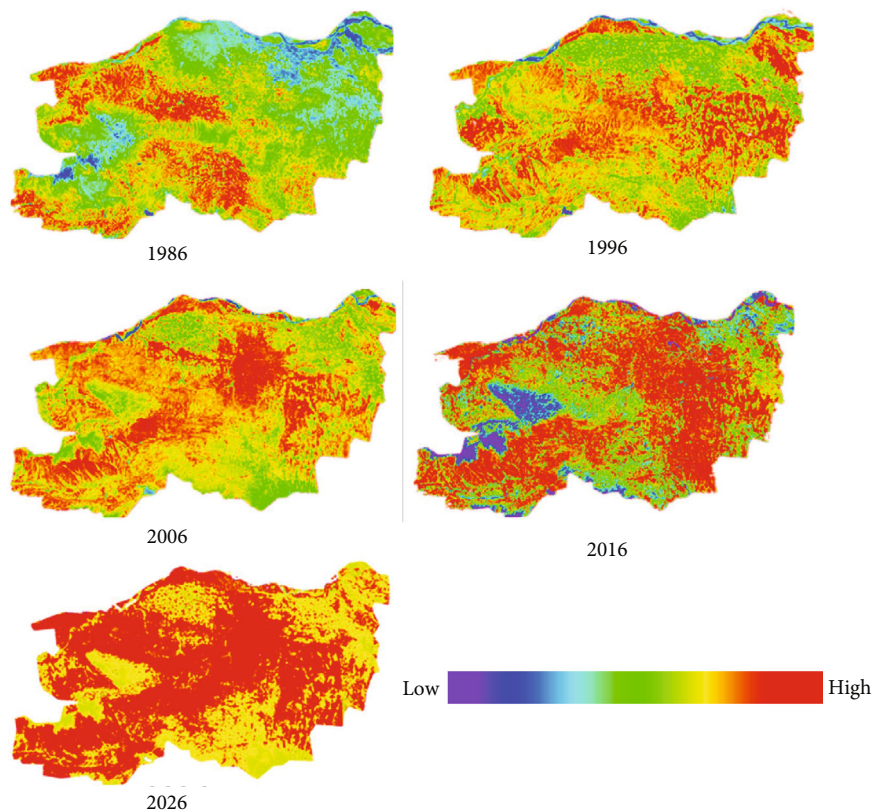


FIGURE 6: Urban heat island distribution maps of Zhengzhou city.

TABLE 8: UHI grade levels of Zhengzhou from 1986 to 2026/ha.

Year	Low	Sub-low	Medium	Sub-high	High
1986	83625.55	103911.49	147552.85	263781.08	153504.82
1996	16351.18	140410.92	295761.75	178092.00	121759.94
2006	16231.06	94541.95	219800.68	193058.52	228743.58
2016	4981.83	39214.67	269041.02	245564.53	193573.74
2026	3487.13	32674.59	253577.21	266035.92	196600.94

People's Park, Henan Provincial Department of Commerce, Zhengzhou People's Hospital, Zhengzhou University, Henan Agricultural University, and Dehua Street. The profiles of LST in a different period were calculated (Figure 10).

Profiles of LST in a different period showed that (1) there are significant jagged jumps in the EW and NS profiles of LST in each period. The urban central area and the suburbs are characterized by "upward convexity" and "depressed." This is mainly due to the complex structure of the underlying surface of the urban surface, and the change of the underlying surface characteristics in a small area makes the surface temperature abrupt; (2) in this study, LST in the central area of the city was generally higher, and there was obvious "bumps" with uneven "peak" and "low valley" morphological features. In the suburbs, the LST was low, while the "jumping" phenomenon with a large jump and fast frequency of change is more significant than that of the city center. The cause of that related to the underlying surface of the central area is mainly

composed of steel, cement, and masonry, and the structure is relatively simple, while the underlying of suburb mainly consists of green land, water area, soil, cement, and masonry, and the structure is relatively complicated; (3) the comparison of various characteristics of EW and NS LST profiles showed that the EW direction changes rapidly and complex than that of NS direction. This phenomenon showed that the underlying surface of EW section of Zhengzhou has a more diverse structure type than NS direction, and changes are more complicated; (4) the perspective of the long-term changes in the distribution pattern of LST, EW, and NS profiles showed that with the continuous development of urbanization construction, the overall average temperature of the city was gradually increasing, while the frequency and jump range of urban surface temperature profile had a decreasing trend. This is mainly because the urbanization construction of Zhengzhou city is gradually improved, the characteristic structure of the underlying surface of the city tends to be stable, and the temperature field structure of the urban thermal environment tends to be simplified.

Buffer analysis is an important spatial analysis method used to determine the proximity of research elements in GIS. In this research, the buffer analysis method was adopted to study the urban thermal environment. With the support of this methodology, the relationship between LST and the location of the city center from 1986 to 2026 was discussed, the spatial characteristics of LST distribution within a certain distance were described, and the pattern of thermal environment space of Zhengzhou city was analyzed. Four circular

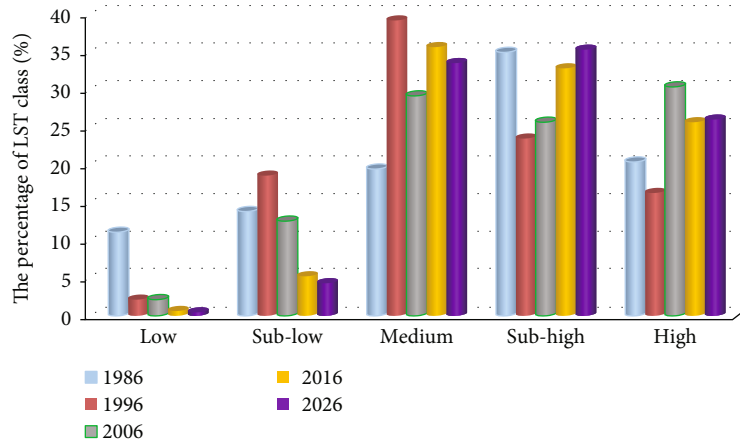


FIGURE 7: Percentage of different UHI levels in a different year.

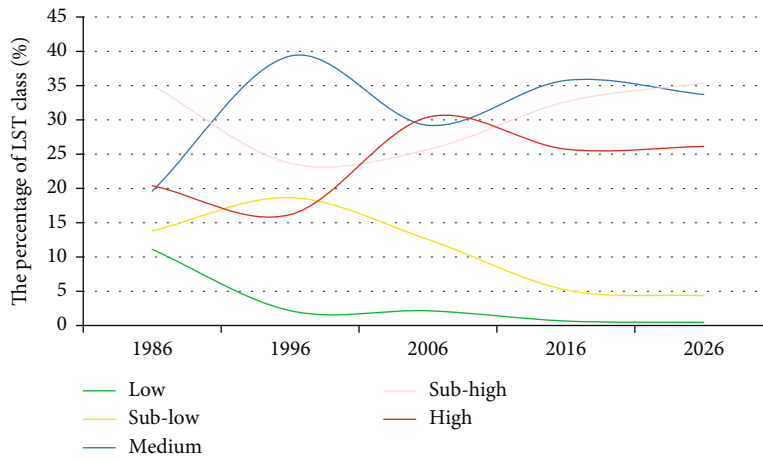


FIGURE 8: Dynamic curve of UHI of Zhengzhou city.

TABLE 9: UHI change in different levels from 1986 to 1996.

Levels	1986	1996	Amount (ha)	Amplitude (%)	Degree (%)
Low	83625.55	16351.18	-67274.37	-80.45	-8.04
Sub-low	103911.49	140410.92	36499.43	35.13	3.51
Medium	147552.85	295761.75	148208.90	100.44	10.04
Sub-high	263781.08	178092.00	-85689.08	-32.48	-3.25
High	153504.82	121759.94	-31744.88	-20.68	-2.07

TABLE 10: UHI change in different levels from 1996 to 2006.

Levels	1986	1996	Amount (ha)	Amplitude (%)	Degree (%)
Low	16351.18	16231.06	-120.12	-0.73	-0.07
Sub-low	140410.92	94541.95	-45868.97	-32.67	-3.27
Medium	295761.75	219800.68	-75961.07	-25.68	-2.57
Sub-high	178092.00	193058.52	14966.52	8.40	0.84
High	121759.94	228743.58	106983.64	87.86	8.79

TABLE 11: UHI change in different levels from 2006 to 2016.

Levels	1986	1996	Amount (ha)	Amplitude (%)	Degree (%)
Low	16231.06	4981.83	-11249.23	-69.31	-6.93
Sub-low	94541.95	39214.67	-55327.28	-58.52	-5.85
Medium	219800.68	269041.02	49240.34	22.40	2.24
Sub-high	193058.52	245564.53	52506.01	27.20	2.72
High	228743.58	193573.74	-35169.84	-15.38	-1.54

TABLE 12: UHI change in different levels from 2016 to 2026.

Levels	1986	1996	Amount (ha)	Amplitude (%)	Degree (%)
Low	4981.83	3487.13	-1494.70	-30.00	-3.00
Sub-low	39214.67	32674.59	-6540.08	-16.68	-1.67
Medium	269041.02	253577.21	-15463.81	-5.75	-0.57
Sub-high	245564.53	266035.92	20471.39	8.34	0.83
High	193573.74	196600.94	3027.20	1.56	0.16

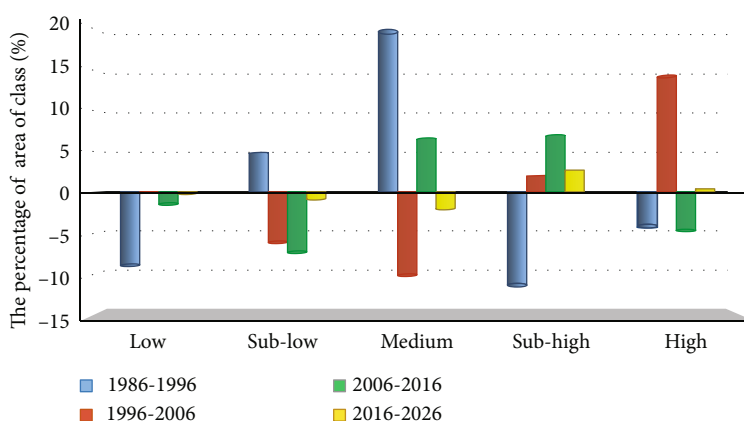


FIGURE 9: UHI change in percentage of different levels from 1986 to 2026.

buffers with a distance of 5 km with the Erqi Memorial Tower taken as the center of the circle was drawn (Figure 11). The LST grade vector data of 1986, 1996, 2006, 2016, and 2026 and buffer vector data were superimposed and analyzed to obtain LST distribution results of five different distances from the center of the circle (Figure 12).

The proportion area of different grades in the different buffer zone (Figure 12) indicates that during the 40 years' urbanization in Zhengzhou, the area of low-temperature zone gradually decreased and the area of intermediate grade temperature zone experienced a process of increasing first and then decreasing; the area of higher temperature zone increased continuously. Therefore, the pattern of urban thermal environment in Zhengzhou city showed a trend of agglomeration to an intermediate level firstly and then change to the extremely high temperature.

By longitudinally comparing the temperature levels in the buffer zone, the results showed that in the 0 km to 10 km buffer zone, the temperature of each grade changes greatly, the area of low-temperature zone gradually decreases, and

the area of high-temperature zone continues to increase rapidly. The area ratio of high temperature is more than 80% in 2016, and this number will close to 90% in 2026 by prediction. In the buffer zone of 10 km to 20 km, the area of low-temperature area is generally declining, the area of medium temperature zone is stable, and the area of high-temperature zone is changed from the least to the most. It is predicted that by 2026, the area of the lower temperature zone will be close to 10 km². In the 20 km to 30 km buffer zone, the area of low temperature and sub-low temperature continues to decline, the area of medium temperature zone and sub-high temperature zone is in a wave-like state, and the area of high-temperature zone increases slightly. In the 30 km to 40 km buffer zone, the medium and sub-high temperature zones are the main temperature grades. The proportion of high-temperature zone increased at the beginning and then stabilized at 20%; the proportion of low-temperature zone is small and in a reduced state. The above phenomenon is mainly due to the continuous acceleration of the urbanization process in Zhengzhou city, the continuous expansion

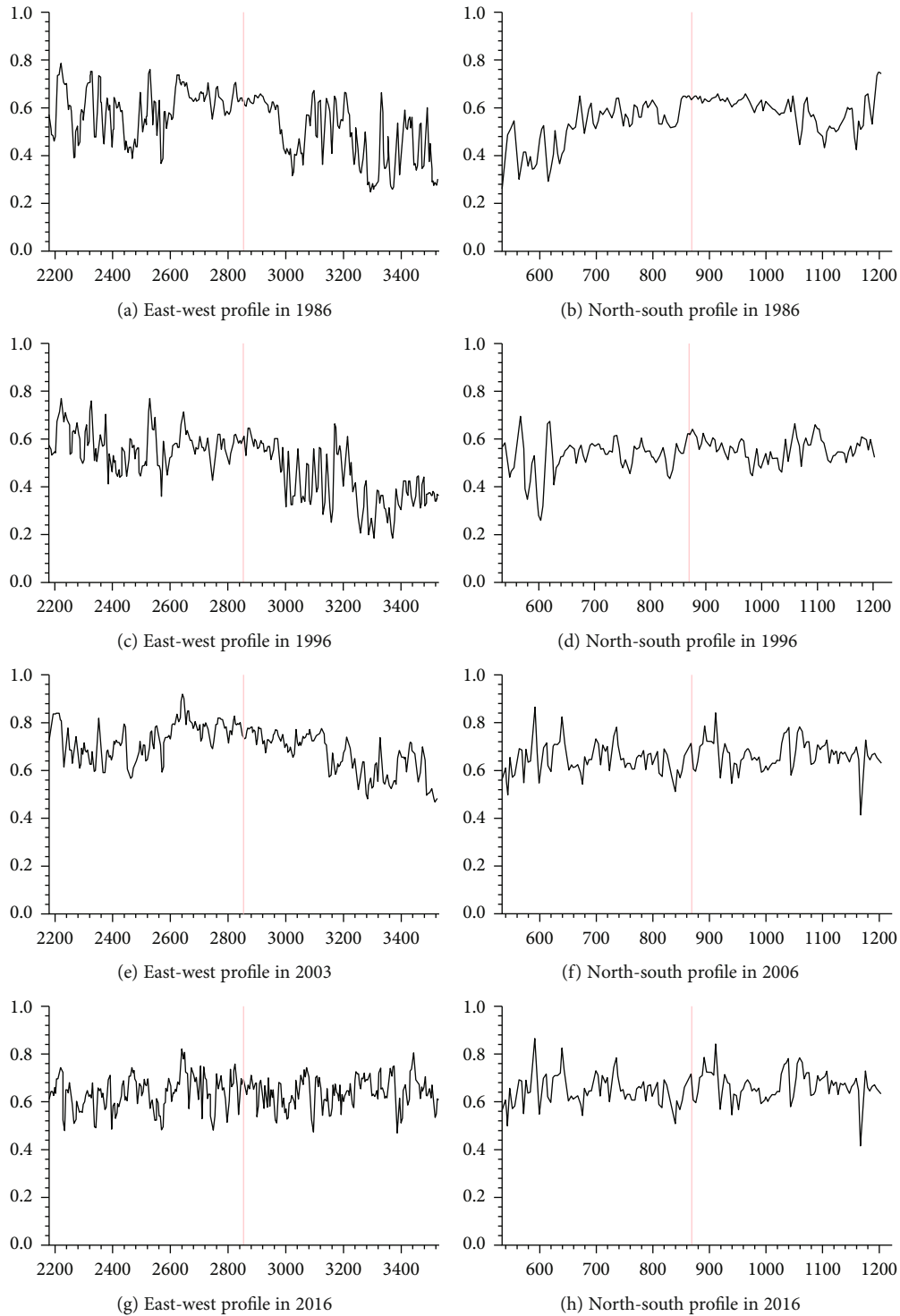


FIGURE 10: Profiles of UHI of Zhengzhou.

of the human-made area around Erqi Tower, and the increasing surface area of artificial land, which lead to continuous temperature rising in the high buffer zone, and the low-temperature zone reduced gradually.

