Improving CYGNSS-based Soil Moisture Coverage through Autocorrelation and Machine Learning-Aided Method

Improving CYGNSS-Based Soil Moisture Coverage Through Autocorrelation and Machine Learning-Aided Method

Yan Jia, *Member, IEEE*, Zhiyu Xiao, Shuanggen Jin, *Senior Member, IEEE*, Qingyun Yan, *Member, IEEE*, Yan Jin, Wenmei Li, *Member, IEEE*, and Patrizia Savi, *Senior Member, IEEE*

***Abstract*—Global Navigation System Reflectometry (GNSS-R) is a microwave remote sensing technology that enables Earth observation by receiving GNSS signals reflected from the Earth's surface. The Cyclone Global Navigation Satellite System (CYGNSS) constellation is a satellite system that uses GNSS-R technology with high temporal resolution and has been a popular data source in soil moisture retrieval in recent years. However, the constant movement of GNSS transmitters and GNSS-R satellites results in potentially chaotic and random observations of the Earth's surface, with many unevenly distributed gaps in the observed data. In this paper, a gap-filling method based on spatial autocorrelation is proposed to interpolate the gaps within these observation datasets, with SM being estimated post-interpolation. The sample set for the model comprises points surrounding the interpolation target, with modeling conducted considering factors of spatial weighting to estimate values at the interpolation target. Different autocorrelation-based gap-filling methods using CYGNSS data can achieve good estimation accuracy, and the data coverage after interpolation is on average 1.8 times greater than before interpolation. The gap-filling method using XGBoost achieves the best performance and offers the highest accuracy in SM estimation, with an average correlation coefficient of 0.8445, and an average RMSE of 0.0457 m^3/m^3. The gap-filling approach can significantly enhance data coverage and facilitate the filling of daily gaps in CYGNSS data with all maintaining high SM estimation accuracy. The estimation of daily missing values using CYGNSS data can fully exploit the embedded surface features in the data's fine resolution and can provide high-resolution SM retrieval.**

***Index Terms*—Cyclone global navigation satellite system (CYGNSS), gap-filling method, GNSS reflectometry (GNSS-R), soil moisture (SM), soil moisture active passive (SMAP).**

Manuscript received 11 January 2024; revised 26 April 2024; accepted 23 June 2024. Date of publication 27 June 2024; date of current version 24 July 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42001375 and Grant 42001362, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20180765, in part by the Nanjing Technology Innovation Foundation for Selected Overseas Scientists under Grant RK032YZZ18003, in part by the Shanghai Leading Talent Project under Grant E056061, and in part Strategic Priority Research Program Project of the Chinese Academy of Sciences under Grant XDA23040100. *(Corresponding author: Shuanggen Jin.)*

Yan Jia, Zhiyu Xiao, Yan Jin, and Wenmei Li are with the Department of Surveying and Geoinformatics, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mail: yan.jia@njupt.edu.cn).

Shuanggen Jin is with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China, and also with the Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China (e-mail: sgjin@shao.ac.cn).

Qingyun Yan is with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China.

Patrizia Savi is with the Department of Electronic and Telecommunication, Politecnico di Torino, 10129 Torino, Italy.

Digital Object Identifier 10.1109/JSTARS.2024.3419779

I. Introduction

GLOBAL navigation satellite system-reflectometry (GNSS-R) is an emerging microwave remote sensing technique that utilizes satellite navigation constellations to monitor the Earth's surface with GNSS reflected signals using bistatic geometry [1], [2], [3]. The reflected signals, transmitted by GNSS satellites, are subsequently scattered forward across the Earth's surface in the specular direction, carrying valuable information about the scattering surface characteristics [4]. By utilizing GNSS signals reflected from scattering surfaces, the geophysical properties at the reflection points can be determined through cross correlation with direct GNSS signals received, or signal replicas [5], [6]. GNSS-R signals, which are typically in the L-band, offer short revisit times and higher spatial resolution with significant potential in remote sensing monitoring and applications.

With the advancement of GNSS-R, the new satellite constellation cyclone GNSS (CYGNSS) has received noteworthy attention with providing long-term time series observational data. The CYGNSS satellites, launched by NASA in December 2016, were initially aimed at observing tropical cyclones by estimating the wind speed within the latitudes of 38°N and 38°S [7]. The positions of the ground observation points observed by the CYGNSS constellation are determined by the positions of the GNSS signal receivers on each satellite, which are continuously changing, as well as by the moving positions of the GNSS signal transmitters [8]. Consequently, the CYGNSS constellation observes the Earth's surface in a pseudorandom manner, and the distribution of the observed ground points roughly follows the trajectory of the CYGNSS satellites. Although the CYGNSS constellation can provide daily surface observation data, there are many blank gaps in the spatial distribution of these reflection points.

The applications of CYGNSS have progressively expanded from its initial use for measuring ocean winds [9], [10] to building retrieval models for soil moisture (SM) [11], [12] and analyzing the effects of influencing factors on reflectivity and SM. Chew et al. [13] pointed out the strong correlation between

CYGNSS reflectivity and soil moisture active passive (SMAP) SM, and achieved accurate SM estimation. Both AL-Khaldi et al. and Clarizia et al. [1], [14], [15] constructed SM estimation models to fit the relationship between reflectivity, vegetation opacity, surface roughness, and SM. Calabia et al. [16] used CYGNSS data to indirectly estimate SM by calculating the Fresnel coefficients. Yang et al. [17] achieved the estimation of SM by coupling reflectivity with SMAP brightness temperature data based on a physical algorithm. Senyurek et al. [11], [12] used multiple ML models to retrieve SM based on CYGNSS data, and the RF model demonstrated good retrieval accuracy. Lei et al. [18] used the RF model to simulate the nonlinear relationship between CYGNSS surface reflectivity and a variety of surface features to obtain SM. Nabi et al. [6] used a convolutional neural network (CNN) model to predict SM using CYGNSS reflectivity as well as other auxiliary data to further improve the accuracy of SM retrieval. Most CYGNSS-based SM estimates use SMAP data as a reference. To accommodate modeling needs, CYGNSS data are often projected onto a grid larger than its resolution, at either 36 or 9 km, followed by spatial or grid averaging within the respective grid. However, many grid values are still missing in specific regional CYGNSS-based SM estimation. The number of grids with missing values increases when CYGNSS data are used in research applications that require higher resolution.

In addition, CYGNSS observations have been discovered to exhibit sensitivity to water bodies even in dense vegetation areas like the Amazon region [19], prompting the use of CYGNSS data in hydrologic studies such as detecting flood inundation scenarios and identifying changes in water bodies. To maintain spatial data continuity, observations have been aggregated over longer intervals [20]. For monitoring changes in the water bodies and detecting inundation, CYGNSS data are typically resampled at a resolution of 0.03° and averaged over 14–21 days to extend the data coverage. Notably, this average time interval is extended up to 30 days or more in studies aiming for an extremely fine spatial resolution of 0.01° or smaller [21], [22], [23], [24], [25]. While these approach helps to fill spatial blank gaps of data, it poses a challenge as aggregated observations that are closely linked in space may be widely separated temporally. In addition, aggregating observations over longer periods can dilute the robustness of the findings and complicate the data analysis.

Indeed, balancing the detailed spatial resolution of CYGNSS data with the necessity for comprehensive coverage is a complex task. In previous CYGNSS-based studies, the observations were commonly resampled or aggregated to ensure adequate data coverage. However, this process has led to the loss of important fine-scale surface features and severely hindered the development of applications that rely on the detailed parameter retrieval possible with CYGNSS data. Therefore, it is essential to address the gaps in CYGNSS data by integrating observations with high spatial resolution.

This work aims to utilize the random nature of CYGNSS reflection points and the concept of spatial autocorrelation. It proposes the use of several observed data points around the target point, as the main variables to establish an autocorrelation model. Both typical regression and ML methods are adopted with considering spatial weighing factor. The constructed autocorrelation-based model is utilized to predict the missing points in CYGNSS data, thereby filling the gaps in daily CYGNSS observations. This approach leverages the concept of spatial autocorrelation to estimate the target points robustly, making full use of the random distribution of CYGNSS reflection points. It significantly improves the daily coverage of SM estimation using CYGNSS data and minimizes the dependence on auxiliary data from other sources. Section II describes the adopted dataset and the data processing procedure; Section III outlines the regression and ML algorithms, as well as the proposed method for constructing the CYGNSS SM estimation model based on spatial autocorrelation; Section IV presents and analyzes the experimental results; and finally, the conclusions are summarized in Section V.

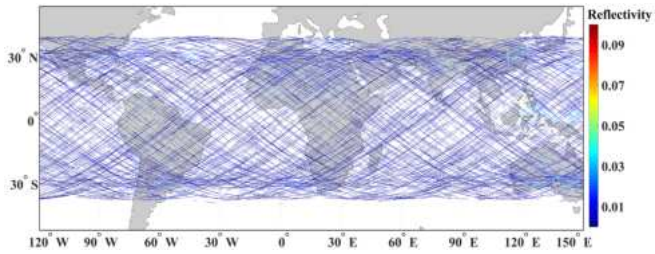

Fig. 1. CYGNSS observations on July 1st, 2023.