The horizontal comparison (in the 10 km, 20 km, 30 km, and 40 km buffer zone) of the temperature levels showed that the change trend of low-temperature zone increased and then

decreased, which displayed as a “n” shape change trend. The area of sub-low temperature zone increased gradually during the period of 2006 to 2026. The overall change range of medium temperature zone during the period from 1986 to 1996 is small, while it is increasing gradually during the period of 2006 to 2026. The sub-high temperature zone occupied a larger proportion in the 10 km buffer zone than

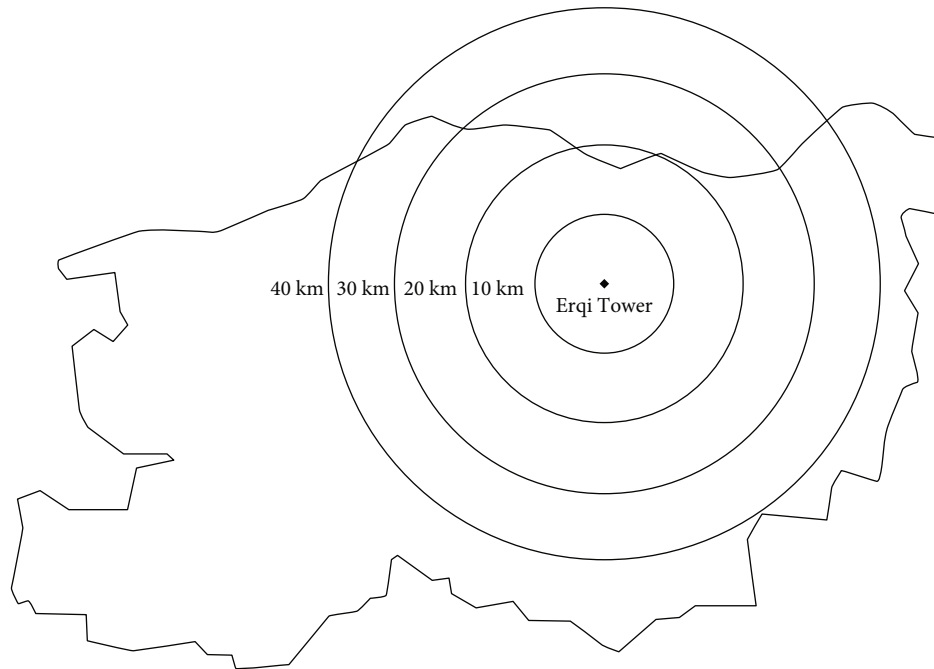


FIGURE 11: The buffer setting of UHI.

other buffer zones from 1986 to 1996. From the period of 1996 to 2006, the proportion of sub-high temperature zones was almost the same in each buffer zone, and from the period of 2006 to 2016, the spatial distribution of sub-high temperature had an opposite pattern of a period from 1986 to 1996. The distribution characteristics of LST are mainly due to the high degree of urbanization in the area closer to the Erqi Tower where the urban underlying surface features are dominated by steel, cement, and masonry which increase the impact of UHI effect.

4.7. Relationship between LULC and UHI. In this subsection, the relationship between LULC and UHI is analyzed using three quantitative indices such as normalized difference vegetation index (NDVI), normalized difference building index (NDBI), and normalized difference waterbody index (NDWI) which indicate relative important LULC types vegetation, built-up area, and waterbody.

NDVI is a standardized way to measure vegetation, which quantifies vegetation by measuring the difference between red (R) and near-infrared (NIR) spectral reflectance value from remotely sensed data. NDVI can be calculated in its formula $NDVI = (NIR - R) / (NIR + R)$. NDBI is another solution for easily calculating of the built-up area because it is simple, rapid, and accurate in urban area mapping. NDBI can be calculated using the formula as $NDBI = (SWIR - NIR) / (SWIR + NIR)$, where SWIR and NIR represent the spectral reflectance value of shortwave infrared and NIR band. The NDWI, which can be calculated using green and NIR spectral reflectance value using its formula $NDWI = (GREEN - NIR) / (GREEN + NIR)$, is most appropriate for waterbody mapping. In this research, there are 30 sets of NDVI, NDBI, and MNDWI index and corresponding UST data were randomly

selected from the data of 1986, 1996, 2006, 2016, and 2026 for linear regression, and the results are shown as

$$T = -3.437 + NDVI + 29.835, \quad (15)$$

$$R^2 = 0.835,$$

$$T = 11 + NDBI + 25.539, \quad (16)$$

$$R^2 = 0.821,$$

$$T = -8.628 + NDWI + 31.481, \quad (17)$$

$$R^2 = 0.868.$$

The regression equations (15), (16), and (17) showed that there was a significant negative correlation between NDVI, NDWI, and LST; the correlation coefficients are 0.835 and 0.868, respectively. Waterbody has a more obvious effect on relieving LST than vegetation. There was a significant positive correlation between NDBI and LST, with a correlation coefficient of 0.821.

5. Analysis and Discussion

The urban expansion intensity level (Table 5) from 1986 to 2026 indicates that the urban growth intensity maintains an increase during this period in general. There was a medium-speed growth stage from 1986 to 1996 and a rapid growth from 1996 to 2006. And according to the predicted results of LULC maps, there will be a slight slow down to some extent in the next 10 years, while the urban growth is still at a medium speed.

What is the relationship between LULC type and UHI? In order to discover this relationship and to effectively and

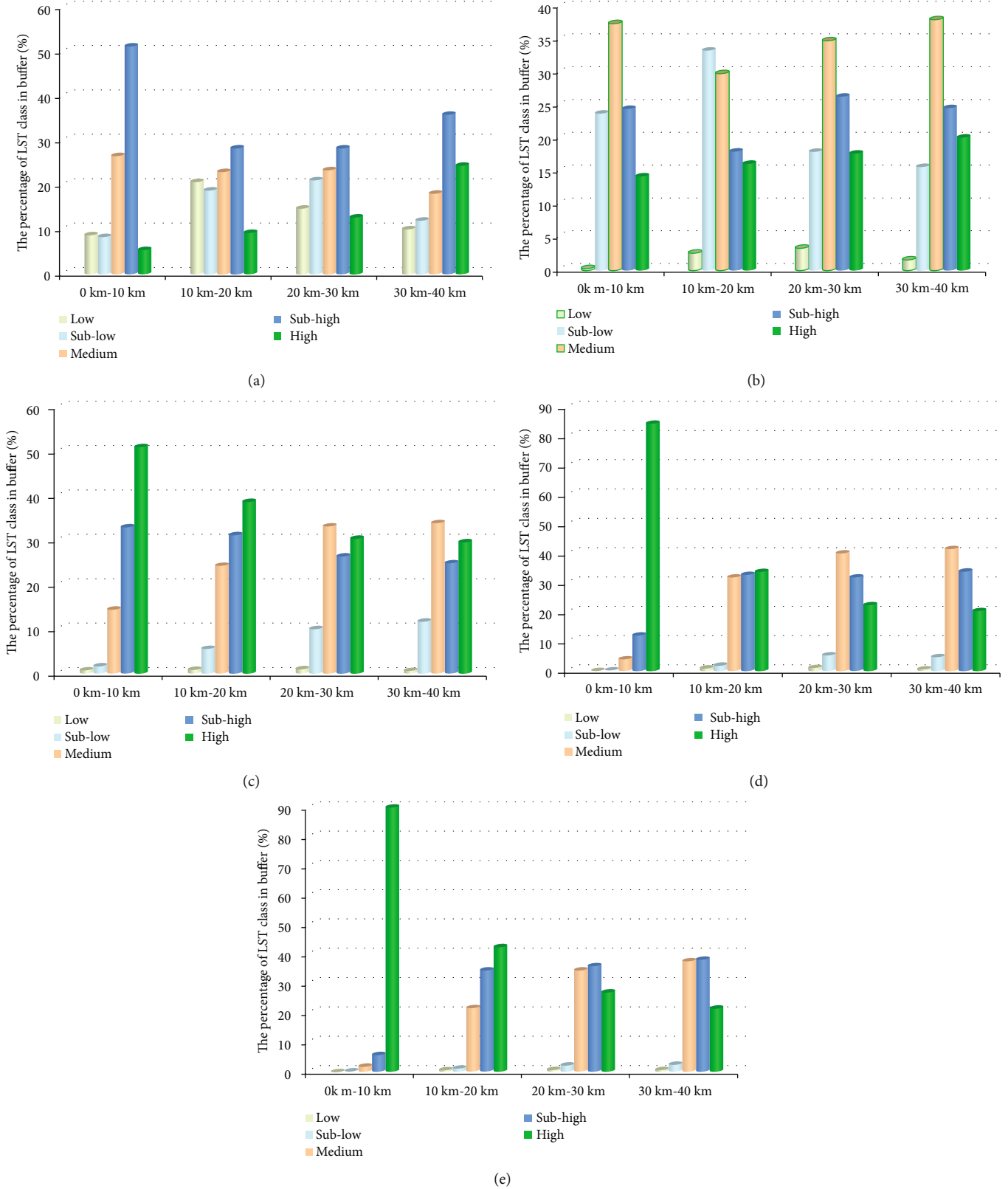


FIGURE 12: Proportion of each UHI levels in each buffer zone. (a) 1986; (b) 1996; (c) 2006; (d) 2016; (e)2026.

reasonably analyze the impact and contribution of different surface cover to urban thermal environment, the artificial surface, vegetation, water, farmland, and bare land vector

data of 1986, 1996, 2006, 2016, and 2026 were sequentially performed on vector data of low-temperature, sub-low temperature, medium temperature, sub-high temperature, and

TABLE 13: Area ratio of land cover type in each LST grade in 1986 (%).

	Low	Sub-low	Medium	Sub-high	High
Building-up	0.23	8.19	52.27	31.23	8.08
Vegetation	0.45	34.61	48.73	12.81	3.40
Water	64.46	23.78	8.48	2.48	0.80
Farmland	0.04	1.96	23.08	36.38	38.54
Bare land	0.77	7.13	39.78	32.10	20.22

TABLE 14: Area ratio of land cover type in each LST grade in 1996 (%).

	Low	Sub-low	Medium	Sub-high	High
Building-up	7.08	10.11	20.71	43.06	19.03
Vegetation	10.69	13.82	17.35	33.06	25.08
Water	58.24	20.20	12.38	6.45	2.73
Farmland	7.04	16.96	30.32	35.74	9.94
Bare land	11.71	13.23	19.47	38.78	16.82

TABLE 15: Area ratio of land cover type in each LST grade in 2006 (%).

	Low	Sub-low	Medium	Sub-high	High
Building-up	0.37	1.59	15.42	39.19	43.42
Vegetation	0.00	6.95	49.21	31.72	12.12
Water	58.48	25.56	15.29	0.48	0.19
Farmland	0.00	0.26	4.73	27.10	67.91
Bare land	0.19	4.09	25.19	33.86	36.68

TABLE 16: Area ratio of land cover type in each LST grade in 2016 (%).

	Low	Sub-low	Medium	Sub-high	High
Building-up	0.14	1.39	16.31	36.47	45.70
Vegetation	3.22	21.29	43.10	22.01	10.38
Water	29.70	53.06	14.04	2.17	1.04
Farmland	0.11	1.14	15.12	25.42	58.21
Bare land	0.32	8.81	23.03	24.36	43.48

TABLE 17: Area ratio of land cover type in each LST grade in 2026 (%).

	Low	Sub-low	Medium	Sub-high	High
Building-up	0.23	1.13	17.78	35.18	45.69
Vegetation	0.14	6.73	46.28	34.38	12.47
Water	7.30	12.04	44.67	19.81	16.18
Farmland	0.42	2.41	28.23	38.24	30.69
Bare land	0.13	1.45	21.08	39.74	37.60

high-temperature zone. And the proportion of different LULC types occupied in different temperature grades was also calculated (Tables 13–17).

The spatial-temporal change of thermal environment from 1986 to 2026 showed that the proportion of LULC types in a temperature grade changed significantly. For example, the built-up area in the high-temperature zone has rapidly increased from less than 10% in 1986 to more than 45% in 2016. In the low-temperature zone, the proportion of water has been close to 65% from the beginning of the study period and has continued to drop to less than 30% in 2016, and it is predicted that the proportion will further decrease significantly by 2026. Priyankara et al.'s research [34] demonstrates similar findings that mean LST has a strong significant positive relationship with a fraction of impervious surface and persistent impervious surface, while a strong negative relationship with a fraction of forest surface and new added impervious surface. Priyankara et al. suggested that more vegetation areas are recommended in both horizontal and vertical directions to reduce the UHI effect.

The change of proportion of the same LULC type in different temperature grades from 1986 to 2026 showed that the changes of the built-up area and bare land area are similar, that is, in the early stage of research, these two land cover types are mainly concentrated in medium temperature zone. With the expansion of the urban area, the proportion of the high-temperature zone increased. When combining the spatial distribution of built-up and bare land in the study area, it fully demonstrates the significant positive impact and contribution of these two LULC types on urban thermal environment. The distribution of vegetation is mainly in a medium level of UHI grades, which indicates that vegetation plays an important role in balancing surface temperature. Waterbody is distributed in a low-temperature zone, and proportion was gradually decreased in the low-, medium-, and high-temperature zone. With the acceleration of urbanization process, the proportion of waterbody in a low-temperature zone will be less than 8% in 2026 by prediction, but it still has an obvious advantage over other land cover types showing that the waterbody is indispensable for reducing UHI effect and maintaining the balance and stability of urban thermal environment. LULC types of distribution on all temperature grades were discussed here, while the contribution of other factors might be differentiated. Ranagalage et al.'s research [35] revealed that mean LST was positively correlated with the increase of fraction ratio of building area and forest area and with the decrease of fraction ratio of agricultural and forest area. Building density is a crucial element in increasing LST.

6. Conclusions

In this research, comprehensive research of LULCCs on urban heat environment assessment was performed using the RS and GIS spatial technique. The spatial and temporal changes of urbanization of Zhengzhou city from 1986 to 2026 were analyzed, and conclusions can be drawn as follows.