II. Dataset and Data Processing

A. CYGNSS Dataset

The CYGNSS constellation comprises eight microsatellites, each of which can receive both direct signals from GPS satellites and reflected signals off the Earth's surface. In the CYGNSS mission, each satellite is equipped with a dual-channel radar capable of receiving up to four signals simultaneously. This configuration enables the collection of data from all eight satellites at a single point in time, which facilitates observations at 32 different points on the Earth's surface to produce delay-Doppler Maps and bistatic radar cross-section metadata [26]. The CYGNSS constellation boasts high temporal resolution with short revisit periods, approximately 7.2 h for the oceans and 1-2 days for land [27]. Theoretically, the spatial resolution of the constellation varies from 0.5 to 25 km for specular reflections in the Fresnel region and diffuse reflections in the glitter region [28]. Since 2019, CYGNSS's sampling interval has been reduced from 1 to 0.5 s, commonly achieving a minimum spatial resolution of 3.5×0.5 km [14].

In this study, we utilized the CYGNSS Level 1 Version 3.1 data product and calculated surface reflectivity using the bistatic radar equation based on the metadata. Since the sampling locations of CYGNSS data are not fixed, we projected the surface reflectivity data onto a 9 km $\times$ 9 km EASE-Grid 2.0. Fig. 1 depicts the distribution of daily CYGNSS observations. Nearly one million CYGNSS data points were obtained and filtered excluding those flagged by the SMAP quality indicator, to enhance model accuracy and yield more precise results [29].

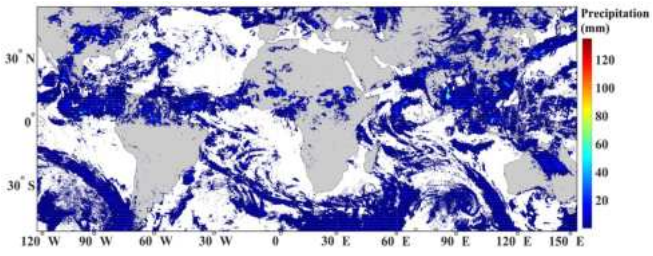

Fig. 2. GPM daily precipitation on July 1st, 2023.

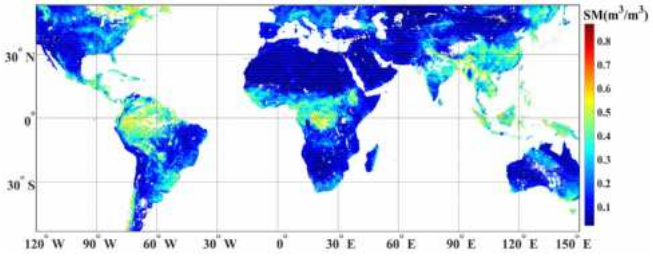

Fig. 3. Distribution of average SMAP SM from July 1st to 3rd, 2023.

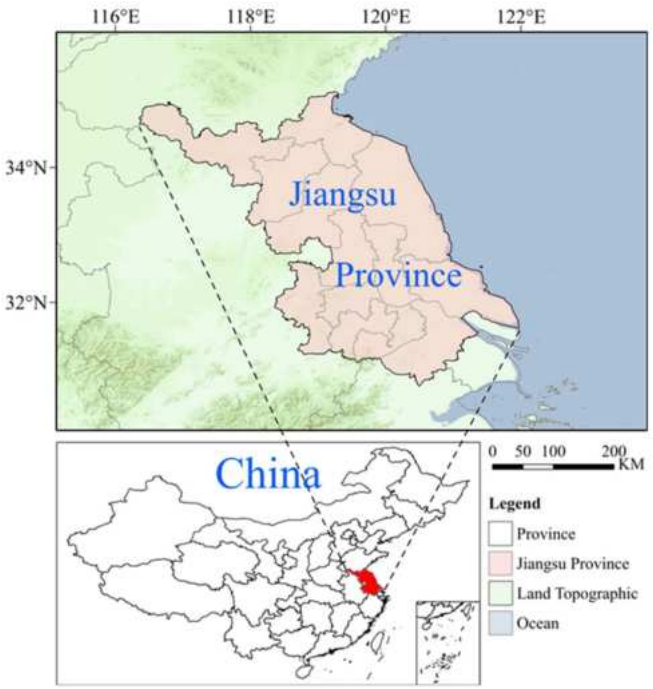

Fig. 4. Study area in Jiangsu Province and Surrounding Regions, China.

The GNSS-R based SM retrieval method would rely on the bistatic radar equation to obtain the ground reflectivity [29]. The coherent component of the surface reflected signal received by the bistatic radar can be described as follows:

$$P_{RL}^{coh} = \left(\frac{\lambda}{4\pi}\right)^2 \frac{P_t G_t G_r}{(R_r + R_t)^2} \Gamma_{RL}(\theta) \tag{1}$$

where λ is the wavelength, P_t is the peak power of the GNSS transmitted signal, G_t is the gain of the transmit antenna, and G_r is the gain of the receive antenna. R_r is the distance from the specular reflection point to the GNSS-R receiver, R_t is the distance from the specular reflection point to the GNSS transmitter, and $\Gamma_{RL}(\theta)$ is the reflectivity at the specular reflection point.

For the incoherent component, it can be described as

$$P_{RL}^{inc} = \frac{\lambda^2 P_t G_t G_r R_{PL}}{(4\pi)^3} \sigma_{RL} \tag{2}$$

where σ_{RL} is the bistatic radar cross-sectional area in square meters and R_{PL} is the Fresnel coefficient.

When the surface is relatively flat and smooth, and the reflected signal received is predominantly coherent, it is considered that $P_{RL}^{coh} = P_{RL}^{inc}$, the ground reflectivity can then be expressed through the following equation:

$$\Gamma_{RL}(\theta) = \frac{\sigma_{RL}(R_r + R_t)^2}{4\pi {R_t}^2 {R_r}^2}. \tag{3}$$

B. GPM Precipitation Data

The global precipitation measurement (GPM) mission is jointly managed by NASA and the Japan Aerospace Exploration Agency. It aims to provide high-quality precipitation observations with both high spatial and temporal resolution on a global scale, utilizing satellite technology. The GPM satellites are equipped with advanced precipitation observation instruments, including dual-frequency radars and microwave radiometers. The data provided by GPM encompass various aspects of precipitation information, including precipitation rates, precipitation patterns, and vertical precipitation profiles. Precipitation rate data are typically measured in terms of hourly or daily amounts (see Fig. 2). GPM satellites boast high resolution and precision, with observations featuring a spatial resolution of approximately 0.1°. In this article, we utilize daily precipitation data from the GPM Level 3 Version 6 (L3 V6) dataset, acquired from the Data Center at NASA's Goddard Space Flight Center. The downloaded daily precipitation data were resampled onto the EASE-Grid 2.0 at a resolution of 9 km, used by SMAP, and employed as one of the auxiliary variables for SM retrieval.

C. SMAP Dataset

The SMAP is a satellite mission launched in January 2015 by NASA in collaboration with the United States Department of Agriculture to globally observe SM with a temporal resolution of three days [30]. SMAP employs an L-band microwave radiometer to collect brightness temperature observations, which are then converted into estimates of SM (see Fig. 3). The daily data products provided by SMAP include gridded SM estimates at 36 and 9 km scales on the EASE grid, as well as vegetation opacity, surface roughness, and other auxiliary data [31]. For this study, we have used the SMAP Level 3 Version 4 SM data product (global daily 9 km EASE grid). It contains a 16-bit binary string known as the SMAP retrieval quality flag, which is an important quality control indicator data that can help to filter the unreliable data.

In this study, the SM values provided by SMAP serve as a reference for both modeling and validating models, aiding in the verification and assessment of SM estimates derived from CYGNSS data, before and after interpolation. Furthermore, the surface roughness and vegetation opacity data from SMAP were used in modeling SM based on CYGNSS observations. The data utilized span from July 1, 2023, to July 30, 2023. To clearly demonstrate the effects of gap-filling, the study areas selected should be of a moderate or small scale. Given the practical need for SM applications in society, our research focused on a provincial level, choosing Jiangsu Province as the primary study area.

D. Study Area

Jiangsu Province is located in eastern China, spanning from 30° to 36°N latitude and 115° to 123°E longitude (see Fig. 4). It boasts a superior geographical location and favorable conditions for economic development, making it one of China's most important industrial production areas. With a total area of 102 600 square km, Jiangsu has a population of over 80 million. The region's cities are known for advanced manufacturing and well-developed high-tech industries. Nanjing, the provincial capital, is recognized as one of China's ancient capitals and is celebrated for its long history and rich cultural heritage. The topography of Jiangsu, together with the surrounding areas, is primarily characterized by coastal plains and the hilly regions found in the south-central part of the province, creating a generally low-lying and rolling landscape. The northern coastal belt of Jiangsu features the typical Yellow Sea-Huaihai Plain, which is among China's most crucial agricultural areas for grain production. The Huai River flows through northern Jiangsu, creating the middle and lower Huai River Plain, a low-lying area prone to flooding. The flat terrain in these areas impacts surface reflectivity only marginally. As a result, the signals received are mainly composed of coherent components. Therefore, SM retrieval results obtained from these flat regions are generally more accurate, which is advantageous for verifying the precision of gap-filling results.

III. Theory and Methods

A. Gap-Filling Method Based on the Spatial Autocorrelation

Spatial correlation generally refers to the relationship between entities or phenomena situated at distinct positions within the space, which can encompass any type of connection within a geographical context. Spatial autocorrelation [32], a term used to describe the spatial relevance of geographical entities to their own features, implies that there is a correlation among similar variables at different spatial positions. If a variable's value at one location is associated with its values at adjacent locations, spatial autocorrelation is said to exist. To a certain degree, spatial autocorrelation is a specialized form of spatial correlation, indicating the level of association among neighboring points in spatial datasets.

The intensity of the GNSS signal reflected from the Earth's surface, as received by the CYGNSS constellation, exhibits a strong correlation with the surface SM. When the surface SM and the surface condition remain constant within a confined spatial area, the received signals ought to remain consistent. Furthermore, physical surface parameters such as SM and texture exhibit significant spatial autocorrelation. In other words, the signals received by CYGNSS are expected to display a high degree of spatial autocorrelation, particularly within a confined area or over a specified spatial extent, where the value at a specific reflection point or grid cell should be consistent and sustain a stable relationship with the values at the surrounding grid cells. For grid cells with missing data, observed values from nearby grid cells can, to some extent, represent the observational results of that location. Accordingly, in this study, to estimate the value for missing grid (Y_j), the model uses the value from nearby data grid and the values of a certain number of surrounding data grids (X_m^j) as input variables. The model is represented as

$$Y_j = f\left(X_1^j,\ X_2^j,\ \ldots,\ X_m^j\right). \tag{4}$$

Here, *j* denotes the position of the modeling points around the point to be measured, and *m* represents the count of neighboring points used to characterize the spatial features of the modeling point. The model of the neighboring modeling points is used to predict and interpolate the points to be estimated, from which the value of Y_j can be obtained.

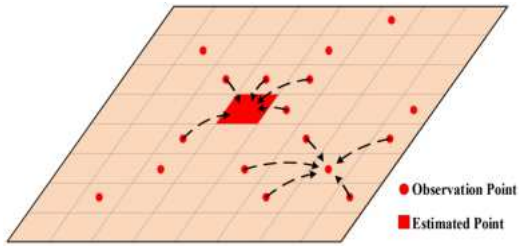

Fig. 5. Spatial autocorrelation-based modeling gap-filling method (9 km × 9 km EASE grid).

As shown in the illustration (see Fig. 5), to establish a training set using existing data points, one would formulate a relationship model between independent variables x and dependent variables y. To ensure the reliability of the constructed dataset, the maximum distance for interpolation in the EASE Grid was set to 2 cells. Once this model is constructed, new independent variables can be input to predict their corresponding dependent variables, which are the values for the points that require interpolation. It is important to note that for interpolating CYGNSS-based data, we adopt a "spreading" method. This approach prioritizes interpolation for grid points that are closer to the established training set, performing predictions and assigning values progressively from nearer to more distant points. Predicted values are then automatically integrated into the training set. This enhanced training set is subsequently used to interpolate the next closest grid points. By iterating this process, we can methodically fill the entire area requiring interpolation.

From this, it can be seen that determining the locations of the modeling points, the quantity of surrounding points, and their assigned weights are crucial factors influencing the model's

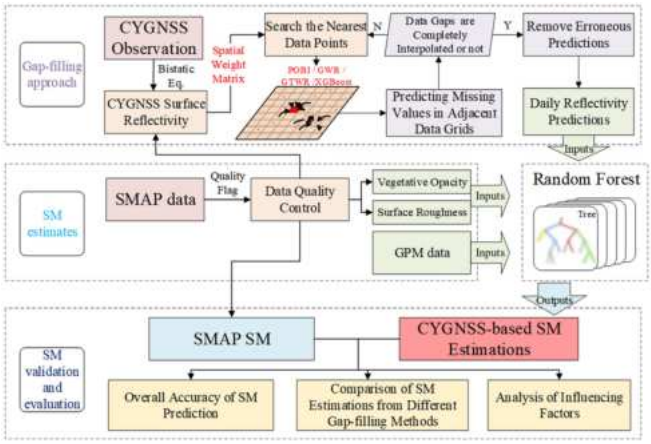

Fig. 6. Flowchart illustrating the proposed process.

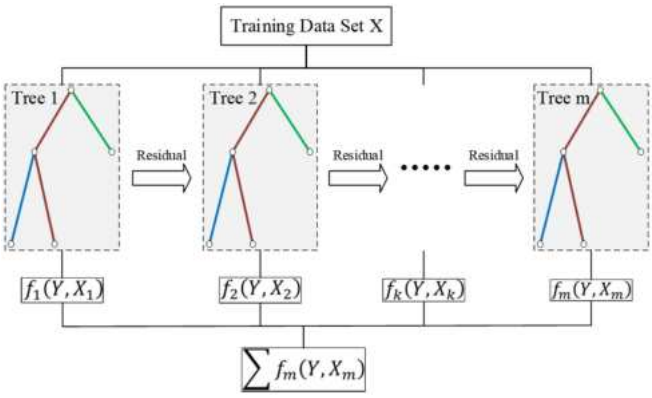

Fig. 7. XGBoost Algorithm.

accuracy. Building on the preceding discussion, this study primarily accomplishes the following:

1) Comparing the accuracy of SM estimation before and after interpolation.
2) Comparing daily SM coverage before and after interpolation.
3) Using regression and ML aided methods to construct autocorrelation models and compare their performances.
4) Analyzing factors that may impact the precision of gap-filling, such as the quantity of samples.

Fig. 5 illustrates the principal algorithms and proposed modeling processes.

As shown in Fig. 6, first, the surface reflectivity is calculated using the bistatic radar equation based on CYGNSS observation data. Values from neighboring points, chosen as modeling points and considering spatial weight factors, are utilized as the training set for the sample. A selected number of values from the vicinity of the modeling points are included in this set. This study incorporates four typical regression and machine learning algorithms to aid modeling and training: Geographically weighted regression (GWR), geographically and temporally weighted regression (GTWR), previously observed behavior interpolation (POBI), and the extreme gradient boosting (XGBoost) model.

B. GWR and GTWR Methods

The GWR and GTWR models are advancements of the conventional ordinary least squares (OLS) model. Serving as a global regression model, OLS often functions as a modeling tool in geographical analysis, where the dependent variable is modeled as a linear function of a set of (n) independent variables, also known as predictor variables. The OLS model can be represented as [33], [34]:

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i. \tag{5}$$

In the OLS model, the estimated regression coefficients are average values for the entire study area and the regression parameters cannot accurately reflect the true spatial characteristics. The GWR is an extension of traditional regression models that allows for local parameter estimation as opposed to direct estimation of global parameters. When variables exhibit significant spatial heterogeneity, the issue of spatial non-stationarity can be well addressed by computing the local parameters of the regression model, thereby optimizing the model's fit. The GWR model can be expressed as [35]

$$y_i = \beta_0 (u_i, v_i) + \sum_k \beta_k (u_i, v_i) x_{ik} + \varepsilon_i. \tag{6}$$

In this model, y_i is the dependent variable for the ith sample point, (u_i, v_i) represents the spatial location of the ith sample point, $\beta_0(u_i, v_i)$ is the constant term estimate at the ith point, $\beta_k(u_i, v_i)$ is the estimated regression coefficient for the kth independent variable at point i, x_{ik} is the independent variable at the ith sample point, and ε_i is a random error term that follows a normal distribution. The GWR formulation considers potential spatial variation in the relationship between the independent and dependent variables, where observational data near location i may have a larger impact on the estimation of $\beta_k(u_i, v_i)$ as opposed to data further away from i. In GWR, weights are applied to observations near the location i; thus, the weight of the observation's changes with i. Observations closer to i are given greater weight than those further away, as follows:

$$\hat{\beta}\ (u_i, v_i) = \left(X^T W\ (u_i, v_i)\ X\right)^{-1} X^T W\ (u_i, v_i)\ y \tag{7}$$

where $\hat{\beta}$ is the estimated value of β, and $W(u_i, v_i)$ is an $n \times n$ diagonal matrix whose diagonal elements represent the geographical weight values for observational data point i.

While the GWR model addresses spatial heterogeneity in estimating regression parameters, building local regression models for parameters, it does not consider heterogeneity over time. In 2010, Huang et al. [36] introduced the temporal scale into the GWR model, proposing the GTWR model. This model incorporates time factors in the calculation of weighting matrices and the estimation of regression coefficients, which can mathematically be expressed as

$$y_i = \beta_0 (u_i, v_i, t_i) + \sum_k \beta_k (u_i, v_i, t_i) x_{ik} + \varepsilon_i. \tag{8}$$

Methods based on GWR have been applied to the downscaling of SM and the study of spatial heterogeneity, yielding favorable

results [37]. It was revealed that the SM exhibited spatial heterogeneity and distinct clustering features. The fitting performance of the GWR model was found to significantly surpass that of the ordinary OLS model, both in terms of goodness of fit and spatial distribution.

C. POBI Method

To better reflect the spatiotemporal distribution characteristics of CYGNSS observations, Chew [8] proposed an interpolation method based on previous observations behavior (POBI). This method integrates spatial interpolation with autoregressive time series analysis, interpolating using previously recorded observations. The autoregressive principle assumes that the predictive variable linearly depends on past values, using observations from previous time steps to forecast future observations. In other words, the future behavior of the output variable is calculated by regressing its past observations.

The POBI method applies the autoregressive principle in every aspect within the spatial domain, regressing previous observations at the target location with proximal observations from surrounding locations. Empirical regression parameters are then used along with any available observations near the point of interest to interpolate the estimated value at the target location. The expression of the POBI method is as follows:

$$y = \frac{\sum_{i=1}^{n} w_i \left(a_i x_i + b_i\right)}{\sum_{i=1}^{n} w_i} \tag{9}$$

$$w_i = r_i^2 \tag{10}$$

where y represents the estimated value at the point of interpolation, x_i denotes the ith available observation point near y, a_i and b_i are regression coefficients for prior observations at the location of x_i relative to y, and w_i indicates the square of correlation coefficient (r_i^2) between them.