Land use and land cover changes of 40 years in Zhengzhou city have been studied and analyzed. In the past forty

years, the significant transform land cover type is the built-up area. The LULC conversion relationship in the study area is extremely complex. The strongest urban growth intensity happened in 1996 to 2006. The study area generally expanded in NEE and SWW orientation most notably from 1986 to 2026.

A combination analysis of natural and thermal environmental has been yielded. During the period from 1986 to 2026, LST in the study area is distributed mainly in medium, sub-high, and high-temperature zone. The thermal environment change process in Zhengzhou is relatively complicated, and the dynamics of spatial-temporal change were dramatical. In the early stage of the study period, the temperature grade trends to medium zone, the middle trends to high-temperature zone, and the later stage has a tendency to change to the medium zone. The temperature changes in the east-west direction were faster than that in the north-south direction. There is a significant correlation between vegetation, water and urban surface temperature, and LST, and a positive correlation between the built-up area and LST.

The apparent drawback of the paper is that our study is limited to daytime UHI due to the limit of the Landsat dataset. Therefore, future works tend to integrate the advantages of Landsat and nighttime light datasets and then extend our study to nighttime.

Abbreviations

LULCCs:	Land use and land cover changes
UHI:	Urban heat island
NEE:	North-northwest
WSW:	West-southwest
CA-Markov:	Cellular Automata Markov
RS:	Remote sensing
LULC:	Land use and land cover
UST:	Urban surface temperature.

Data Availability

The data availability statement: all datasets used in this paper including the primary remotely sensed data, boundary vector data, and processed results can be obtained from the hyperlink: https://pan.baidu.com/s/1d9GwkEpwryWYgHU_78qV5g, with password: c8rp. Researchers who are interested in this topic can download the data from the above hyperlink or contact the corresponding author to obtain source data to conduct secondary analysis.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

P.L. performed the conceptualization; P.L. and RM. H. did the methodology; P.L. performed the formal analysis; P.L. helped in writing and preparing the original draft; P.L. and RM. H. contributed to the writing, reviewing, and editing of the manuscript; P.L. performed the project administration;

SJ. J. and XF. L. conducted the experiment; Pei Liu proposed the idea, wrote the original manuscript and the following revisions, and provided the funding. P.L. and SJ. J. executed all the experiments; YP. L., RM. H., and HW. Z. contributed to the revisions and provided valuable comments. Pei Liu, Shoujun Jia, Ruimei Han, Yuanping Liu, Xiaofeng Lu, and Hanwei Zhang contributed equally to this work.

Acknowledgments

We would like to thank Dr. Xiaoqian Cheng, graduate student Weijia Zhao, and Changhu Wang for their assistance with the preparation of the datasets and some field works used in this study. The authors also express their thanks to USGS for making Landsat datasets available. This research was funded by the National Natural Science Foundation of China Grant number (No. 41601450), the Cooperative Exchange Program between the National Natural Science Foundation of China and the Royal Society of UK (No. 42011530174), the Key Technology R and D Program of Henan Province (No. 182102310860), and China Scholarship Council Grant (No. 201808410212).

References

- [1] P. Fu and Q. Weng, "A time series analysis of urbanization induced land use and land cover change and its impact on land surface temperature with Landsat imagery," *Remote Sensing of Environment*, vol. 175, pp. 205–214, 2016.
- [2] S. Pauleit, R. Ennos, and Y. Golding, "Modeling the environmental impacts of urban land use and land cover change—a study in Merseyside, UK," *Landscape and Urban Planning*, vol. 71, no. 2–4, pp. 295–310, 2005.
- [3] Q. Ren, C. He, Q. Huang, and Y. Zhou, "Urbanization impacts on vegetation phenology in China," *Remote Sensing*, vol. 10, p. 1095, 2018.
- [4] M. Sapena and L. Á. Ruiz, "Analysis of land use/land cover spatio-temporal metrics and population dynamics for urban growth characterization," *Environment and Urban Systems*, vol. 73, pp. 27–39, 2019.
- [5] J. Peng, J. Ma, Q. Liu et al., "Spatial-temporal change of land surface temperature across 285 cities in China: an urban-rural contrast perspective," *Science of the Total Environment*, vol. 635, pp. 487–497, 2018.
- [6] I. Manakos, M. Tomaszewska, I. Gkinis et al., "Comparison of global and continental land cover products for selected study areas in South Central and Eastern European region," *Remote Sensing*, vol. 10, no. 12, p. 1967, 2018.
- [7] M. Melchiorri, A. Florczyk, S. Freire, M. Schiavina, M. Pesaresi, and T. Kemper, "Unveiling 25 years of planetary urbanization with remote sensing: perspectives from the Global Human Settlement Layer," *Remote Sensing*, vol. 10, no. 5, p. 768, 2018.
- [8] J. S. Powers, "Changes in soil carbon and nitrogen after contrasting land-use transitions in northeastern Costa Rica," *Ecosystems*, vol. 7, no. 2, pp. 134–146, 2004.
- [9] T. M. Tena and P. N. A. Mwaanga, "Impact of land use/land cover change on hydrological components in Chongwe River Catchment," *Sustainability*, vol. 11, no. 22, p. 6415, 2019.
- [10] Q. Jiang, Q. Fu, and Z. Wang, "Comprehensive evaluation of regional land resources carrying capacity based on projection

- pursuit model optimized by particle swarm optimization,” *Transactions of the Chinese Society of Agricultural Engineering*, vol. 27, no. 11, pp. 319–324, 2011.
- [11] P. J. Crist, T. W. Kohley, and J. Oakleaf, “Assessing land-use impacts on biodiversity using an expert systems tool,” *Landscape Ecology*, vol. 15, no. 1, pp. 47–62, 2000.
- [12] I. Roustafar, M. Sarif, R. Gupta et al., “Spatiotemporal analysis of land use/land cover and its effects on surface urban heat island using Landsat data: a case study of metropolitan city Tehran (1988–2018),” *Sustainability*, vol. 10, no. 12, p. 4433, 2018.
- [13] Y. Weng, “Spatiotemporal changes of landscape pattern in response to urbanization,” *Landscape and Urban Planning*, vol. 81, no. 4, pp. 341–353, 2007.
- [14] D. Dissanayake, T. Morimoto, Y. Murayama, and M. Ranagalage, “Impact of landscape structure on the variation of land surface temperature in sub-Saharan region: a case study of Addis Ababa using Landsat data (1986–2016),” *Sustainability*, vol. 11, no. 8, p. 2257, 2019.
- [15] H. Shen, L. Huang, L. Zhang, P. Wu, and C. Zeng, “Long-term and fine-scale satellite monitoring of the urban heat island effect by the fusion of multi-temporal and multi-sensor remote sensed data: a 26-year case study of the city of Wuhan in China,” *Remote Sensing of Environment*, vol. 172, pp. 109–125, 2016.
- [16] S. Chapman, J. E. M. Watson, A. Salazar, M. Thatcher, and C. A. McAlpine, “The impact of urbanization and climate change on urban temperatures: a systematic review,” *Landscape Ecology*, vol. 32, no. 10, pp. 1921–1935, 2017.
- [17] Y. Cai, H. Zhang, P. Zheng, and W. Pan, “Quantifying the impact of land use/land cover changes on the urban heat island: a case study of the natural wetlands distribution area of Fuzhou City, China,” *Wetlands*, vol. 36, no. 2, pp. 285–298, 2016.
- [18] D. Armson, “The effect of tree shade and grass on surface and globe temperatures in an urban area,” *Urban Forestry and Urban Greening*, vol. 11, no. 3, pp. 245–255, 2012.
- [19] S. Cui and H. H. Y. Hong, “Progress of the ecological security research,” *Acta Ecologica Sinica*, vol. 4, pp. 861–866, 2015.
- [20] A. Cook, “From resource scarcity to ecological security: exploring new limits to growth,” *Environmental Health Perspectives*, vol. 114, p. A190, 2006.
- [21] A. Arnfield, “Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island,” *International Journal of Climatology*, vol. 23, no. 1, pp. 1–26, 2003.
- [22] M. H. Saputra and H. Lee, “Prediction of land use and land cover changes for North Sumatra, Indonesia, using an artificial-neural-network-based cellular automaton,” *Sustainability*, vol. 11, no. 11, p. 3024, 2019.
- [23] G. Chander and B. Markham, “Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 11, pp. 2674–2677, 2003.
- [24] Z. Dai, J. M. Guldmann, and Y. Hu, “Spatial regression models of park and land-use impacts on the urban heat island in central Beijing,” *Science of the Total Environment*, vol. 626, pp. 1136–1147, 2018.
- [25] E. Windahl and K. . Beurs, “An intercomparison of Landsat land surface temperature retrieval methods under variable atmospheric conditions using in situ skin temperature,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 51, pp. 11–27, 2016.
- [26] Q. Weng, M. K. Firozjaei, M. Kiavarz, S. K. Alavipanah, and S. Hamzeh, “Normalizing land surface temperature for environmental parameters in mountainous and urban areas of a cold semi-arid climate,” *Science of The Total Environment*, vol. 650, Part 1, pp. 515–529, 2019.
- [27] R. Yan, Y. Cai, C. Li, X. Wang, and Q. Liu, “Hydrological responses to climate and land use changes in a watershed of the Loess Plateau, China,” *Sustainability*, vol. 11, no. 5, p. 1443, 2019.
- [28] L. Chu, T. Sun, T. Wang, Z. Li, and C. Cai, “Evolution and prediction of landscape pattern and habitat quality based on CA-Markov and InVEST model in Hubei section of Three Gorges Reservoir Area (TGRA),” *Sustainability*, vol. 10, no. 11, p. 3854, 2018.
- [29] Y. Yu, M. Yu, L. Lin et al., “National green GDP assessment and prediction for China based on a CA-Markov land use simulation model,” *Sustainability*, vol. 11, no. 3, p. 576, 2019.
- [30] Z. Qiao, G. Tian, L. Zhang, and X. Xu, “Influences of Urban Expansion on Urban Heat Island in Beijing during 1989–2010,” *Advances in Meteorology*, vol. 2014, 11 pages, 2014.
- [31] Z.-l. Hu, D. U. Pei-jun, and D.-z. Guo, “Analysis of urban expansion and driving forces in Xuzhou City based on remote sensing,” *Journal of China University of Mining and Technology*, vol. 17, no. 2, pp. 267–271, 2007.
- [32] Y. Yang, Y. Liu, Y. Li, and G. Du, “Quantifying spatio-temporal patterns of urban expansion in Beijing during 1985–2013 with rural-urban development transformation,” *Land Use Policy*, vol. 74, pp. 220–230, 2018.
- [33] G. Shi, N. Jiang, Y. Li, and B. He, “Analysis of the dynamic urban expansion based on multi-sourced data from 1998 to 2013: a case study of Jiangsu Province,” *Sustainability*, vol. 10, no. 10, p. 3467, 2018.
- [34] P. Priyankara, M. Ranagalage, D. Dissanayake, T. Morimoto, and Y. Murayama, “Spatial process of surface urban heat island in rapidly growing Seoul metropolitan area for sustainable urban planning using Landsat data (1996–2017),” *Climate*, vol. 7, no. 9, p. 110, 2019.
- [35] M. Ranagalage, Y. Murayama, D. Dissanayake, and M. Simwanda, “The impacts of landscape changes on annual mean land surface temperature in the tropical mountain city of Sri Lanka: a case study of Nuwara Eliya (1996–2017),” *Sustainability*, vol. 11, no. 19, p. 5517, 2019.

Research Article

Spatiotemporal Fusion of Remote Sensing Image Based on Deep Learning

Xiaofei Wang  and Xiaoyi Wang

School of Electronic Engineering, Heilongjiang University, Harbin 150080, China

Correspondence should be addressed to Xiaofei Wang; nk_wxf@hlju.edu.cn

Received 23 April 2020; Revised 21 May 2020; Accepted 17 June 2020; Published 29 June 2020

Academic Editor: Sang-Hoon Hong

Copyright © 2020 Xiaofei Wang and Xiaoyi Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

High spatial and temporal resolution remote sensing data play an important role in monitoring the rapid change of the earth surface. However, there is an irreconcilable contradiction between the spatial and temporal resolutions of the remote sensing image acquired from a same sensor. The spatiotemporal fusion technology for remote sensing data is an effective way to solve the contradiction. In this paper, we will study the spatiotemporal fusion method based on the convolutional neural network, which can fuse the Landsat data with high spatial but low temporal resolution and MODIS data with low spatial but high temporal resolution, and generate time series data with high spatial resolution. In order to improve the accuracy of spatiotemporal fusion, a residual convolution neural network is proposed. MODIS image is used as the input to predict the residual image between MODIS and Landsat, and the sum of the predicted residual image and MODIS data is used as the predicted Landsat-like image. In this paper, the residual network not only increases the depth of the superresolution network but also avoids the problem of vanishing gradient due to the deep network structure. The experimental results show that the prediction accuracy by our method is greater than that of several mainstream methods.

1. Introduction

Due to the limitation of the hardware technology of the remote sensing satellite and the cost of satellite launching, it is difficult for the same satellite to obtain the remote sensing image with both high spatial and temporal resolutions. Landsat series satellites can obtain multispectral data with a spatial resolution of 30 m. While multispectral images reflect the spectral information of ground features, when performing classification and other processing, unlike hyperspectral, which has rich dimensions, dimensionality reduction processing is required. Although there are many dimensionality reduction methods, it can achieve dimensionality reduction of hyperspectral images [1]. But multispectral image imaging is more convenient, making it widely used in many fields. With this feature, Landsat data has been widely used in the exploration of earth resources, management of agriculture, forestry, animal husbandry, and natural disaster and environmental pollution monitoring [2–4]. However, the 16-day visit circle of the Landsat satellite and the impact of cloud

pollution limit their potential use in monitoring and researching the land surface dynamic changes. On the other hand, Moderate-resolution Imaging Spectroradiometer (MODIS) on Terra/Aqua satellite has a visit circle per 1–2 days, which has a high temporal resolution and can be applied in vegetation phenology [5, 6] and other fields. However, the spatial resolution of MODIS data is 250–1000 m, which has a poor representation of the details of the ground objects and is not enough to observe the heterogeneous landscape.

In 1995, Vignolles et al. [7] first proposed to generate high spatiotemporal resolution data by using spatiotemporal fusion technology. Since then, different types of spatiotemporal fusion methods have been emerging. The spatiotemporal fusion technology of remote sensing images is fused with the spatial features of high spatial but low temporal resolution images and the temporal features of low spatial but high temporal resolution images to generate time series images with high spatial resolution. According to the principle, the existing spatiotemporal fusion models can be divided into three

types: reconstruction based, spatial unmixing based, and learning based.