In the interpolation process, since the POBI method relies on prior data for the construction of local regression models, there is a certain requirement for the length of prior data. In the interpolation calculation using the POBI method, longer data series yield more reliable regression relations in the modeling. By iterating through proximal points around the point of interest and forming pairs with each neighboring point, regression calculations are performed for each pair to generate several predicted values for the interpolation point. The correlation of each pair to a certain extent reflects the reliability of the estimated value, hence a weighted average method using the correlation as weights is adopted to integrate these estimates, obtaining the final interpolated value for the interpolation point.

D. XGBoost Method

Boosting is an ensemble learning technique that constructs a strong model by combining multiple weak learners. The idea is to combine several weak classifiers, with each one attempting to correct the errors of its predecessor [38], [39], [40]. Extreme Gradient Boosting, or XGBoost, is an ensemble method based on gradient boosted trees and represents an optimized, distributed gradient boosting technique. XGBoost improves upon traditional gradient descent algorithms, offering faster performance enhancements than other ensemble algorithms that use gradient boosting, and is recognized as an advanced estimator with extremely high performance in both classification and regression.

The XGBoost method is based on classification and regression trees (CART), using them to build weak classifiers. After the first CART is constructed, the number of trees is continuously increased in subsequent iterations, gradually forming a strong estimator composed of an ensemble of numerous tree models (see Fig. 7). When constructing the first tree, a portion of the samples from the dataset is randomly selected to serve as the training samples for modeling. After modeling is complete, an evaluation is performed on the training samples, and then the samples with large prediction biases from the model are fed back to the original dataset. In later iterations, samples with larger prediction biases are given more weight, and the new models are more inclined to deal with these difficult-to-predict samples.

The XGBoost method employs a specific objective function to minimize loss, while introducing regularization terms to limit model complexity, which prevents overfitting. Moreover, it uses a greedy algorithm to control the growth of CART, constraining the outcome of the trees. The greedy algorithm focuses on achieving local optima to reach a global optimum, and when all leaf nodes have achieved optimality, the entire tree reaches the optimal performance of the model.

IV. Results and Discussion

A. CYGNSS-Based SM Estimation Without Gap Filling

The CYGNSS observation data from July 1 to July 30, 2023 were selected for the study area to test the feasibility of interpolation. Surface reflectivity is calculated from the CYGNSS observational data using the bistatic radar equation, which is then interpolated using the concept of spatial autocorrelation combined with methods such as POBI, GWR, GTWR, and XGBoost. The objective is to fill in the missing surface reflectivity data in areas not covered by the CYGNSS constellation. To quantitatively evaluate the interpolation results, SM estimates were obtained using a random forest (RF) model and compared with the SM reference values released by SMAP. In the CYGNSS-based SM estimation procedure, the inputs are preinterpolation/postinterpolation surface reflectivity, vegetation opacity, surface roughness, and daily GPM precipitation data, with the output being SM reference values.

Due to the limitations of the SMAP constellation, its data products do not provide full coverage, and their temporal resolution is approximately 2 to 3 days. Therefore, this study adopts the principle of nearest supplementation to process the data, using the closest observation to the missing location from the past three days (including the current day) for supplementation. To achieve more intensive spatial coverage, each SMAP SM datum is reprocessed by averaging over every three consecutive days, hereby referred to as "3-day data." The spatial distribution of CYGNSS data at this scale remains very sparse and uneven. To ensure a more uniform distribution of CYGNSS data within

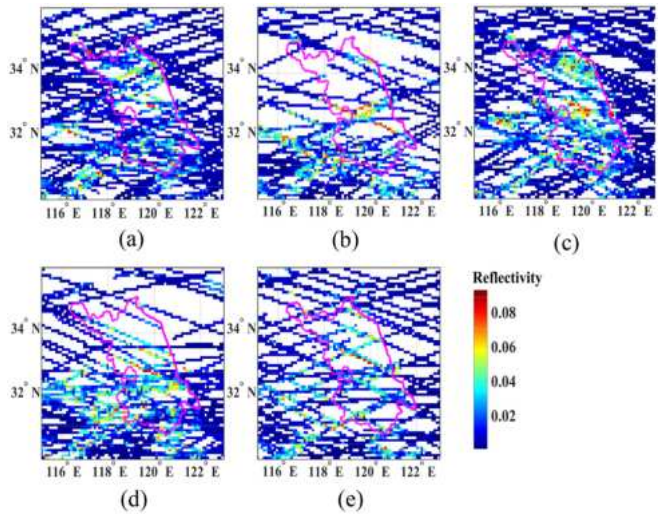

Fig. 8. Daily CYGNSS reflectivity observations before interpolation: (a) July 1st, (b) July 7th, (c) July 13th, (d) July 19th, and (e) July 25th.

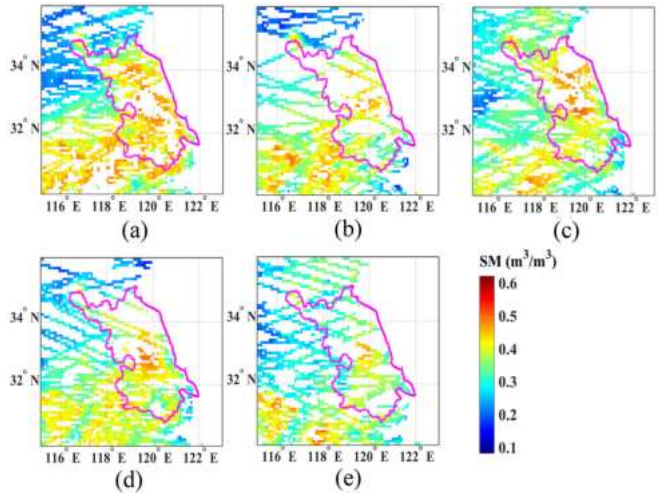

Fig. 9. SM daily estimation before interpolation: (a) July 1st, (b) July 7th, (c) July 13th, (d) July 19th, and (e) July 25th.

the study area and to minimize the impact of time, a 3-day time window was used, which also agrees with the operations of SMAP data.

Figs. 8 and 9 show the processed daily data of CYGNSS surface reflectivity and the daily SM data obtained from retrieval, respectively. To improve the readability of the charts, Figs. 8 and 9 display the results of the averaged CYGNSS surface reflectivity and the estimated SM values from SMAP at six-day intervals.

As shown in Fig. 10, the Pearson correlation coefficient R fluctuates between 0.8721 and 0.9451, with a monthly average of 0.9092. The RMSE varies from 0.0295 to 0.0434 m^3/m^3, with a monthly average of 0.0368 m^3/m^3. The results presented by the model indicate that estimating SM with this combination of input variables is reliable. The same RF modeling approach will also be used in the accuracy assessment of subsequent gap-filling results.

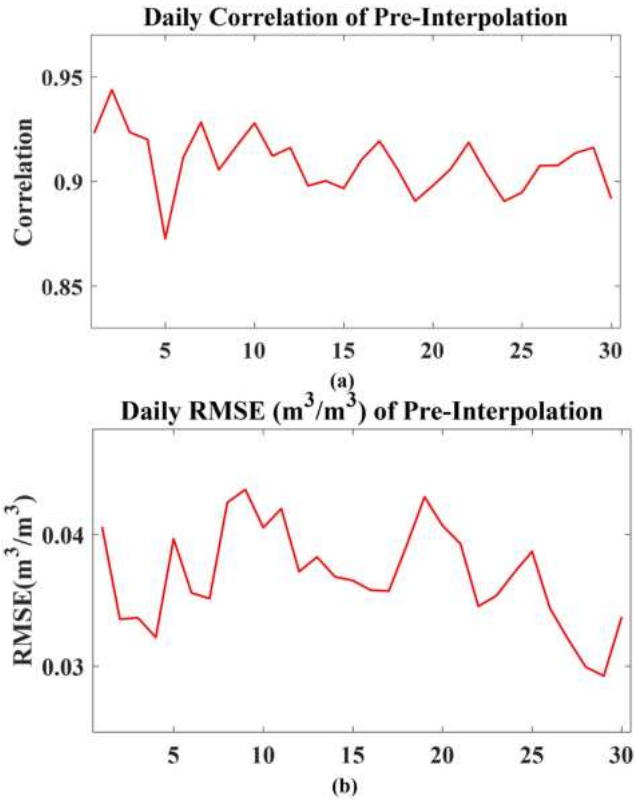

Fig. 10. Accuracy assessment of daily SM estimations based on preinterpolation data. (a) Change of correlation between retrieved SM and SMAP SM. (b) Change of daily RMSE between retrieved SM and SMAP SM.

B. Evaluation of CYGNSS Gap-Filling Method

In this study, we assessed the effectiveness of XGBoost-aided by comparing it with three other competing methods: POBI, GWR, and GTWR, all of which have demonstrated promising results in the SM gap-filling method.

The results of SM estimation, obtained after interpolating CYGNSS reflectivity using different techniques, are presented in Fig. 11 Panels (a) shows SM estimates of preinterpolation data, (b) and (c) display the outcomes derived from GWR with Gaussian and exponential kernels, respectively. Panels (d), (e), and (f) illustrate the SM retrieval results obtained using POBI, GTWR, and XGBoost methods, respectively. The blank areas in the figure represent either bodies of water or regions where the quality of SMAP data is deemed unreliable.

Tables I and II present the correlation coefficients and RMSE between the SM estimates from daily observations before and after interpolation and the SMAP SM reference, respectively. Red values in the Tables I and II indicates the model that achieves the best correlation coefficient for each day. All these methods have achieved good results. Apart from the overall performance decline of each method due to the influence of observations at specific times, the correlation coefficients of the other results range between 0.8 and 0.9, which is consistent with the observations. In addition, XGBoost exhibits better correlation and smaller RMSE compared to several other methods most of the time. All methods demonstrate considerable and stable

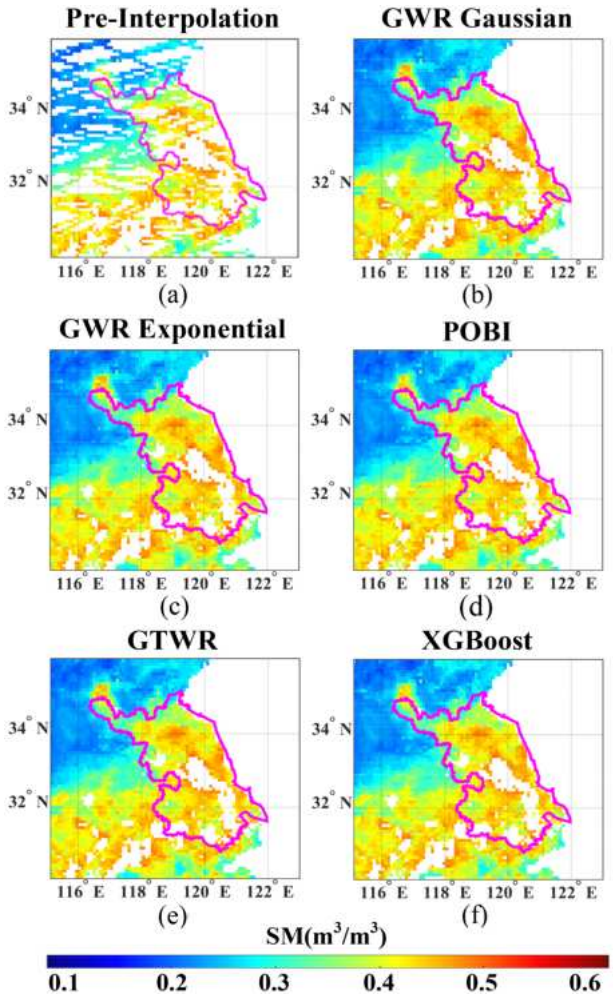

Fig. 11. Comparing SM estimation results from CYGNSS data before and after interpolation: (a) Preinterpolation, (b) GWR with Gaussian kernel. (c) GWR with exponential kernel. (d) POBI. (e) GTWR. (f) XGBoost.

accuracy, indicating that the interpolated results are plausible over extended time series as well.

Fig. 12 illustrates the differences in SM estimation using reflectivity data before interpolation and after interpolation with XGBoost, GWR, and GWRT methods, compared to the reference SM data from SMAP. All these methods yield high correlations with the SMAP SM data, with correlation coefficients above 0.9. Specifically, the correlation coefficient for XGBoost is 0.9156, and the RMSE is 0.0336 m^3/m^3, which are the closest to the accuracy of the preinterpolation monthly mean SM estimation. In contrast, the POBI method yields the least accurate results.

The SM estimates derived from reflectivity data interpolated using the XGBoost method show a significant tendency to overestimate in regions with lower SM levels, while underestimation is observed in areas with higher SM levels. This pattern is consistent with estimates made using observations before interpolation, indicating that the issue is inherent to the estimation process rather than the gap-filling method and could be addressed in future SM retrieval models.

TABLE I
COMPARISON OF DAILY CORRELATION COEFFICIENTS FOR SM ESTIMATES BEFORE AND AFTER REFLECTIVITY INTERPOLATION

Days/ (R)	Pre-interpolation	After-interpolation				
		POBI	GWR Gaus	GWR Exp	GTWR	XGBoost
1	0.9231	0.8791	0.8782	0.8767	0.8769	0.8829
2	0.9437	0.8943	0.8913	0.8925	0.8934	0.8982
3	0.9235	0.8296	0.8309	0.8308	0.8317	0.8394
4	0.9200	0.8183	0.8280	0.8255	0.8257	0.8297
5	0.8726	0.7240	0.7527	0.7521	0.7532	0.7456
6	0.9116	0.8255	0.8294	0.8289	0.8277	0.8331
7	0.9283	0.8555	0.8573	0.8556	0.8558	0.8571
8	0.9056	0.8512	0.8480	0.8468	0.8481	0.8538
9	0.9170	0.8637	0.8657	0.8655	0.8678	0.8698
10	0.9279	0.8814	0.8828	0.8834	0.8826	0.8846
11	0.9122	0.8569	0.8572	0.8600	0.8595	0.8617
12	0.9161	0.8338	0.8411	0.8391	0.8425	0.8394
13	0.8980	0.8238	0.8243	0.8248	0.8231	0.8342
14	0.9002	0.8201	0.8233	0.8243	0.8202	0.8292
15	0.8967	0.8121	0.8146	0.8181	0.8195	0.8203
16	0.9103	0.8690	0.8718	0.8715	0.8728	0.8726
17	0.9194	0.8807	0.8872	0.8853	0.8849	0.8875
18	0.9060	0.8802	0.8827	0.8814	0.8824	0.8819
19	0.8906	0.8655	0.8689	0.8683	0.8672	0.8677
20	0.8980	0.8712	0.8728	0.8722	0.8695	0.8708
21	0.9059	0.8618	0.8679	0.8676	0.8659	0.8676
22	0.9186	0.8794	0.8862	0.8851	0.8850	0.8922
23	0.9036	0.8459	0.8553	0.8552	0.8546	0.8549
24	0.8905	0.8090	0.8086	0.8086	0.8095	0.8171
25	0.8948	0.7600	0.7626	0.7622	0.7613	0.7659
26	0.9075	0.7869	0.7837	0.7856	0.7895	0.8004
27	0.9077	0.7949	0.8008	0.7984	0.7975	0.8004
28	0.9136	0.8194	0.8124	0.8144	0.8148	0.8230
29	0.9162	0.8177	0.8307	0.8296	0.8300	0.8329
30	0.8918	0.8054	0.8149	0.8161	0.8167	0.8230

Fig. 13 displays the distribution of the correlation coefficient (R) between SM estimates obtained using observational data without interpolation over a month and SM estimates acquired from data interpolated by different methods, in comparison with SMAP SM reference values. The blank areas in the map represent permanent bodies of water or regions with missing reference values. Black boxes outline regions with relatively minor variations in SM, where the correlation between the estimation results and the SM reference is notably significant. In other regions, however, the precision of SM estimation after interpolation using the XGBoost method decreases to varying extents. These results confirm the strong spatial autocorrelation inherent in SM, indicating that gap-filling outcomes are more accurate in regions with stronger autocorrelation.

TABLE II
DAILY RMSE IN M^3/M^3 FOR SM ESTIMATES BEFORE AND AFTER REFLECTIVITY INTERPOLATION

Days/ (R)	Pre-interpolation	After-interpolation				
		POBI	GWR Gaus	GWR Exp	GTWR	XGBoost
1	0.0406	0.0498	0.0498	0.0496	0.0502	0.0489
2	0.0335	0.0442	0.0444	0.0448	0.0445	0.0434
3	0.0334	0.0496	0.0497	0.0494	0.0498	0.0483
4	0.0324	0.0474	0.0467	0.0466	0.0467	0.0459
5	0.0396	0.0552	0.0527	0.0528	0.0528	0.0535
6	0.0357	0.0472	0.0473	0.0472	0.0468	0.0464
7	0.0354	0.0472	0.0475	0.0474	0.0473	0.0467
8	0.0425	0.0516	0.0527	0.0521	0.0521	0.0512
9	0.0433	0.0541	0.0541	0.0545	0.0539	0.0534
10	0.0403	0.0505	0.0502	0.0502	0.0499	0.0497
11	0.0420	0.0517	0.0514	0.0515	0.0517	0.0512
12	0.0372	0.0490	0.0481	0.0481	0.0474	0.0481
13	0.0384	0.0463	0.0458	0.0458	0.0460	0.0451
14	0.0366	0.0449	0.0448	0.0447	0.0447	0.0439
15	0.0364	0.0459	0.0456	0.0459	0.0453	0.0458
16	0.0358	0.0420	0.0419	0.0417	0.0417	0.0415
17	0.0358	0.0423	0.0418	0.0414	0.0418	0.0414
18	0.0391	0.0441	0.0438	0.0437	0.0434	0.0439
19	0.0427	0.0468	0.0461	0.0465	0.0461	0.0464
20	0.0405	0.0463	0.0461	0.0460	0.0465	0.0458
21	0.0391	0.0479	0.0471	0.0476	0.0473	0.0468
22	0.0345	0.0413	0.0407	0.0406	0.0409	0.0394
23	0.0353	0.0413	0.0403	0.0401	0.0403	0.0404
24	0.0370	0.0458	0.0457	0.0459	0.0458	0.0453
25	0.0387	0.0532	0.0532	0.0531	0.0532	0.0526
26	0.0340	0.0474	0.0473	0.0473	0.0470	0.0459
27	0.0319	0.0443	0.0441	0.0441	0.0442	0.0435
28	0.0298	0.0401	0.0411	0.0412	0.0407	0.0398
29	0.0292	0.0406	0.0397	0.0397	0.0396	0.0392
30	0.0337	0.0414	0.0402	0.0404	0.0404	0.0396