The basic principle of reconstruction-based methods is to calculate the reflectance of the center fusion pixel through a weighting function which takes full account of the spectral, temporal, and spatial information in similar pixels. Gao et al. [8] first proposed the spatiotemporal adaptive reflection fusion model (STARFM), which uses a pair of MODIS and ETM+ reflectance images at known time phase and MODIS reflectance images at predicted time phase to generate 30 m spatial resolution image. Hilker et al. [9] proposed a spatiotemporal fusion algorithm for mapping reflectance change (STAARCH) based on tasseled cap change. The algorithm can not only generate 30 m spatial resolution ETM+ like images but also detect highly detailed surface classes. However, the fusion accuracy of STARFM and STAARCH are highly related to the surface landscape heterogeneity, resulting in low fusion accuracy for heterogeneous area. David et al. [10] considered the influence of bidirectional reflectance effect, proposing a semiphysical method to generate fused Landsat ETM+ reflectance using MODIS and Landsat ETM+ data. Zhu et al. [11] based on STARFM considering the reflectivity difference between different sensor imaging systems due to different orbital parameters, band bandwidth, spectral response curve and other factors, the transfer coefficient between different sensor differences is increased, and the enhanced STARFM (ESTARFM) model is proposed to improve the fusion accuracy of complex surface area (heterogeneous area) to a certain extent. The model uses two sets of MODIS and ETM+ reflectance images and MODIS reflectance images to generate 30 m spatial resolution ETM+ like image. Wang and Atkinson [12] proposed a spatiotemporal fusion algorithm consisting of three parts: regression model fitting (RM Fitting), spatial filtering (SF), and residual compensation (RC), referred to as Fit-FC; this method only uses a pair of known high-low resolution image pair as input and can better predict the spatial change between images in different periods. Chimani et al. [13] proposed a simple and intuitive method and has two steps. First, a mapping is established between two MODIS images where one is at an earlier time, t_1 , and the other one is at the time of prediction, t_p . Second, this mapping is applied to a known Landsat image at t_1 to generate a predicted Landsat image at t_p .

Spatial-temporal fusion methods based on spatial unmixing performs spatial unmixing of pixels in known low-resolution images and applies the classification results to high-resolution images at the unknown time to predict high-resolution images. Zhukov et al. [14] proposed a spatiotemporal fusion method that considers the spatial variability of pixel reflectivity based on the assumption that the pixel reflectivity does not change drastically between neighbor pixels. This method introduces window technology to predict the high-resolution reflectance of each type of feature. This method is not ideal for the farmland area which changes dramatically in a short time. Wu [15] proposed a spatial and temporal data fusion approach (STDFA) based on the assumption that the temporal variation characteristics of the class reflectivity are consistent with the intraclass pixel reflectivity. This method extracts the temporal change infor-

mation of ground features from time-series low spatial resolution images and performs classification and density segmentation on known two periods of high spatial resolution images to obtain classified images, so as to obtain class average reflectance for image fusion. On the basis of [15], Wu and Huang [16] comprehensively considered the spatial variability and temporal variation of pixel reflectivity and proposed an enhanced STDFA; the method solves the problem of missing remote sensing data. Hazaymeh and Hassan [17] proposed a relatively simple and more efficient algorithm; the Spatiotemporal Image-Fusion Model (STI-FM) applies clustering to the images first, and, for each cluster, performs a separate prediction. Zhu et al. [18] proposed a flexible spatiotemporal data fusion (FSDAF) based on spectral demixing analysis and thin-plate spline interpolation. The algorithm uses less input data, which is suitable for heterogeneous areas and can effectively preserve the low-resolution details of the image during the prediction period.

In remote sensing image processing, learning-based methods are more commonly used for classification of ground features [19]. In recent years, spatiotemporal fusion methods based on learning have been widely concerned. In 2012, Huang and Song [20] first introduced sparse representation technology into the process of spatiotemporal fusion and proposed a sparse representation based on a spatiotemporal reflectance fusion model (SPSTFM), which uses the MODIS and ETM+ images of the front and back phases at the predicted time phase. First, use high- and low-resolution difference images to train a couple dictionary representing high- and low-resolution features, and then use a low-resolution image to predict high-resolution image. Song and Huang [21] proposed a sparse representation of spatiotemporal reflectance fusion model using only a pair of known high- and low-resolution image pair, which first enhanced MODIS image by sparse representation to obtain a transition image, then predicted image is generated by combining known high-resolution image with transition image through high-pass modulation. The model reduces the number of known image pair that needs to be inputted, so that the algorithm can be applied in the case of lack of data and has more general applicability. Spatiotemporal fusion method based on feature learning considers the spatial information of changing image. However, there are some limitations in previous methods based on sparse representation. First, the image features need to be designed, which brings complexity and instability to performance. Secondly, the method does not consider the large amount of actual remote sensing data but only develops and validates the algorithm for small-scale research areas.

The convolutional neural network (CNN) [22] model has a simple structure and can be used to solve the problems of target recognition [23] and image classification [24] in computer vision. In recent years, CNN has also been used in the field of superresolution. As the pioneer CNN model for SR, superresolution convolutional neural network (SRCNN) [25] predicts the nonlinear LR-HR mapping via a fully convolutional network and significantly outperforms classical non-DL methods. In the field of remote sensing, Song et al. [26] proposed a five-layer convolutional neural network

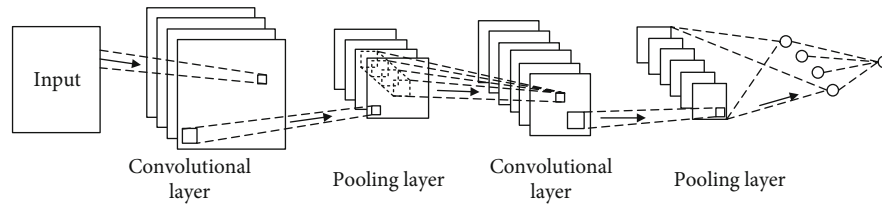


FIGURE 1: Flowchart of CNN network structure.

(CNN) spatiotemporal fusion model. This model is similar to [21] and is a two-stage model. It learns the CNN nonlinear mapping between MODIS and Landsat images and combines high-pass modulation with a weighting strategy to predict Landsat-like images. Liu et al. [27] proposed a two-stream convolutional neural network *StfNet*, which not only considered the temporal dependence of remote sensing images but also introduced temporal constraint, the network takes a coarse difference image with the neighboring fine image as inputs and the corresponding fine difference image as output, the method can restore spatial details greater. At present, there are two main problems faced by learning-based spatiotemporal fusion methods: first, the deep-seated network can improve the prediction accuracy; however, the deep-seated network will lead vanishing gradient or convergence difficulty and second, it is difficult to obtain two pairs of suitable prior image pairs as the input of network training. For example, *StfNet* is a fusion method using two pairs of prior images as input. Considering the above two points, we propose a spatiotemporal fusion model based on residual convolution neural network. The model can only uses a pair of prior images as train input. The MODIS image is very similar to the predicted Landsat image. In other words, the low-frequency information of low-resolution image is similar to that of high-resolution image. In fact, the low-resolution image and the high-resolution image only lack the residual of the high-frequency part. If only train the high-frequency residual between the high resolution and the low resolution, which does not need to spend too much time in the low-frequency part. And can deepen the network structure to avoid problems such as gradient disappearance. For this reason, we introduce the idea of ResNet [28] and set up a spatiotemporal fusion framework of remote sensing image suitable for a small-sample training set for CNN. Considering the time dependence between the image sequences, we use the MODIS-Landsat image pairs of the front and back phases of the prediction image to construct the prediction network, respectively. The experimental results show that compared with benchmark methods, the spectral color and spatial details of our method are closer to the real Landsat image.

The rest of this paper is divided into three sections. In Section 2, the principle of residual CNN is introduced. Section 3 provides the experimental verification process and results. Section 4 gives the conclusion.

2. Methods

In this paper, we use CNN and ResNet to construct a dual stream network to predict Landsat-like images. The principles involved are briefly introduced as follows.

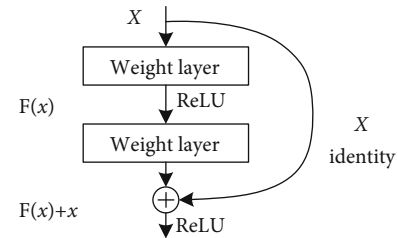


FIGURE 2: Residual learning unit.

2.1. CNN. Convolutional neural network (CNN) is one of the most representative network models in the deep learning method [29]. With the continuous development of deep learning techniques in recent years, it has achieved very good results in the field of image processing. Compared with the traditional data processing algorithm, CNN avoids the complicated preprocessing work such as manually extracting data from the data, so that it can be directly used in the original data.

CNN is a nonfully connected multilayer neural network, as shown in Figure 1. The main structure consists of a convolutional layer, pooling layer, activation layer, and fully connected layer [30]. The convolutional layer, pooling layer, and activation layer are the feature extraction layers of CNN, which are used to extract the signal features. The fully connected layer is the CNN classifier. Since this paper mainly uses the deep convolution network to extract the spatial characteristics of the remote sensing image, the feature extraction layer of deep convolutional neural networks is analyzed.

2.2. Residual Learning. If the input of a neural network is x , and the expected output is $H(x)$; $H(\cdot)$ is the expected mapping. If we want to learn such a model, the training difficulty will be greater; if we have learned the more saturated accuracy (or when we find that the error in the lower layer becomes larger), then the next learning goal will be transformed into the learning of identity mapping, that is, to make the input x an approximate output $H(x)$, which is in order to keep in the later hierarchy without causing a drop in accuracy.

As shown in the residual network structure diagram in Figure 2, input x is directly transferred to the output as the initial result through "shortcut connections," and the output result is $H(x) = F(x) + x$. When $F(x) = 0$, then $H(x) = x$, which is the constant mapping mentioned above. Therefore, ResNet is equivalent to changing the learning goal, not a complete output of learning, but the difference between the goal value $H(x)$ and x , that is, the so-called residual $F(x) =$

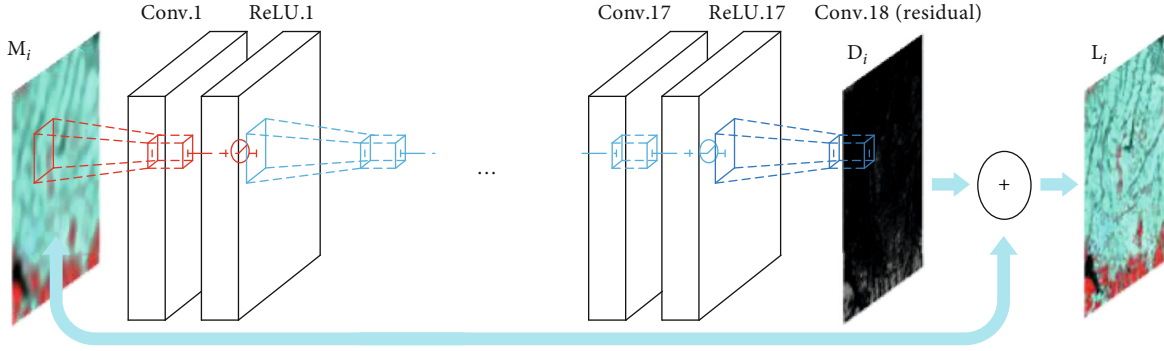


FIGURE 3: Flowchart illustrating the proposed scheme of two-stream residual learning in CNN.

$H(x) - x$. Therefore, the later training goal is to approach the residual result to 0, so that with the deepening of the network, the accuracy does not decline.

2.3. Spatiotemporal Fusion Using Residual Learning in CNN. In this paper, Landsat image is regarded as high spatial but low temporal resolution data; MODIS image is regarded as high temporal but low spatial resolution data. We express the Landsat image and MODIS image at t_i as \mathbf{L}_i and \mathbf{M}_i , respectively. If there are two pairs of prior images, the two-stream residual CNN network uses the known Landsat-MODIS image pair at t_1 and t_3 , and MODIS image at t_2 to predict Landsat-like image.

2.3.1. Training Stage. In the training stage, in order to build an nonlinear mapping model between MODIS and Landsat-MODIS residual images, we first up sample the spatial resolution of \mathbf{M}_i to the same size as \mathbf{L}_i . Then, the Landsat and MODIS images at the same time are differenced to obtain a residual image \mathbf{D}_i . Thus, we expect to learn a mapping function $f(x)$ which approximates \mathbf{D}_i . Pixel value in \mathbf{D}_i are likely to be zero or small. We want to predict this residual image. The loss function now becomes $1/2\|y - f(x)\|^2$, where $f(x)$ is the network prediction. We divide the high- and low-resolution images corresponding on the same time into overlapping image patches. Define the set of high- and low-resolution samples as \mathbf{X} and \mathbf{Y} , where the corresponding samples are \mathbf{x} and \mathbf{y} . The overlapping segmentation is performed here to increase the number of training samples. After predicting the residual image, the ground truth Landsat image is obtained by the sum of the input MODIS image and the predicted residual image.

In the network, the loss layer has three inputs: residual estimation, input MODIS image, and Landsat image. The loss is calculated as the Euclidean distance between the reconstructed image and the real Landsat image. In order to achieve the purpose of high-precision spatiotemporal fusion, we use a very deep convolutional network. We use 18 layers where layers except the first and the last are of the same type: 64 filters of the size $3 \times 3 \times 64$, where a filter operates on 3×3 spatial region across 64 channels (feature maps). The first layer operates on the input image. The last layer, used for image reconstruction, consists of a single filter of size $3 \times 3 \times 64$. The process structure is shown in Figure 3.

Training was performed by using back-propagation-based minibatch gradient descent to optimize regression targets. We set the momentum parameter to 0.9. Training is regularized by weight loss (l_2 penalty multiplied by 0.0001).