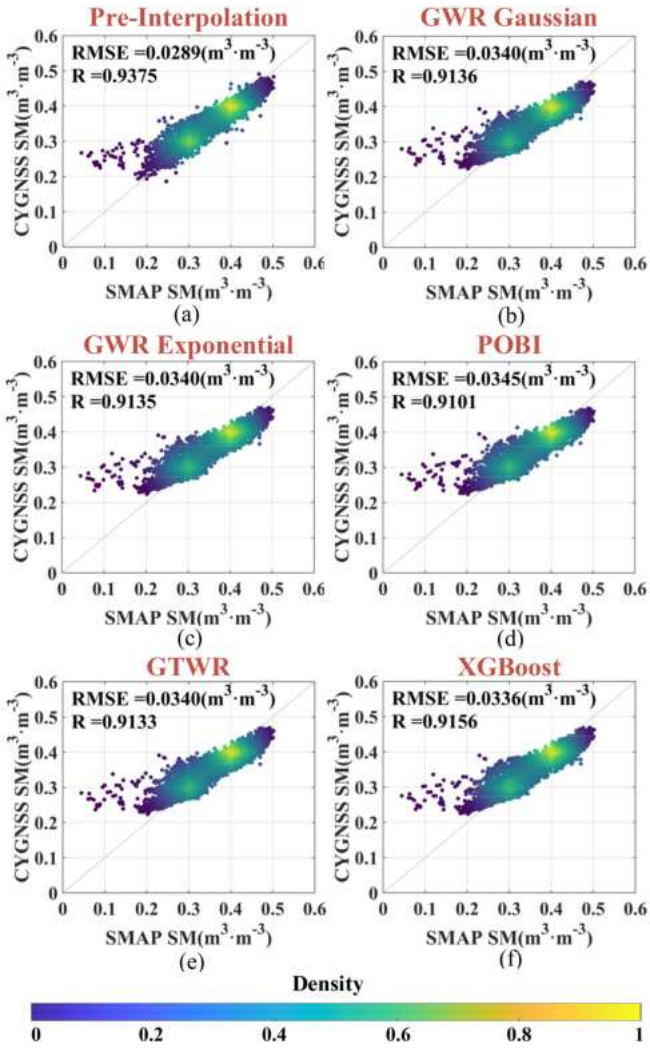

Fig. 12. Scatterplot of monthly mean SM estimates from CYGNSS-based SM with SMAP reference SM. (a) Preinterpolation. (b) GWR with Gaussian kernel. (c) GWR with exponential kernel. (d) POBI. (e) GTWR. (f) XGBoost.

Fig. 14 illustrates the deviation between the monthly average retrieval results obtained after interpolation of all methods and the monthly average SM from SMAP. The positive bias indicates that the SMAP SM values are higher than the retrieval results. It can be observed that, in most parts of the study area, the retrieved SM from interpolated reflectivity tends to be underestimated to varying degrees. However, a small portion of the coastal area, where SM values are low, exhibits an overestimation. This observation is consistent with the conclusions obtained from Fig. 12. The bias in SM estimation results after reflectivity interpolation remains virtually unchanged compared to the preinterpolation results in regions with flat terrain. In regions characterized by low hilly terrain (central and southwestern parts), the bias is slightly more pronounced than in the preinterpolation results. Overall, the biases are distributed between -0.05 and 0.05 m^3/m^3 in most areas and are close to 0 in the left plain area. This suggests that the SM obtained by retrieval after interpolation closely matches the SMAP reference values, allowing for effective SM distribution following the proposed interpolation strategy.

The coverage of data is significantly enhanced after interpolation, and Fig. 12 illustrates the improvement in daily data coverage postinterpolation. As previously indicated in Figs. 8 and 9, the reflection points of the CYGNSS constellation are distributed randomly due to its reflection characteristics. Therefore, the daily observed data varies in regional coverage, leading to fluctuations in the enhanced regional coverage provided by the interpolated data. For instance, in July 2023, the regional coverage of interpolated SM estimates is 1.8 times that of the coverage before interpolation (see Fig. 15). This indicates that employing the concept of autocorrelation in combination with the proposed gap-filling methods such as POBI, GWR, and XGBoost can significantly increase the daily data coverage.

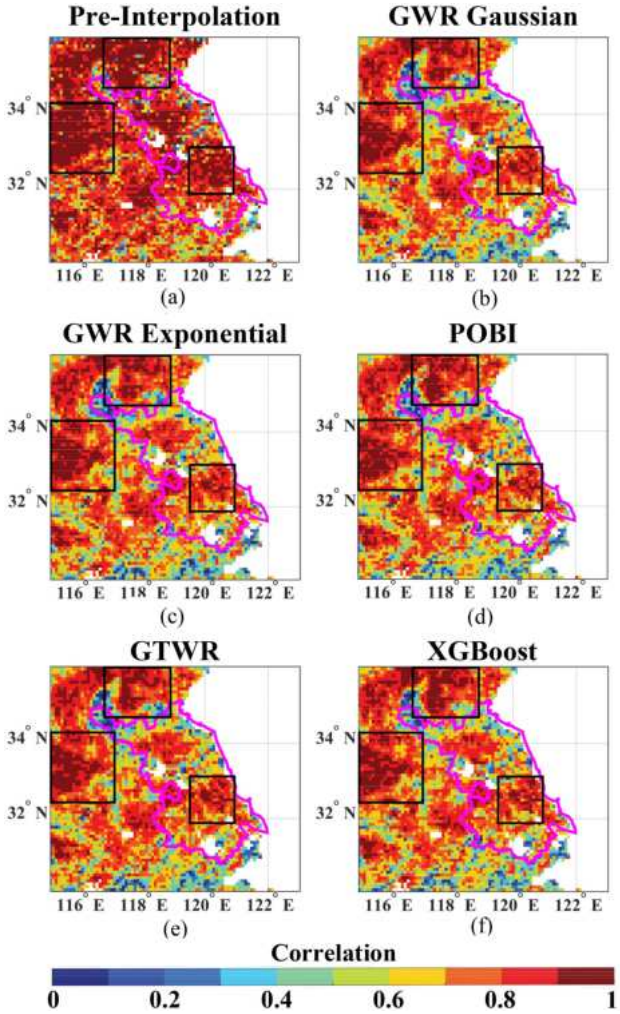

Fig. 13. Distribution of correlation coefficients for SM estimates from CYGNSS observations versus SMAP SM before and after interpolation. (a) Preinterpolation. (b) GWR with Gaussian kernel. (c) GWR with exponential kernel. (d) POBI. (e) GTWR. (f) XGBoost.

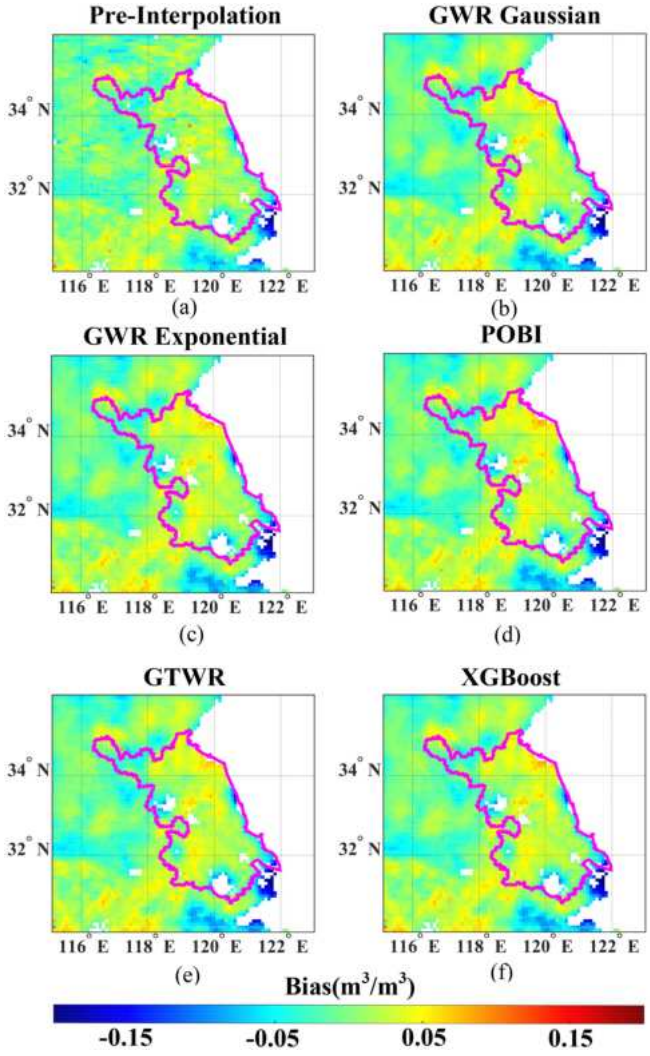

Fig. 14. SM estimation bias between CYGNSS observations and SMAP SM before and after interpolation. (a) Preinterpolation. (b) GWR with Gaussian kernel. (c) GWR with exponential kernel. (d) POBI. (e) GTWR. (f) XGBoost.

C. Influencing Factors for Proposed Gap-Filling Method

When constructing models based on the concept of spatial autocorrelation, a key issue is determining the number of neighboring points to use for building the sample data for modeling. To study the impact of the number of neighboring point samples on the model and the final SM estimation results, we modeled using the four to seven nearest observation points, respectively. The accuracy results obtained from interpolation and prediction were then compared with SMAP SM reference values, as shown in Fig. 16.

In Fig. 16, "GWR Gaus" denotes the GWR method that employs a Gaussian function as the kernel, whereas "GWR Exp" refers to the GWR variant using an Exponential function for the kernel. This figure demonstrates the accuracy of gap-filling results from prediction models that were constructed using varying numbers of neighboring points for retrieving monthly average SM estimates. It is observed that there is no significant change in the accuracy of interpolation-based retrieval for methods such as

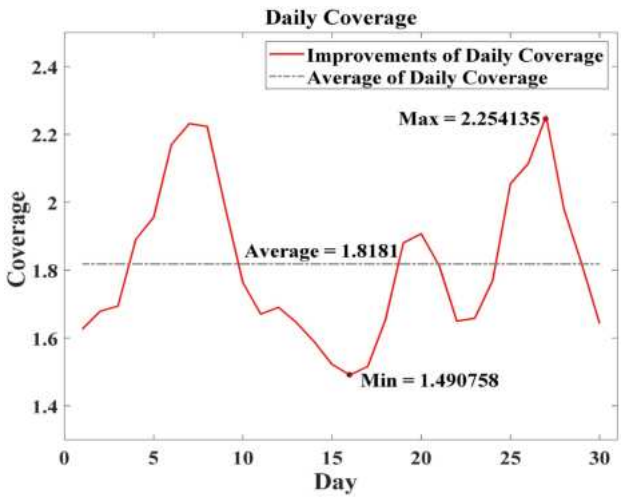

Fig. 15. Daily coverage with proposed gap-filling method (one month).

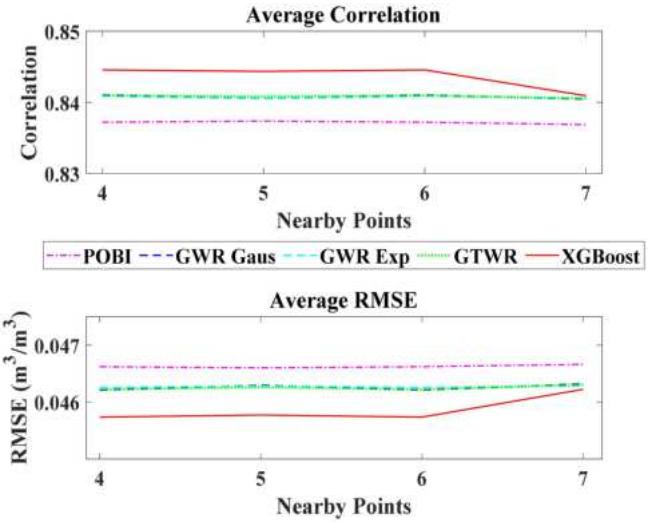

Fig. 16. Comparison of monthly average accuracy in SM estimates from observation interpolation using varied neighboring point counts.

Previously Observed Behavior Interpolation (POBI), GWR, and GTWR as the count of neighboring points increases. However, the precision of the XGBoost method appears to be somewhat impacted, indicating that an excess of information may adversely affect the model's fitting process. Additionally, a higher number of neighboring points correlates with increased computational time required for interpolation. While the POBI method, which is grounded in linear regression, remains largely unaffected by this increase, more complex methods that necessitate model fitting, like GWR, GTWR, and XGBoost, must account for this computational consideration. With regard to the metrics of correlation coefficient R or root mean square error (RMSE), the XGBoost method consistently achieves the highest levels of precision.

V. Conclusion

This paper integrates spatial autocorrelation with a variety of statistical methods, including POBI, GWR, and GTWR, as well as machine learning techniques like XGBoost, to interpolate surface reflectivity in the study area with the objective of filling gaps where CYGNSS observations are missing. The interpolated surface reflectivity is used to estimate SM using an RF model, which is compared with SM reference values released by SMAP. The accuracy of the gap-filling results is evaluated using the Pearson correlation coefficient (R) and RMSE. The results show that the proposed XGBoost-aided SM estimation approach maintain accuracy with filling of daily gaps in CYGNSS data.

Data sets constructed using a varying number of neighboring points, ranging from 4 to 7 nearest observations around the target point, were compared to assess the SM retrieval accuracy of the various methods. The results indicate that gap-filling using the XGBoost method closely matches the observed data in accuracy. The article also analyzes the impact of observation data on daily gap-filling results, reasons for anomalies, and fluctuations in accuracy at specific time points due to interpolation. Since the gap-filling process estimates missing values from surrounding known points, it cannot detect the maximum and minimum values in the missing areas. This limitation can lead to a centralizing tendency in the interpolated data. In addition, due to the distribution of observation data, there can be a decrease in gap-filling accuracy, as noted on the 5th and 25th days, where predicting local peaks overlaid by the overall prediction average is challenging.

Prior to gap-filling, the mean correlation coefficient for SM estimation in the study area was 0.9091, and the mean RMSE was 0.0368 m^3/m^3. After gap-filling, the SM data exhibited an average correlation coefficient R of 0.8445 and an RMSE of 0.0457 m^3/m^3. These results are satisfactory and indicate only a slight decrease in estimation accuracy compared to the data before interpolation. Furthermore, the coverage of the interpolated data increased by an average of 1.8 times. The method proposed in this paper significantly improves data coverage while ensuring data accuracy, thus addressing the issue of missing daily CYGNSS observations in certain areas. The effective gap-filling method enhances data quality, which is crucial for other applications, such as consistent and accurate environmental monitoring, land cover analysis, and climate change research. The impact of land cover types will be conducted in future works covering large-scale and extensive areas.

Acknowledgment

The authors would like to thank to the SMAP and CYGNSS teams for providing the dataset used in this study.

References

[1] M. P. Clarizia, N. Pierdicca, F. Costantini, and N. Floury, "Analysis of CYGNSS data for soil moisture retrieval," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, pp. 2227–2235, Jul. 2019.

[2] S. Jin, G. P. Feng, and S. Gleason, "Remote sensing using GNSS signals: Current status and future directions," *Adv. Space Res.*, vol. 47, no. 10, pp. 1645–1653, doi: 10.1016/j.asr.2011.01.036, 2011.

[3] H. Kim and V. Lakshmi, "Use of cyclone global navigation satellite system (CyGNSS) observations for estimation of soil moisture," *Geophysical Res. Lett.*, vol. 45, pp. 8272–8282, 2018.

[4] A. Egido et al., "Airborne GNSS-R polarimetric measurements for soil moisture and above-ground biomass estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, pp. 1522–1532, May 2014.

[5] N. Najibi and S. Jin, "Physical reflectivity and polarization characteristics for snow and ice-covered surfaces interacting with GPS signals," *Remote Sens.*, vol. 5, no. 8, pp. 4006–4030, 2013, doi: 10.3390/rs5084006.

[6] M. Nabi, V. Senyurek, A. C. Gurbuz, and M. Kurum, "Deep learning-based soil moisture retrieval in CONUS using CYGNSS delay–Doppler maps," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6867–6881, 2022.

[7] C. Chew, R. Shah, C. Zuffada, G. Hajj, D. Masters, and A. J. Mannucci, "Demonstrating soil moisture remote sensing with observations from the U.K. TechDemoSat-1 satellite mission," *Geophysical Res. Lett.*, vol. 43, pp. 3317–3324, 2016.

[8] C. Chew, "Spatial interpolation based on previously-observed behavior: A framework for interpolating spaceborne GNSS-R data from CYGNSS," *J. Spatial Sci.*, vol. 68, pp. 155–168, 2023.

[9] M. P. Clarizia, C. S. Ruf, P. Jales, and C. Gommenginger, "Spaceborne GNSS-R minimum variance wind speed estimator," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, pp. 6829–6843, Nov. 2014.

[10] G. Foti et al., "Spaceborne GNSS reflectometry for ocean winds: First results from the U.K. TechDemoSat-1 mission," *Geophysical Res. Lett.*, vol. 42, pp. 5435–5441, 2015.

[11] V. Senyurek, F. Lei, D. Boyd, M. Kurum, A. C. Gurbuz, and R. Moorhead, "Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS," *Remote Sens.*, vol. 12, 2020, Art. no. 1168.