2.3.2. Prediction Stage. There are two pairs of prior Landsat-MODIS images and the MODIS image on prediction date, we aim to fuse them to predict the Landsat-like image on prediction date. Denote the prior dates as t_1 and t_3 , the prediction date as t_2 , we predict \mathbf{L}_2 based on the residual learning CNN. \mathbf{M}_i , \mathbf{L}_i , and \mathbf{D}_i are divided into patches, and their corresponding image patches are $\mathbf{m}_i^k, \mathbf{l}_i^k$, and \mathbf{d}_i^k , respectively. Taking \mathbf{m}_1^k as the input of CNN, the label is \mathbf{l}_1^k , and the sum of the residual image \mathbf{d}_1^k and \mathbf{m}_1^k is used as the prediction. In this paper, the number of network layers is set to 18. In the process of reconstruction, input \mathbf{m}_2^k into the trained network and get the predicted \mathbf{l}_2^k . Similarly, \mathbf{l}_2^{3k} can be predicted by Landsat-MODIS image pair at t_3 . Considering the temporal correlation between the image at the predicted time and the reference image, we use the corresponding temporal weight when reconstructing each image patch. Finally, the high spatial resolution image patch at the predicted time is obtained:

$$\mathbf{l}_2^k = \omega_1^k * \mathbf{l}_2^{1k} + \omega_3^k * \mathbf{l}_2^{3k}, \quad (1)$$

where \mathbf{l}_2^{1k} and \mathbf{l}_2^{3k} are the k th predicted patch using \mathbf{L}_1 and \mathbf{L}_3 as the reference image, respectively, ω_1^k and ω_3^k are the corresponding weight, and determined as follows:

$$\omega_i^k = \frac{1/v_i^k}{(1/v_1^k) + (1/v_3^k)} \quad (i = 1, 3). \quad (2)$$

The local weight is calculated by the sum U of normalized difference vegetation index (NDVI) [31] and normalized difference built-up index (NDBI) [32], where v_i^k is used to measure the change degree between MODIS images at two times, and it is the absolute average change of U in \mathbf{m}_{ij}^k , where \mathbf{m}_{ij}^k represents the MODIS image change at different times. After each image patch is reconstructed one by one, it is restored to the whole image. In order to ensure the continuity of the reconstructed image, there is an overlap between adjacent patches, and the pixel value of the overlapped part of the

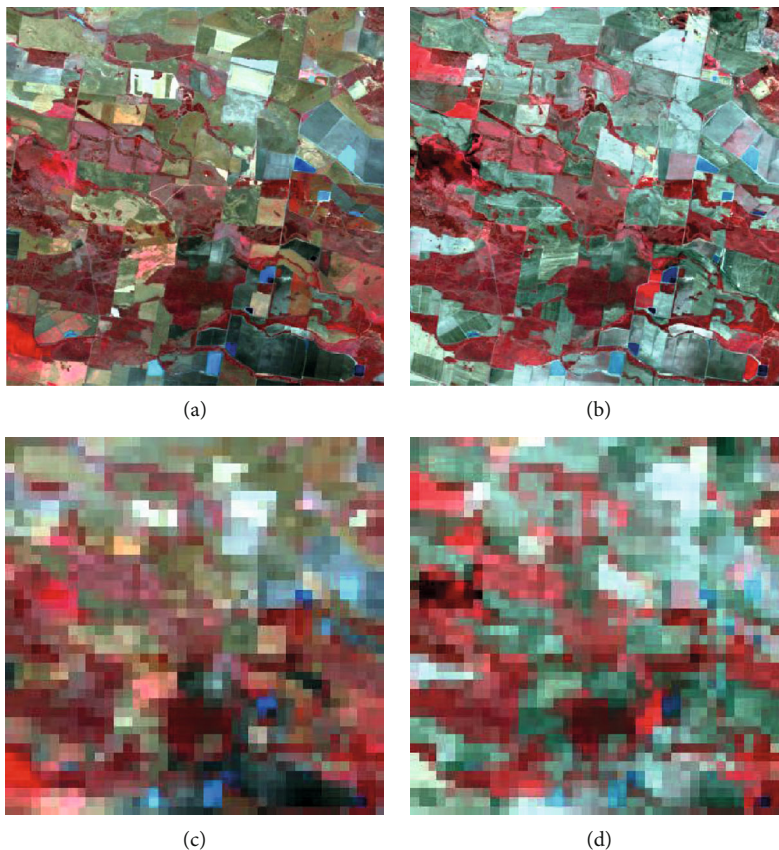


FIGURE 4: 30 m Landsat 8 and 500 m MODIS images for the first dataset (green, red, and NIR bands as RGB). (a) and (b) are 30 m Landsat images on 2 July 2013 and 17 August 2013, respectively, (c) and (d) are the corresponding 500 m MODIS images for (a) and (b).

image patch is taken as the mean value when the whole image is restored.

3. Experiments

Two datasets were used in the experiments. The first dataset contains two pairs of MODIS-Landsat images, and the second dataset contains three pairs of MODIS-Landsat images. Both areas are located in Coleambally, New South Wales, Australia. MODIS data uses the surface reflectance of MOD09A1 (500 m) and MOD09Q1 (250 m) for 8-day synthetic products. We up sampled all MODIS in the dataset to the same resolution as the Landsat image of the corresponding date. Compared with natural images, remote sensing images have a large size and rich details. Therefore, the remote sensing images are overlapped and divided into patches to obtain the training set. In the paper, the images of the two areas are overlapped divided into 33×33 image patches. The above image patch set is used as the train set and prediction set. We compare our method with the mainstream and advanced methods (including STARFM, FSDAF, Fit-FC, STDFA, STI-FM, HCM, ESTARFM, SPSTFM, and *StfNet*), which will be described in detail in this section.

3.1. Experiment on the First Dataset. In order to verify the applicability of our proposed spatial-temporal fusion method based on residual convolutional neural network for one prior

Landsat-MODIS image pair, we use a single-stream network to verify and compare the same data as the input of STARFM, FSDAF, Fit-FC, STDFA, STI-FM, and HCM.

In this experiment, two pairs of Landsat and MODIS surface reflectance images covering a $20 \text{ km} \times 20 \text{ km}$ area in Coleambally are used. The two pair images were acquired on 2 July 2013 and 17 August 2013. Figure 4 shows the 30 m Landsat images (upper row) and 500 m MODIS images (lower row) using green-red-NIR as RGB composite image. Then, we use the bicubic interpolation method to downscale the 500 m MODIS image into 30 m. Our experimental task used the pair of Landsat-MODIS images on 2 July 2013 and the MDOIS image on 17 August 2013 to predict the 30 m Landsat-like image on 17 August 2013. At the same time, STARFM, FSDAF, Fit-FC, STDFA, STI-FM, and HCM are tested with the same input in this experiment, and the true 30 m Landsat image acquired on 17 August 2013 is used as the reference to evaluate the accuracy of fusion results.

Figure 5 shows the fusion results by four methods (STARFM, Fit-FC, FSDAF, STDFA, STI-FM, HCM, and our method). Obviously, the prediction accuracy by our method is greater. For example, the highlighted areas in the bottom left part of subarea S, for Fit-FC, STDFA, FSDAF, STI-FM, HCM, and STARFM, some dark green pixels are incorrectly predicted as purple pixels. In addition, the highlighted areas in the bottom right part of the subarea, Fit-FC, STDFA, FSDAF, STI-FM, HCM, and STARFM

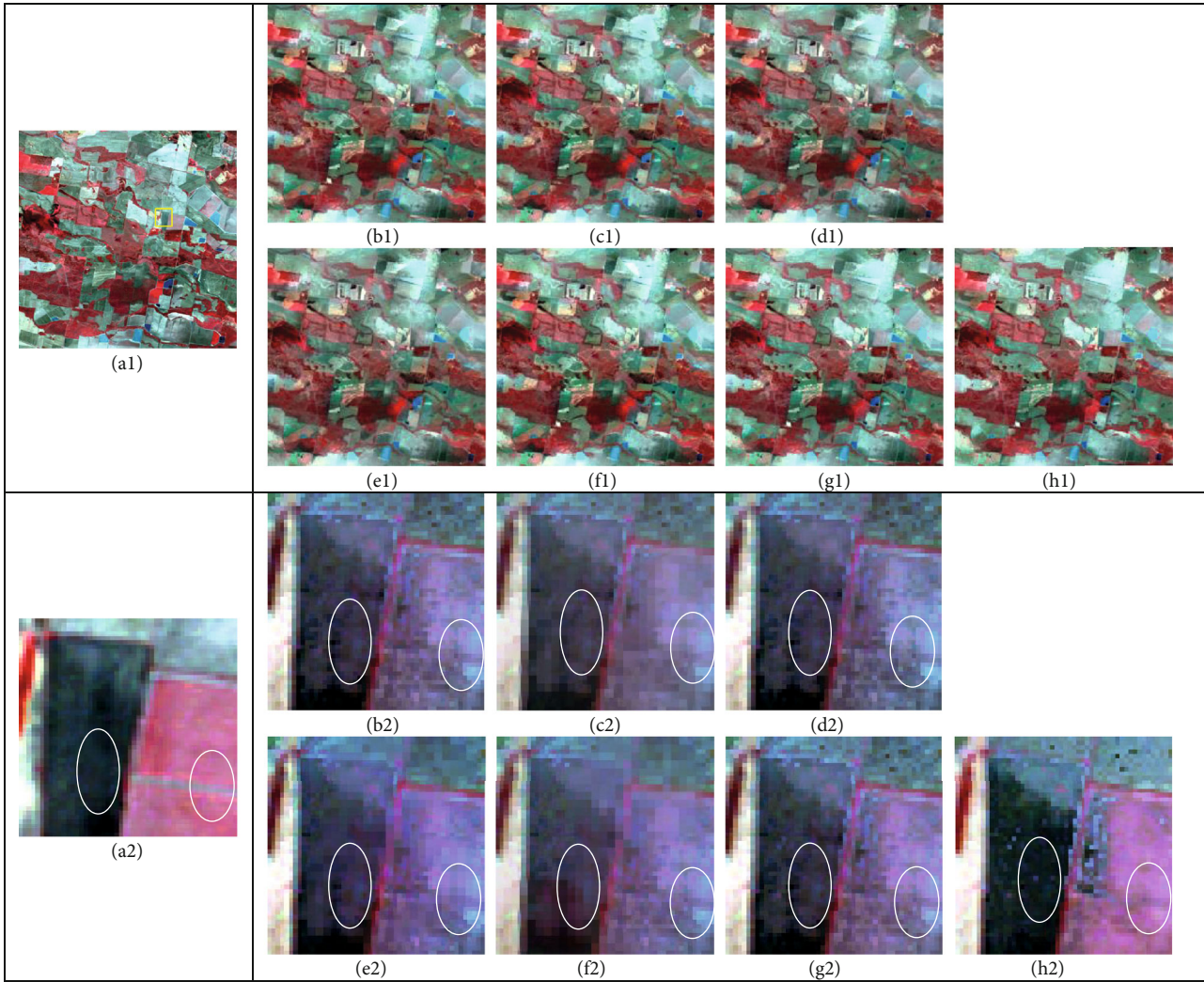


FIGURE 5: 30 m Landsat 8 results for the first dataset (green, red, and NIR bands as RGB). (a1) is the 30 m true Landsat 8 image on 17 August, 2013. (b1) is the 30 m Fit-FC-derived Landsat 8 images on 17 August, 2013. (c1) is the 30 m STDFA-derived Landsat 8 images on 17 August, 2013. (d1) is the 30 m FSDAF-derived Landsat 8 images on 17 August, 2013. (e1) is the 30 m STI-FM-derived Landsat 8 images on 17 August, 2013. (f1) is the 30 m HCM-derived Landsat 8 images on 17 August, 2013. (g1) is the 30 m STARFM-derived Landsat 8 images on 17 August, 2013. (h1) is the 30 m our method-derived Landsat 8 images on 17 August, 2013. Row 1 shows the results for the whole area, and Row 2 is the results for a heterogeneous subareas (S) marked in (a1).

incorrectly predicted some red pixels as purple and blue pixels. However, our method is closer to the reference image. The main reason is that the Fit-FC method directly applies the known linear coefficients of the low-resolution image to fit the high-resolution image on the prediction period. Therefore, when the spatial resolution difference between the high- and low-resolution images is large, there will be obvious “block effect,” for example, the spatial resolution difference between Landsat image and MODIS image is nearly 17 times. STDFA assumes that the temporal variation characteristics of the same surface coverage class in coarse pixels are consistent, but there may be inconsistencies in practical applications, so the fusion result is affected. The accuracy of the FSDAF spatiotemporal fusion algorithm is low, which is mainly caused by two aspects: The prediction accuracy of FSDAF is worse, which is mainly caused by two aspects: at first, the known high-resolution data needs to be classified,

and the classification accuracy by unsupervised classification method (such as the K -means method) will have a certain impact on the results; at second, when the spatial resolution difference between high- and low-resolution data is large, the endmember (that is, high-resolution pixel) represented area will be more refined. When the number of categories is small, the fusion result will be relatively smooth, and when the number of categories is large, the fitting accuracy will also be reduced (such as in low-resolution pixels, if the richness of a certain category is low, the total prediction error will be increase). STI-FM is susceptible to interference from outliers, so when the spatial characteristics change significantly, the prediction effect is not good. The method of using gradation mapping is greatly affected by the heterogeneous region, so HCM failed to show the best performance in this experiment. STARFM considers the similarity of neighboring pixels, so the prediction accuracy is relatively stable. However, the

TABLE 1: Quantitative assessment for the Coleambally dataset.

	Bands	Fit-FC	STDFA	FSDAF	STI-FM	HCM	STARFM	Our method
RMSE	Blue	0.0088	0.0089	0.0091	0.0098	0.0094	0.0086	0.0081
	Green	0.0113	0.0115	0.0117	0.0124	0.0120	0.0110	0.0105
	Red	0.0150	0.0152	0.0154	0.0164	0.0159	0.0146	0.0139
	NIR	0.0240	0.0244	0.0248	0.0268	0.0258	0.0233	0.0219
	SWIR1	0.0328	0.0331	0.0339	0.0363	0.0351	0.0318	0.0299
	SWIR2	0.0308	0.0317	0.0319	0.0344	0.0331	0.0299	0.0279
	Mean	0.0205	0.0208	0.0211	0.0227	0.0219	0.0199	0.0187
CC	Blue	0.8761	0.8773	0.8674	0.8463	0.8574	0.8833	0.8967
	Green	0.8786	0.8928	0.8705	0.8510	0.8612	0.8853	0.8976
	Red	0.8925	0.8931	0.8856	0.8687	0.8775	0.8982	0.9085
	NIR	0.7713	0.7722	0.7550	0.7171	0.7367	0.7850	0.8098
	SWIR1	0.8327	0.8331	0.8207	0.7925	0.8072	0.8429	0.8627
	SWIR2	0.8393	0.8398	0.8269	0.7978	0.8130	0.8497	0.8697
	Mean	0.8484	0.8489	0.8377	0.8122	0.8255	0.8574	0.8742
UIQI	Blue	0.8653	0.8662	0.8564	0.8350	0.8462	0.8728	0.8867
	Green	0.8680	0.8687	0.8597	0.8400	0.8503	0.8750	0.8879
	Red	0.8835	0.8841	0.8763	0.8591	0.8681	0.8895	0.9008
	NIR	0.7644	0.7652	0.7488	0.7125	0.7313	0.7775	0.8011
	SWIR1	0.8251	0.8259	0.8132	0.7855	0.7999	0.8352	0.8557
	SWIR2	0.8313	0.8319	0.8190	0.7903	0.8053	0.8418	0.8622
	Mean	0.8396	0.8403	0.8289	0.8037	0.8168	0.8486	0.8657

premise of STARFM is that the spectrum of similar pixels in the neighborhood is constant and there is no land cover change during the observation period, which makes the model susceptible to environmental and phenological changes, resulting in large prediction errors, especially for heterogeneous areas. Our method uses deep convolutional neural networks to more effectively extract the features of low-resolution images and residual images and constructs a mapping relationship between low-resolution images and residual images through a residual learning network. This mapping relationship is nonlinear mapping and is more in line with the change of ground features. In addition, the number of layers in the network is deepened through residual learning, which strengthens the robustness of the network. Therefore, the experimental results based on our method have better visual effects.

Table 1 lists the objective evaluation results of four fusion methods and uses three common fusion evaluation methods of remote sensing image, including root mean square error (RMSE) [33], correlation coefficient (CC) [34], and universal image quality index (UIQI) [35]. The ideal values for RMSE, CC, and UIQI are 0, 1, and 1, respectively. From Table 1, we can see that for the six bands of all fusion results, the fusion results of our method have smaller RMSE and larger CC and UIQI. Our method is compared with other six methods (STARFM, Fit-FC, FSDAF, STDFA, STI-FM, and HCM); the gain of the mean CC is 0.0259, 0.0365, 0.0168, 0.0253, 0.0620, and 0.0487, and the gain of the mean UIQI is 0.0261, 0.0368, 0.0175, 0.0254, 0.0620, and 0.0489, respec-

tively. The mean RMSE is reduced by 0.0018, 0.0024, 0.0012, 0.0021, 0.0040, and 0.0032, respectively. In addition, the fusion result based on our method is better than STATFM, and STARFM is better than Fit-FC, the rest of the sequence is STDFA>FSDAF>HCM>STI-FM. The main reason is when the spatial resolution difference between high- and low-resolution images is large, Fit-FC directly applies the fitting coefficients of low-resolution images into high-resolution images, which causes large errors; FSDAF also has similar fitting errors. STDFA assumes that the temporal variation characteristics of the same surface coverage class in coarse pixels are consistent, but there may be inconsistencies in practical applications, so the fusion result is affected. Although STARFM considers the similarity of neighboring pixels, the reconstruction method of each pixel cannot consider the continuity of the image. STI-FM is susceptible to interference from outliers, so when the spatial characteristics change significantly, the prediction effect is not good. HCM using gradation mapping is greatly affected by the heterogeneous region. Our method can better restore the continuity of the image by reconstructing the image patch.

3.2. Experiment on the Second Dataset. In this experiment, three pairs of Landsat-MODIS images covering 30 km × 30 km area of Coleambally are used to verify the applicability of our method for two pairs of prior images. The three pairs of images were acquired on 6 April, 2012, 12 May, 2012, and 20 July, 2012, respectively. Figure 6 shows the 30 m Landsat

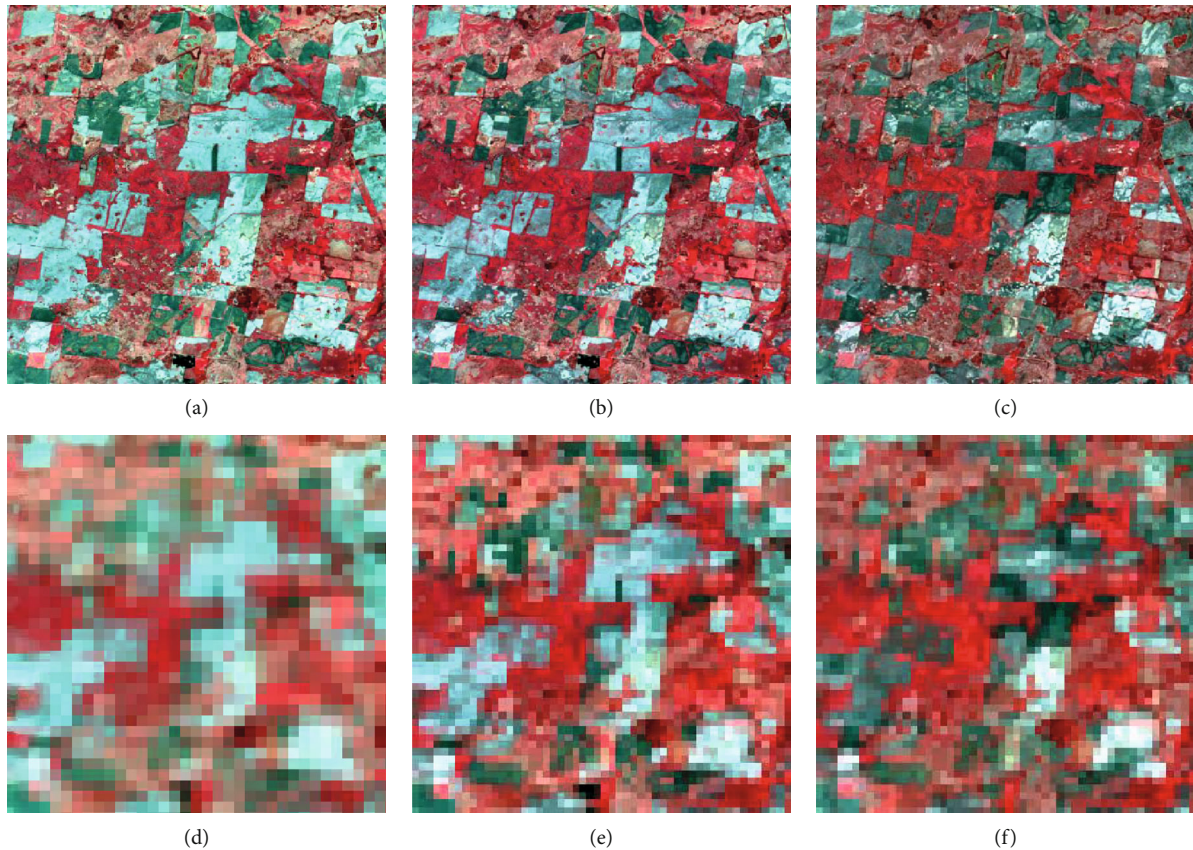


FIGURE 6: 30 m Landsat and 500 m MODIS images for the second dataset (green, red, and NIR bands as RGB). (a), (b), and (c) are 30 m Landsat images on 6 April, 2012, 12 May, 2012, and 20 July, 2012, respectively, and (d–f) are the corresponding 500 m MODIS images for (a–c).

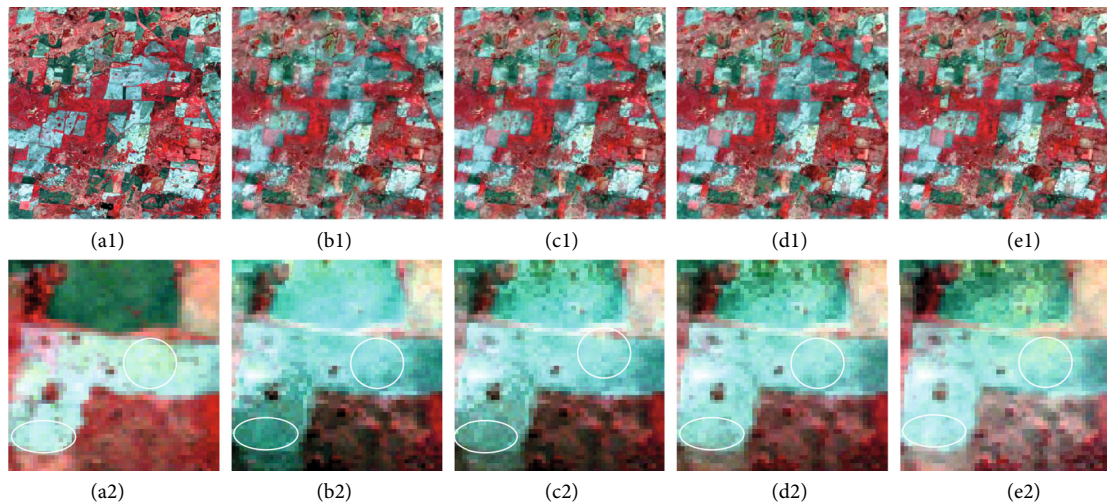


FIGURE 7: 30 m Landsat results for the second dataset (green, red, and NIR bands as RGB). (a1) is the 30 m true Landsat image on July 11, 2012. (b1) is the 30 m ESTARFM-derived Landsat images on 12 May, 2012. (c1) is the 30 m SPSTFM-derived Landsat images on 12 May, 2012. (d1) is the 30 m *StfNet*-derived Landsat images on 12 May, 2012. (e1) is the 30 m our method-derived Landsat images on 12 May, 2012. Row 1 shows the results for the whole area, and Row 2 is the results for the heterogeneous sub-areas (S1) marked in (a1).

image (upper row) and 500 m MODIS image (lower row) using green-red-NIR as RGB composite image. This experiment is to verify the accuracy of our methods based on two pairs of prior images, we use the two image pairs on 6 April,

2012 and 20 July, 2012, and MODIS image on 12 May, 2012 to predict the Landsat-like image on 12 May, 2012.

Figure 7 shows the 30 m prediction results on August 12, 2012 based on the four methods (ESTARFM, SPSTFM,

StfNet, and our method). It is worth noting that the ESTARFM result is the worst, *StfNet* is better than SPSTFM, and our method is better than *StfNet*. For example, the highlighted areas in the bottom left of subarea S, ESTARFM, and SPSTFM incorrectly predicted some light green pixels as dark green pixels. Although the prediction by *StfNet* is similar to the reference image, but there are some yellow pixels which had been incorrectly predicted to be blue pixels. However, for our method, the prediction is closer to the true reference image. Compared with the three benchmark methods, our method provides excellent performance. The main reason is ESTARFM assumes that during the observation period, the conversion coefficients between high- and low-resolution images remain unchanged, but in actual conditions, land types and coverage will change, so this assumption is not applicable in the areas with significant changes. SPSTFM utilizes sparse representation and dictionary learning approaches in the signal domain to increase prediction accuracy for land cover change and heterogeneous region. Although the network structure, compared with our method, SPSTFM only applicable for small-scale regions and cannot extract sufficient image features. Although *StfNet* can produce more accurate prediction results by deep network, but the data contained in the training process is too large and the network is hard to converge, which also has a certain impact on the prediction accuracy. Our method using residual learning network can only learn the difference information between high- and low-resolution images. Since the low-frequency information between high- and low-resolution images is similar, if we directly learn mapping relationship between high- and low-resolution images, it will increase the amount of calculation, which also introduces errors. Through residual learning, not only the nonlinear mapping relationship between high- and low-frequency information can be directly learned, but also the network layers can be deepened, which enhances the accuracy and stability of the network structure.

Table 2 shows the comparison results in RMSE, CC, and UIQI. From Table 2, we can see that for six bands, the fusion results by our method can obtain smaller average RMSE and larger CC and UIQI. It is easy to find that our method is better than *StfNet*, the *StfNet* is better than SPSTFM, and ESTARFM is the worst among the four approaches. Specifically, the CC gains of our method over ESTARFM, SPSTFM, and *StfNet* are 0.0508, 0.0257, and 0.0134, and the UIQI gains are 0.0467, 0.0238, and 0.0126, respectively. The main reason is that ESTARFM assumes that the conversion coefficients remain unchanged during the observation period, but there are land cover types in this area, such as subregion S, so the conversion coefficients are not consistent, so the prediction results are greatly biased. SPSTFM takes the image patch as the reconstruction unit and considers the continuity between adjacent pixels, so it has strong robustness in dealing with complex surface changes. However, due to the instability of forcing the same sparse coefficient of high- and low-resolution dictionaries to construct the mapping relationship, the performance in this experiment is worse than our method. *StfNet* has a deep network layers; however, it is difficult to converge

TABLE 2: Quantitative assessment for the second dataset.

	Bands	ESTARFM	SPSTFM	<i>StfNet</i>	Our method
RMSE	Blue	0.0118	0.0112	0.0109	0.0104
	Green	0.0139	0.0132	0.0128	0.0122
	Red	0.0217	0.0208	0.0203	0.0196
	NIR	0.0361	0.0332	0.0317	0.0301
	SWIR1	0.0306	0.0288	0.0279	0.0268
	SWIR2	0.0369	0.0356	0.0348	0.0340
	Mean	0.0252	0.0238	0.0231	0.0222
CC	Blue	0.8512	0.8727	0.8837	0.8964
	Green	0.8520	0.8729	0.8833	0.8951
	Red	0.8537	0.8756	0.8864	0.8990
	NIR	0.7131	0.7601	0.7824	0.8055
	SWIR1	0.8519	0.8735	0.8843	0.8955
	SWIR2	0.8725	0.8900	0.8987	0.9077
	Mean	0.8324	0.8575	0.8698	0.8832
UIQI	Blue	0.8425	0.8622	0.8723	0.8847
	Green	0.8431	0.8624	0.8719	0.8834
	Red	0.8297	0.8483	0.8574	0.8688
	NIR	0.7040	0.7500	0.7714	0.7940
	SWIR1	0.8455	0.8655	0.8753	0.8858
	SWIR2	0.8529	0.8667	0.8735	0.8810
	Mean	0.8196	0.8425	0.8537	0.8663

due to directly training the mapping relationship between high- and low-resolution images, which also leads to network instability. Our method through residual network not only improves the stability of the network but also enhances the accuracy of the fusion results.

4. Conclusion

In this paper, we propose a residual convolution neural network to predict Landsat-like image, and the method can be applied to the case where there is only a pair of prior images. This method mainly includes two steps: firstly, use the known MODIS-Landsat image pair to train the residual convolutional neural network and secondly, input MODIS image at predicted phase to reconstruct Landsat-like image. Compared with the several benchmark algorithms (STARFM, FSDAF, Fit-FC, ESTARFM, SPSTFM, and *StfNet*), our method has the advantages of learning algorithm, which takes the image patch as the reconstruction unit and considers the continuity between adjacent pixels. Training the residual to construct the depth network not only enhances the stability of the network but also improves the prediction accuracy.

The spatiotemporal fusion methods based on learning have greater prediction accuracy for heterogeneous regions. In this paper, we use a multilayer convolution neural network to extract spatial features. In the future work, we will try to design more effective methods to extract spatial features to improve the recognition ability of change information. In recent years, deep learning has received extensive attention.

Deep learning needs a lot of data to train model. Because of the characteristics of large amount of data and rich information in remote sensing data, we can use the “big data” characteristics of remote sensing data to train more effective mapping relationship between MODIS and Landsat images by deep learning training, so as to improve the prediction accuracy. In addition, although the spatiotemporal fusion models based on learning have outstanding performance, but the calculation time is longer, which is also a “common failure” based on the learning method. Therefore, our future work will follow the idea of improving the accuracy of fusion results and reducing computational complexity.