[12] V. Senyurek, F. Lei, D. Boyd, A. C. Gurbuz, M. Kurum, and R. Moorhead, "Evaluations of machine learning-based CYGNSS soil moisture estimates against SMAP observations," *Remote Sens.*, vol. 12, 2020, Art. no. 3503.
[13] C. Chew and E. Small, "Soil moisture sensing using spaceborne GNSS reflections: Comparison of CYGNSS reflectivity to SMAP soil moisture," *Geophysical Res. Lett.*, vol. 45, pp. 4049–4057, 2018.
[14] M. M. Al-Khaldi, J. T. Johnson, A. J. O'Brien, A. Balenzano, and F. Mattia, "Time-series retrieval of soil moisture using CYGNSS," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, pp. 4322–4331, Jul. 2019.
[15] Q. Yan, W. Huang, S. Jin, and Y. Jia, "Pan-tropical soil moisture mapping based on a three-layer model from CYGNSS GNSS-R data," *Remote Sens. Environ.*, vol. 247, 2020, Art. no. 111944.
[16] A. Calabia, I. Molina, and S. Jin, "Soil moisture content from GNSS reflectometry using dielectric permittivity from Fresnel reflection coefficients," *Remote Sens.*, vol. 12, 2020, Art. no. 122.
[17] T. Yang, W. Wan, Z. Sun, B. Liu, S. Li, and X. Chen, "Comprehensive evaluation of using TechDemoSat-1 and CYGNSS data to estimate soil moisture over mainland China," *Remote Sens.*, vol. 12, Art. no. 1699, 2020.
[18] F. Lei et al., "Quasi-global machine learning-based soil moisture estimates at high spatio-temporal scales using CYGNSS and SMAP observations," *Remote Sens. Environ.*, vol. 276, 2022, Art. no. 113041, doi: 10.1016/j.rse.2022.113041.
[19] C. S. Ruf et al., "A new paradigm in earth environmental monitoring with the CYGNSS small satellite constellation," *Sci. Rep.*, vol. 8, 2018, Art. no. 8782.
[20] C. Chew, J. T. Reager, and E. Small, "CYGNSS data map flood inundation during the 2017 Atlantic hurricane season," *Sci. Rep.*, vol. 8, 2018, Art. no. 9336.
[21] M. M. Al-Khaldi et al., "Inland water body mapping using CYGNSS coherence detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, pp. 7385–7394, Sep. 2021.
[22] C. Gerlein-Safdi and C. S. Ruf, "A CYGNSS-based algorithm for the detection of inland waterbodies," *Geophysical Res. Lett.*, vol. 46, pp. 12065–12072, 2019.
[23] P. Ghasemigoudarzi, W. Huang, O. De Silva, Q. Yan, and D. Power, "A machine learning method for inland water detection using CYGNSS data," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2020, Art. no. 8001105.
[24] M. Morris, C. Chew, J. T. Reager, R. Shah, and C. Zuffada, "A novel approach to monitoring wetland dynamics using CYGNSS: Everglades case study," *Remote Sens. Environ.*, vol. 233, 2019, Art. no. 111417.
[25] N. Rodriguez-Alvarez, E. Podest, K. Jensen, and K. C. McDonald, "Classifying inundation in a tropical wetlands complex with GNSS-R," *Remote Sens.*, vol. 11, 2019, Art. no. 1053.
[26] Y. Jia et al., "Temporal-spatial soil moisture estimation from CYGNSS using machine learning regression with a preclassification approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4879–4893, 2021.
[27] C. S. Ruf et al., "New ocean winds satellite mission to probe hurricanes and tropical convection," *Bull. Amer. Meteorol. Soc.*, vol. 97, pp. 385–395, 2016.
[28] O. Eroglu, M. Kurum, D. Boyd, and A. C. Gurbuz, "High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks," *Remote Sens.*, vol. 11, 2019, Art. no. 2272.
[29] F. Tang and S. Yan, "CYGNSS soil moisture estimations based on quality control," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 8022105.
[30] H. Carreno-Luengo, G. Luzi, and M. Crosetto, "Sensitivity of CyGNSS bistatic reflectivity and SMAP microwave radiometry brightness temperature to geophysical parameters over land surfaces," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 1, pp. 107–122, Jan. 2019.
[31] J. Chaubell, S. Yueh, D. Entekhabi, and J. Peng, "Resolution enhancement of SMAP radiometer data using the Backus Gilbert optimum interpolation technique," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 284–287.
[32] W. R. Tobler, "A computer movie simulating urban growth in the Detroit region," *Econ. Geogr.*, vol. 46, pp. 234–240, 1970.
[33] A. S. Fotheringham, C. Brunsdon, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Hoboken, NJ, USA: Wiley, 2003.
[34] A. Stewart Fotheringham, M. Charlton, and C. Brunsdon, "The geography of parameter space: An investigation of spatial non-stationarity," *Int. J. Geographical Inf. Syst.*, vol. 10, pp. 605–627, 1996.
[35] C. Brunsdon, A. S. Fotheringham, and M. E. Charlton, "Geographically weighted regression: A method for exploring spatial nonstationarity," *Geographical Anal.*, vol. 28, pp. 281–298, 1996.
[36] B. Huang, B. Wu, and M. Barry, "Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices," *Int. J. Geographical Inf. Sci.*, vol. 24, pp. 383–401, 2010.
[37] P. Song, J. Huang, and L. R. Mansaray, "An improved surface soil moisture downscaling approach over cloudy areas based on geographically weighted regression," *Agricultural Forest Meteorol.*, vol. 275, pp. 146–158, 2019.
[38] T. Chen and C. Guestrin, "XgBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794.
[39] Z. Li, F. Guo, F. Chen, Z. Zhang, and X. Zhang, "Wind speed retrieval using GNSS-R technique with geographic partitioning," *Satell. Navigation*, vol. 4, no. 1, 2023, Art. no. 4.
[40] S. Jin et al., "Remote sensing and its applications using GNSS reflected signals: Advances and prospects," *Satell. Navigation*, vol. 5, 2024, Art. no. 19, doi: 10.1186/s43020-024-00139-4.

Yan Jia (Member, IEEE) received the double M.S. degree in telecommunications engineering and computer application technology from Politecnico di Torino, Turin, Italy, and Henan Polytechnic University, Jiaozuo, China, in 2013. She received the Ph.D. degree in electronics engineering from Politecnico di Torino, Turin, Italy, in 2017.

She is currently working with Nanjing University of Posts and Telecommunications. In 2013, she was with the Department of Electronics and Telecommunications, Politecnico di Torino, Torino, Italy, where she performed research on the GNSS system construction and GNSS antenna analysis. In 2014, she worked on the SMAT Project, mainly focusing on the retrieval of soil moisture and vegetation biomass content by GNSS-R. Her research interests include microwave remote sensing, soil moisture retrieval, Global Navigation Satellite System Reflectometry applications to land remote sensing and antenna design.

Zhiyu Xiao received the B.S. degree in surveying and mapping engineering from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2022. He is currently working toward the master's degree in surveying and mapping science and technology at Nanjing University of Posts and Telecommunications.

He was engaged in the research of remote sensing data classification based on machine learning. He has been working on CYGNSS-based soil moisture retrieval and related applications with Nanjing University of Posts and Telecommunications. His main research interests include machine learning based applications and program development.

Shuanggen Jin (Senior Member, IEEE) was born in Anhui, China, in 1974. He received the B.Sc. degree in geodesy from Wuhan University, Wuhan, China, in 1999 and the Ph.D. degree in geodesy from the University of Chinese Academy of Sciences, Beijing, China, in 2003.

He is the Vice President and a Professor with Henan Polytechnic University, Jiaozuo, China, and also a Professor at Shanghai Astronomical Observatory, CAS, Shanghai, China. His main research areas include satellite navigation, remote sensing and space/planetary exploration. He has published more than 500 papers in peer-reviewed journals and proceedings, ten patents/software copyrights and 10 books/monographs with more than 8000 citations and H-index>50.

Dr. Jin has been the President of International Association of Planetary Sciences (IAPS) (2015–2019), the President of the International Association of CPGPS (2016–2017), Chair of IUGG Union Commission on Planetary Sciences (UCPS) (2015–2023), Editor-in-Chief of International Journal of Geosciences, Editor of Geoscience Letters, Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE & REMOTE SENSING and *Journal of Navigation*, Editorial Board member of Remote Sensing, GPS Solutions and *Journal of Geodynamics*. He has received 100 Talent Program of CAS, Leading Talent of Shanghai, IAG Fellow, IUGG Fellow, a fellow of Electromagnetics Academy, a World Class Professor of Ministry of Education and Cultures, Indonesia, the Chief Scientist of National Key R&D Program, China, a Member of Russian Academy of Natural Sciences, a member of European Academy of Sciences, a member of Turkish Academy of Sciences, and a member of Academia Europaea.

Qingyun Yan (Member, IEEE) was born in Haimen, China. He received the B.Eng. degree in electronic science and engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2014, and the M.Eng. and Ph.D. degrees in electrical engineering from the Memorial University of Newfoundland, St. John's, NL, Canada, in 2015 and 2020, respectively.

He is now with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology. His research interests include tsunami, sea ice, and land remote sensing using Global Navigation Satellite System Reflectometry.

Dr. Yan was a recipient of the 2019 IEEE GRSS Letters Prize Paper Award from the IEEE Geoscience and Remote Sensing Society.

Yan Jin received the B.S. degree in information and computation science and the M.S. degree in applied mathematics, from Chang'an University, Xi'an, China, in 2011 and 2014, respectively, and the Ph.D. degree in cartography and geographical information system from the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, University of Chinese Academy of Sciences, Beijing, China, in 2018.

She is currently a Lecturer with the School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing, China. Her current research interests focus on scale transformation, data fusion, geostatistics, and remote sensing applications.

Wenmei Li (Member, IEEE) received the M.S. degree in cartography and GIS from Nanjing University, in 2010, and the Ph.D. degree from China Academy of Forestry Sciences, in 2013, respectively.

She is an Associate Professor with the School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications. She is working for her postdoctoral studies (2018-) in Nanjing University of Posts and Telecommunications. Her research interests include deep learning, optimization, image reconstruct, and their application in land remote sensing.

Patrizia Savi (Senior Member, IEEE) received the laurea degree in electronic engineering from the Politecnico di Torino, Turin, Italy, in 1985.

In 1986, she was a Consultant in Alenia (Caselle Torinese, Italy) where she conducted research on the analysis and design of dielectric radomes. From 1987 to 1998, she was a Researcher with the Italian National Research Council. In 1998, she joined the Electronic Department, Politecnico di Torino, as an Associate Professor. She currently teaches a course on electromagnetic field theory. Her areas of interest include dielectric radomes, frequency-selective surfaces, waveguide discontinuities and microwave filters, high-altitude platform propagation channels, Global Navigation Satellite System Reflectometry for soil moisture retrieval. Recently, she is focused on the analysis and characterization at microwave frequency of novel materials (polymers and cements with carbon nanotubes, graphene or biochar as fillers) for various applications.

Prof. Savi is currently senior member of IEEE Society and a member of SIEM (Società Italiana di Elettromagnetismo).