Data Availability

Data is not available for the following reasons: In this paper, we received remote sensing data from Institute of Remote Sensing Applications Chinese Academy of Sciences and conducted an experiment, but without the consent of Institute of Remote Sensing Applications Chinese Academy of Sciences, the author cannot judge whether data is available or not.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

- [1] B. Rasti, D. Hong, R. Hang et al., “Feature extraction for hyperspectral imagery: the evolution from shallow to deep (Overview and Toolbox),” *IEEE Geoscience and Remote Sensing Magazine*, vol. 8, no. 3, pp. 63–92, 2020.
- [2] M. C. Anderson, R. G. Allen, A. Morse, and W. P. Kustas, “Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources,” *Remote Sensing of Environment*, vol. 122, pp. 50–65, 2012.
- [3] F. D. van der Meer, H. M. A. van der Werff, F. J. A. van Ruitenbeek et al., “Multi- and hyperspectral geologic remote sensing: A review,” *International Journal of Applied Earth Observation & Geo information*, vol. 14, no. 1, pp. 112–128, 2012.
- [4] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, “Learnable manifold alignment (LeMA): a semi-supervised cross-modality learning framework for land cover and land use classification,” *ISPRS Journal of Photogrammetry & Remote Sensing*, vol. 147, pp. 193–205, 2019.
- [5] S. Ganguly, M. A. Friedl, B. Tan, X. Zhang, and M. Verma, “Land surface phenology from MODIS: characterization of the collection 5 global land cover dynamics product,” *Remote Sensing of Environment*, vol. 114, no. 8, pp. 1805–1816, 2010.
- [6] X. Zhang, M. A. Friedl, C. B. Schaaf et al., “Monitoring vegetation phenology using MODIS,” *Remote Sensing of Environment*, vol. 84, no. 3, pp. 471–475, 2003.
- [7] C. Vignolles, M. Gay, G. Flouzat, and P. Puyou-Lascassies, “Spatiotemporal connection of remote sensing data concerning agricultural production modelisation at a middle scale,” in *1995 International Geoscience and Remote Sensing Symposium, IGARSS '95. Quantitative Remote Sensing for Science and Applications*, Firenze, Italy, Italy, July 1995.
- [8] F. Gao, J. Masek, M. Schwaller, and F. Hall, “On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, pp. 2207–2218, 2006.
- [9] T. Hilker, M. A. Wulder, N. C. Coops et al., “A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS,” *Remote Sensing of Environment*, vol. 113, no. 8, pp. 1613–1627, 2009.
- [10] P. R. David, J. Junchang, L. Philip, S. Crystal, H. Matt, and L. Erik, “Mufti-temporal MODIS-Landsat fusion for relative and prediction of Landsat data,” *Remote Sensing of Environment*, vol. 112, no. 6, pp. 3112–3130, 2008.
- [11] X. Zhu, J. Chen, F. Gao, X. Chen, and J. G. Masek, “An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions,” *Remote Sensing of Environment*, vol. 114, no. 11, pp. 2610–2623, 2010.
- [12] Q. Wang and P. M. Atkinson, “Spatio-temporal fusion for daily Sentinel-2 images,” *Remote Sensing of Environment*, vol. 204, pp. 31–42, 2018.
- [13] C. Kwan, B. Budavari, F. Gao, and X. Zhu, “A hybrid color mapping approach to fusing MODIS and Landsat images for forward prediction,” *Remote Sensing*, vol. 10, no. 4, pp. 520–529, 2018.
- [14] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, “Unmixing-based multisensor multiresolution image fusion,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1212–1226, 1999.
- [15] M. Wu, “Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model,” *Journal of Applied Remote Sensing*, vol. 6, no. 13, p. 063507, 2012.
- [16] M. Wu, W. Huang, Z. Niu, and C. Wang, “Generating daily synthetic Landsat imagery by combining Landsat and MODIS data,” *Sensors*, vol. 15, no. 9, pp. 24002–24025, 2015.
- [17] K. Hazaymeh and Q. K. Hassan, “Spatiotemporal image-fusion model for enhancing the temporal resolution of Landsat-8 surface reflectance images using MODIS images,” *Journal of Applied Remote Sensing*, vol. 9, no. 1, p. 096095, 2015.
- [18] X. Zhu, E. H. Helmer, F. Gao, D. Liu, J. Chen, and M. A. Lefsky, “A flexible spatiotemporal method for fusing satellite images with different resolutions,” *Remote Sensing of Environment*, vol. 172, pp. 165–177, 2016.
- [19] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, “CoSpace: common subspace learning from hyperspectral-multispectral correspondences,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4349–4359, 2019.
- [20] B. Huang and H. Song, “Spatiotemporal reflectance fusion via sparse representation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3707–3716, 2012.
- [21] H. Song and B. Huang, “Spatiotemporal satellite image fusion through one-pair image learning,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 1883–1896, 2013.
- [22] D. Yu and L. Deng, “Deep learning and its applications to signal and information processing exploratory,” *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 145–154, 2011.
- [23] F. H. C. Tivive and A. Bouzerdoum, “A new class of convolutional neural Networks (SICoNNets) and their application of face detection,” in *Proceedings of the International Joint Conference on Neural Networks, 2003*, pp. 2157–2162, Portland, OR, USA, July 2003.
- [24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Image Net classification with deep convolutional neural networks,” in

- Proceedings of the 2012 Advances in Neural Information Processing Systems*, pp. 1097–1105, Lake Tahoe, Nevada, USA, 2012, Curran Associates, Inc..
- [25] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.
- [26] H. Song, Q. Liu, G. Wang, R. Hang, and B. Huang, “Spatiotemporal satellite image fusion using deep convolutional neural networks,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 821–829, 2018.
- [27] X. Liu, C. Deng, J. Chanussot, D. Hong, and B. Zhao, “StfNet: a two-stream convolutional neural network for spatiotemporal image fusion,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6552–6564, 2019.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, June 2016.
- [29] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [30] M. Liang and X. Hu, “Recurrent convolutional neural network for object recognition,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, June 2015.
- [31] E. F. Lambin and A. H. Strahler, “Indicators of land-cover change for change-vector analysis in multitemporal space at coarse spatial scales,” *International Journal of Remote Sensing*, vol. 15, no. 10, pp. 2099–2119, 2007.
- [32] C. He, P. Shi, D. Xie, and Y. Zhao, “Improving the normalized difference built-up index to map urban built-up areas using a semiautomatic segmentation approach,” *Remote Sensing Letters*, vol. 1, no. 4, pp. 213–221, 2010.
- [33] Z. Zhang and R. S. Blum, “A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application,” *Proceedings of the IEEE*, vol. 87, no. 8, pp. 1315–1326, 1999.
- [34] K. D. Kim and J. H. Heo, “Comparative study of flood quantiles estimation by nonparametric models,” *Journal of Hydrology*, vol. 260, no. 1-4, pp. 176–193, 2002.
- [35] Z. Wang and A. C. Bovik, “A universal image quality index,” *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

Research Article

Regional Patch Detection of Road Traffic Network

Xia Zhu , Weidong Song , and Lin Gao 

Cartography and Geographic Information Engineering, Liaoning Technical University, Fuxin 12300, China

Correspondence should be addressed to Xia Zhu; zhuxia1201@126.com

Received 30 January 2020; Revised 19 February 2020; Accepted 24 February 2020; Published 2 June 2020

Guest Editor: Sang-Hoon Hong

Copyright © 2020 Xia Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Road traffic network (RTN) structure plays an important role in the field of complex network analysis. In this paper, we propose a regional patch detection method from RTN via community detection of complex network. Firstly, the refined Adapted PageRank algorithm, which combines with the influence factors of the location property weight, the geographic distance weight and the road level weight, is used to calculate the candidate ranking results of key nodes in the RTN. Secondly, the ranking result and the shortest path distance as two significant impact factors are used to select the key points of the RTN, and then the Adapted K-Means algorithm is applied to regional patch detection of the RTN. Finally, based on the experimental data of Zhangwu road traffic network, the analysis results are as follows: Zhangwu is divided into 9 functional structures with key node locations as the core. Regional patch structure is divided according to key points, and the RTN is actually divided into nine small functional communities. Nine functional regional patches constitute a new network structure, maintaining connectivity between the regional patches can improve the overall efficiency of the RTN.

1. Introduction

Road traffic network is an important infrastructure to serve the economy, society, and the public and is the backbone of the comprehensive transportation system. China is a big country in road traffic, with a total length of 4.8465 million kilometers at all levels, which constitutes a complex network system. The community is the dominant structure that exhibits different features and multifold functions of complex networks; accordingly, community detection is of critical importance in network science. In recent years, Newman [1] and Boccaletti [2] introduce the structure, dynamics, and function of complex networks. Newman [3] introduces the communities, modules, and large-scale structure in networks. Communities have intrinsic interest, they may correspond to functional units within a networked system, an example of the kind of link between structure and function that drives much of the present excitement about networks. Costa et al. [4] analyze and model real-world phenomena with complex networks. Fortunato et al. [5] present that community detection in networks is one of the most popular topics of modern network science. Communities, or clusters,

are usually groups of vertices having higher probability of being connected to each other than to members of other groups, though other patterns are possible. Yang et al. [6, 7] adapt the concept of community detection to the correlation network of urban traffic state and propose a new perspective to identify the spatial correlation patterns of traffic state. Real networks exhibit heterogeneous nature with nodes playing far different roles in structure and function, to identify vital nodes is thus very significant [8]. Liao et al. [9] introduce the problem of ranking the nodes and the edges in complex networks that is critical for a broad range of real-world problems, because it affects how we access online information and products, how success and talent are evaluated in human activities, and how scarce resources are allocated by companies and policymakers, among others. Agryzkov et al. [10] provide a method to establish a ranking of nodes in an urban network, with the main characteristic that is able to consider the importance of data obtained from the urban networks in the process of computing the centrality of every node. Yu et al. [11] say the identification of clusters or communities in complex networks is a reappearing problem. Kim and Kim [12] propose an algorithm that uses an interaction

optimization process to detect community structures in complex networks.

As community detection of complex network is one of the research hotspots, we proposed a Regional patch detection (RPD) analysis. Based on the idea of community detection of complex network, the regional patch detection of road traffic network (RTN) is put forward, which transcends the restriction of administrative division on RTN and fundamentally identifies the special common regional patch structure of RTN. Regional patch detection is applied to the RTN, the deep characteristic structure aggregation of the RTN is analyzed, the restrictions of the administrative division on the RTN are crossed, and the characteristic sharing area is detected according to the trend of the road network, ancillary facilities of the road network, service facilities, distance, and other influencing factors. To evaluate the rationality of traffic network planning, the rationality of road network structure design and the maintenance and maintenance of traffic network are theoretical and practical problems that need to be solved urgently.

The remaining of this paper is organized as follows: Section 2 constructs the road traffic network (RTN), considering the influence factor of weight and the regional patch detection of road traffic network are introduced. Section 3 presents the experimental results of a road traffic network (RTN) in Zhangwu of China. Finally, conclusions are provided in the Conclusion section.

2. Regional Patch Detection (RPD) from Road Traffic Network (RTN)

2.1. The Construction of RTN. The structure of RTN is always defined as followed due to its complexity. The RTN defined:

$$RTN = (N, E) \quad (1)$$

whereas: $N = \{n_i | i = 1, 2, \dots, N\}$ is the set of nodes in the RTN, $N = |N|$ is the number of nodes in RTN; $E = \{e_{ij} | i \neq j, i, j \in \{1, 2, \dots, N\}\}$ is the set of edges. $e_{ij} = 1$, if there is an edge between nodes i and j , otherwise $e_{ij} = 0$; weight factor: if each edge is given a corresponding weight, the network is called a weighted network; otherwise, it is called an unweighted network. An unweighted network can also be regarded as an equal-weight network in which each edge has a weight of 1.

2.2. The Weighting Factors. The influence of features, distance, and level of road traffic network play a key role in the geographical characteristics and traffic operation of road network. So we set the corresponding weight influence factors to analyze the core position distribution of the road traffic network. There are three weighting factors for the RTN:

2.2.1. The Location Property. The location property is an important factor in measuring the importance of a segment. Location characteristics are often related to the presence of facilities such as those in the hotel industry and the commercial sector. Suppose vector v is of size $N \times 1$, N is the number of road nodes in the RTN. An element $v_i = 1 (1 \leq i \leq N)$ that

TABLE 1: Road level weight distribution.

Road level	National road (G)	Provincial road (S)	County road (X)	Rural road (Y)	Village road (C)
Weight	0.6875	0.1875	0.075	0.025	0.025

represents the segment i is relatively important, and there are important ground feature service facilities around it. Otherwise, $v_i = 0$. The property matrix F defined as:

$$F = (v, v, \dots, v)_{1 \times N} \quad (2)$$

2.2.2. The Geographic Distance. For the geographical, a distance-decay characteristic displays significantly in the RTN. The correlation amplitude generally decreases with the increase of the distance between two nodes. Assumption, distance weight matrix $W = (w_{ij})_{i,j=1}^N$, d_{ij} is the shortest path length between segments i and j in RTN and the unit is km:

$$w_{ij} = \begin{cases} \frac{1}{d_{ij}} & i \neq j \\ 0 & i = j \end{cases}, 1 \leq i, j \leq N \quad (3)$$

2.2.3. The Road Level. The traffic road network is classified to five categories according to the road administrative level, which is divided into national road (G), provincial road (S), county road (X), rural road (Y), and village roads (C). According to the traffic volume, task, and nature of China's highway engineering technical standards (JTG B01-2014), different levels of traffic networks have different annual average daily traffic volume (ADT) [13]. Thus, the weights of different levels of traffic network are set as shown in Table 1.

Suppose, road level matrix $L = (l_{ij})_{i,j=1}^N$, element l_{ij} refers to the weighted value of road level in Table 1.

2.3. Regional Patch Detection (RPD). This community detection result is as a reference, we define the concept of geographic space road traffic network regional patch detection (RPD):

The regional patch is a subset of the RTN in the range of dividing the RTN at different levels. A regional patch contains nodes and the sections between nodes. The nodes in the regional patch are closely related to each other and have strong topological structure similarity and feature facility property similarity. The relation between regional patches is relatively sparse, and the close regional patches differ greatly, while the distant regional patches may have similar characteristics.

2.4. RPD Method. Complex networks have scale-free, small-world, and community characteristics. The characteristics of RTN are generally scale-free and small-world. Therefore, the community structure of RTN is analyzed based on the complex network, and the aggregation of road network is deeply explored.

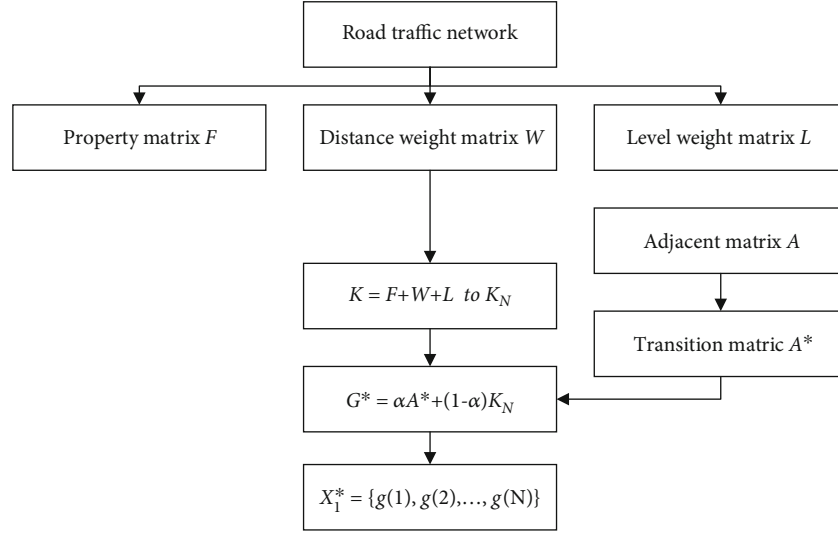


FIGURE 1: The Adapted PageRank algorithm.

TABLE 2: The Adapted K-Means Algorithm process.

Input: Road traffic network $N = \{n_1, n_2, \dots, n_m\}$; Regional patch center (Decision Graph) k .
Process:
1: Select the regional patch center from the RTN as the initial value $\{n_1, n_2, \dots, n_k\}$
2: repeat
3: $C_i = \emptyset (1 \leq i \leq k)$
4: for $i = 1, 2, \dots, n$ do
5: $\arg \min_{C_k \in C} \ \hat{S}_i - C_k\ ^2, i = 1, 2, \dots, N$
6: $C_k = 1/ Z_k \sum_{\hat{S}_i \in Z_k} \hat{S}_i, k = 1, 2, \dots, K$
7: Divide the sample n_i into corresponding regional patch $C_{\lambda_i} = C_{\lambda_i} \cup \{x_i\}$
8: end for
Export: Regional patch detection $C = \{C_1, C_2, \dots, C_k\}$

The Google matrix of PageRank, expressed in G , is defined as:

$$G = \alpha A^* + \frac{(1-\alpha)}{N} I_{N \times N} \quad (4)$$

In the paper, the refined Adapted PageRank algorithm is used to calculate the ranking results of key nodes of the RTN. The matrix $(1 - \alpha/N)I_{N \times N}$ by adding a property matrix, a distance weight matrix, and a road level matrix. The weight influence factor set in Section 2.2 is applied to calculate the candidate location distribution of key nodes in the RTN.

Then, with the property matrix F , the distance weight matrix W , and the road level matrix L , a novel weighted

matrix K is defined, where k_j is the j th column of matrix K :

$$K = (k_1, k_2, \dots, k_j, \dots, k_N) = F + W + L \quad (5)$$

In order to make matrix K have column irreducibility and stochastically, every column vector k_j of matrix K should be normalized, Where $\|k_j\|$ is the norm of column vector k_j .

A new matrix G^* in Equation. (6), the term $(1 - \alpha/N)I_{N \times N}$ is replaced by $(1 - \alpha)K_N$. Calculate $X_1^* = \{g(1), g(2), \dots, g(N)\} (\lambda = 1)$ of G^* are the rank of key nodes.

$$G^* = \alpha A^* + (1-\alpha)K_N \quad (6)$$

The calculation rank of key nodes is described in Figure 1.

In this paper, the center points of regional patch detection are selected by the method of two-dimension decision graph. Thus, the number of centers can be determined by taking ρ and δ into consideration, ρ as the horizontal axis and δ as the vertical axis:

$$\gamma_i = \rho_i \delta_i, i \in I_S \quad (7)$$

The ρ_i , which is the i th vector X_1^* , the ranking value of key nodes. δ_i represents the minimum shortest path length of RTN. The large γ , it may be the regional patch center. Therefore, it is necessary to arrange $\{\gamma_i\}_{i=1}^N$ in descending order, and then intercept a number of data points from front to back as the regional patch centers. Therefore, C nodes distributed in the upper right of the graph were selected as the regional patch centers.

In the RTN, based on the Adapted K-Means algorithm to calculate the regional patches. The Adapted K-Means algorithm process are shown in Table 2. The Adapted PageRank algorithm uses the characteristic influence factor of RTN to calculate the ranking results of key nodes. The influencing factors include the location property, the geographic



FIGURE 2: Zhangwu road traffic network.

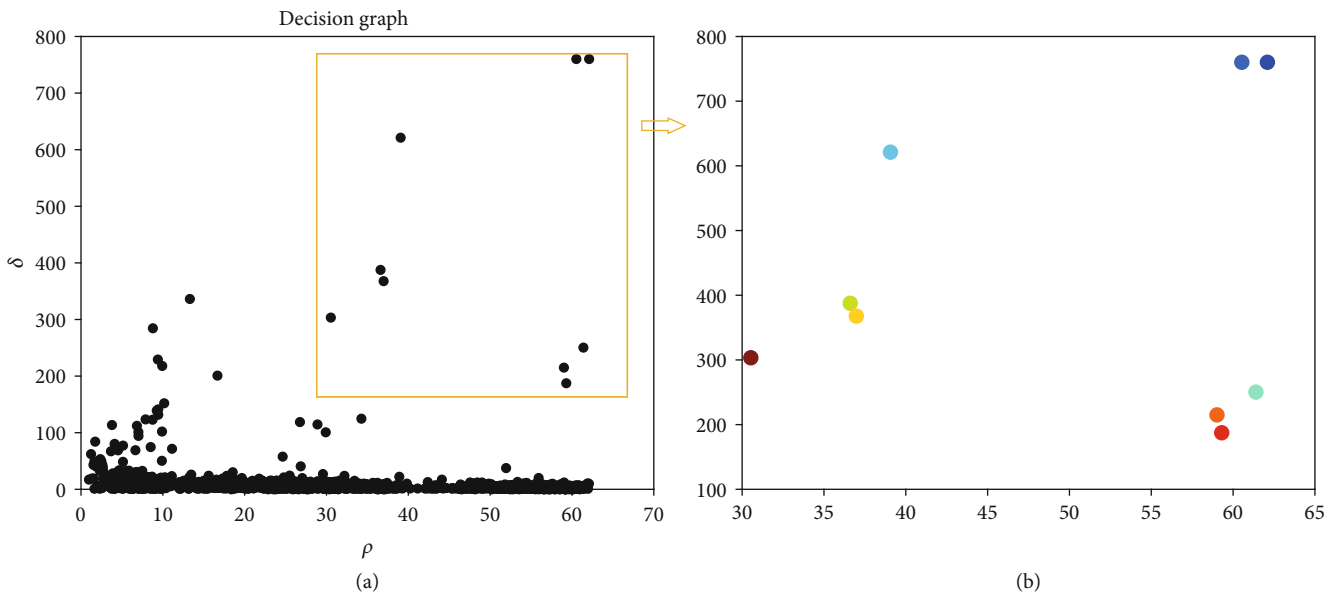


FIGURE 3: Initial cluster center selection results: (a) decision graph; (b) initial cluster center.

distance, and the road level. The method of two-dimension decision graph include the ranking value of key nodes as the horizontal axis and the minimum shortest path length as the vertical axis. The Adapted K-Means algorithm selected the center point of regional patch detection according to the two-dimension decision graph, so as to implement the regional patch detection of RTN. So, the RTN regional patch detection procedure as:

Step 1: in the RTN, calculate the adjacent matrix A , the property matrix F , the distance weight matrix W , and the road level matrix L

Step 2: perform the Adapted PageRank algorithm to the ranking value of key nodes

Step 3: draw a decision graph with ρ_i and δ_i , find the number of regional patches C and corresponding initial centroids

Step 4: perform the Adapted K-Means algorithm (Table 2) to detection the regional patches

Important key node locations are selected as the central point of regional patch aggregation in the RTN, and the Adapted K-Means algorithm is used to calculate the regional patch distribution. The key nodes of the RTN are taken as the basis of regional patch detection to identify the key locations of the overall RTN and divide the functional structure of the RTN. In the regional patches, the edge intensity is defined to quantify the relationship between each pair of connected

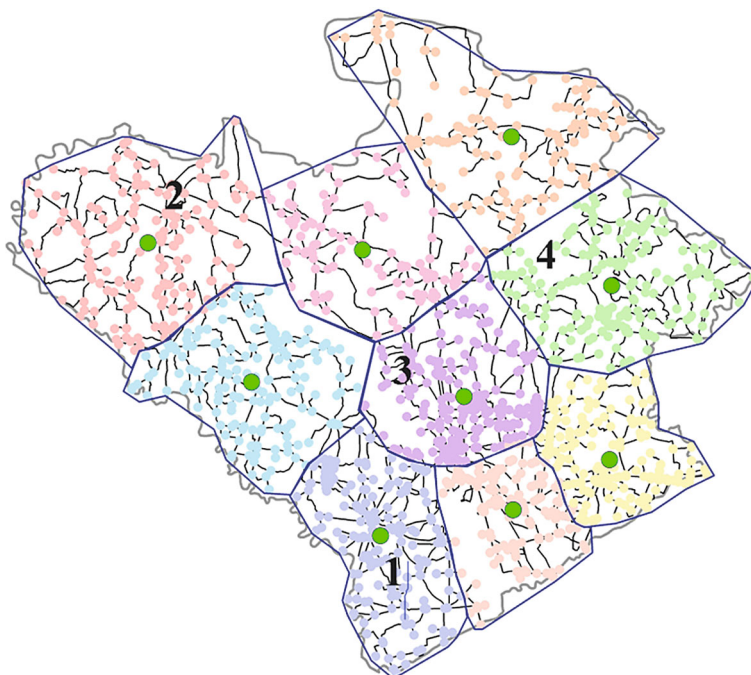


FIGURE 4: Zhangwu road traffic network regional patch detection.

nodes, and the vertical connected by the edges with higher intensities are denoted as core nodes, while the others are denoted as marginal nodes.

The regional patch can identify the key functional structure distribution of the RTN effectively and further understand the structural characteristics of the RTN. Those key nodes are as the center, the regional patch characteristics of the whole network are extracted. From the perspective of geographic space, the characteristic regional patches are detected across the administrative boundaries. The regional patch represents a novel research area and lays the foundation for in-depth analysis of the RTN.

3. Experiments and Analysis

The RTN comprised of national road (G), provincial road (S), county road (X), rural road (Y), and village road (C) in Zhangwu. The total number of nodes is 1750 and the total number of edges is 2053 in Zhuangwu road traffic network. As shown in Figure 2, Zhangwu road traffic network.

In the RTN, the number of nodes representing the national roads and provincial roads are only a small portion of number of nodes of the entire network. And most nodes representing the national roads and provincial roads have relatively low values of degree, and the nodes with large degree are mainly located on the county roads, rural roads, and village roads. Use the RTN regional patch detection to calculate the Zhangwu road traffic network. Get the location information of Zhangwu, $v_i = 1$ and $v_i = 0$ to distinguish the location of important and common features. After defining the property matrix F , computing the distance weight matrix W , and classification of the road level matrix L , we construct a refined matrix G^* .

Then, a vector, which represents the importance of road nodes in the RTN, is obtained. Figure 3 shows the result of the selection of the regional patch center of the RTN. Figure 3(a) shows that the center point of selection is shown in the orange box on the upper right side of the decision graph. The nine road nodes in Figure 3(b) are the number of differentiated regional patches. After this process, the number of regional patches were determined as the output parameters of the Adapted K-Means algorithm.

It can be seen from the decision graph that we calculate the classes of regional patch structures, the nodes with the same color are grouped into the same regional patch, and the number of regional patch structures in the RTN of Zhangwu is 9. As shown in Figure 4, the blue area is the Label 1 regional patch structure; pale orange pink area is the label 2 regional patch structure, and so on. Road nodes within a regional patch indicate that traffic state series on them have relatively high correlation compared with other road nodes in other regional patch. As shown in the results of regional patch detection, road nodes within a regional patch are necessarily connected to each other. The regional patch structure of the RTN in Zhangwu is drawn by the blue border, and the whole county is divided into nine regional patch structures. The structure state of the RTN is analyzed from the characteristics. From the figure, we can clearly find the central position of the nine regional patch. As geographical space structure, the connectivity and convenience of the nine regional patch can greatly improve the RTN of Zhangwu. The location of the center point in the connection diagram highlights the connected critical route, providing theoretical reference for the trend of improving the connectivity of the RTN, and the location of the key maintenance and improvement of the RTN can be determined on the road network

connected by regional patch structure. The RTN of Zhangwu can be better improved and maintain the overall network operating efficiency.

4. Conclusion

Based on the community detection of complex network, the application of Zhangwu RTN is analyzed, the regional patch structure of Zhangwu road traffic network is analyzed, the common aggregation regional patch of key points is drawn, and the functional structure of RTN is deeply analyzed. Key points are identified and analyzed to identify the important locations of the overall RTN. Based on the important locations, the common structure is drawn, and the network is divided into multiple functional regional patch. In RTN, the results of the regional patch detection of the road network system are determined by the influence of the property of ground object facilities, the distance between nodes, and the level distribution of the road network. However, community detection is rarely used in complex road traffic network. Road traffic network (RTN) rely on communities to play multiple roles and embody specific features, and different investigation levels may have different results and perspectives; hence, community detection is of critical importance for a better understanding of road traffic network.

Moreover, the modeling and analyzing methods are applied to the road network in Zhangwu. The results show that the method has a good regional patch detection effect and has great potential in identifying the aggregation degree of road network structure considering the special structural complexity of RTN from the perspective of geographic space. The analysis results can provide an effective theoretical basis for traffic management and operation analysis. However, the results can provide a basis for the regional patch identification of RTN, and the traffic network system can be studied according to the regional patch distribution trend. It provides a theoretical basis for the maintenance and reconstruction of traffic network in the future.

Data Availability

The Zhangwu County Traffic Network data used to support the findings of this study have not been made available because the original road network data is the traffic situation of the real county location in China, and it is real and effective reality data. It can reflect the real data of China's geographical location. Therefore, it cannot be made public. However, the results of the later results can be referenced and applied.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This paper is supported by the Liaoning Province Doctoral Research, Liaoning Provincial Natural Science Fund Project

Key Project (20170520141), and Liaoning Provincial Public Welfare Research Fund Project (20170003).

References

- [1] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [2] S. Boccaletti et al., "Complex networks: structure and dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
- [3] M. E. J. Newman, "Communities, modules and large-scale structure in networks," *Nature Physics*, vol. 8, no. 1, pp. 25–31, 2012.
- [4] L. da Fontoura Costa, O. N. Oliveira, G. Travieso et al., "Analyzing and modeling real-world phenomena with complex networks: a survey of applications," *Advances in Physics*, vol. 60, no. 3, pp. 329–412, 2011.
- [5] S. Fortunato and D. Hric, "Community detection in networks: a user guide," *Physics Reports*, vol. 659, pp. 1–44, 2016.
- [6] Y. Yang, L. Jia, Y. Qin, S. Han, and H. Dong, "Understanding structure of urban traffic network based on spatial-temporal correlation analysis," *Modern Physics Letters B*, vol. 31, no. 22, p. 1750230, 2017.
- [7] Y. Yang, J. Cao, Y. Qin, L. Jia, H. Dong, and A. Zhang, "Spatial correlation analysis of urban traffic state under a perspective of community detection," *International Journal of Modern Physics B*, vol. 32, no. 12, p. 1850150, 2018.
- [8] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, "Vital nodes identification in complex networks," *Physics Reports*, vol. 650, pp. 1–63, 2016.
- [9] H. Liao, M. S. Mariani, M. Medo, Y.-C. Zhang, and M.-Y. Zhou, "Ranking in evolving complex networks," *Physics Reports*, vol. 689, pp. 1–54, 2017.
- [10] T. Agryzkov, L. Tortosa, and J. F. Vicent, "New highlights and a new centrality measure based on the adapted pageRank algorithm for urban networks," *Applied Mathematics and Computation*, vol. 291, pp. 14–29, 2016.
- [11] M. Yu, A. Hillebrand, P. Tewarie et al., "Hierarchical clustering in minimum spanning trees," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 2, article 023107, 2015.
- [12] P. Kim and S. Kim, "Detecting community structure in complex networks using an interaction optimization process," *Physica A: Statistical Mechanics and its Applications*, vol. 465, pp. 525–542, 2017.
- [13] X. Zhu, W. Song, and L. Gao, "Topological characteristics and vulnerability analysis of rural traffic network[J]," *Journal of Sensors*, vol. 9, 2019.