



Article

Using Robust Regression to Retrieve Soil Moisture from CyGNSS Data

Qi Liu ¹, Shuangcheng Zhang ^{1,2,*} , Weiqiang Li ^{3,4} , Yang Nan ⁵ , Jilun Peng ³, Zhongmin Ma ¹ and Xin Zhou ¹

¹ College of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054, China; liuqi.vip@chd.edu.cn (Q.L.)

² State Key Laboratory of Geo-Information Engineering, Xi'an 710054, China

³ Institute of Space Sciences (ICE-CSIC), 08193 Barcelona, Spain; weiqiang@ice.csic.es (W.L.)

⁴ Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain

⁵ School of Marine Science and Technology, Tianjin University, Tianjin 300072, China

* Correspondence: shuangcheng369@chd.edu.cn

Abstract: Accurate global soil moisture (SM) data are crucial for modeling land surface hydrological cycles and monitoring climate change. Spaceborne global navigation satellite system reflectometry (GNSS-R) has attracted extensive attention due to its unique advantages, such as faster revisit time, lower payload costs, and all-weather operation. GNSS signal reflected at L-band also has significant advantages for SM estimation. Usually, SM is estimated based on the sensitivity of GNSS-R reflectivity to SM, but the noise in observations can significantly impact SM estimation results. A new SM retrieval method based on robust regression is proposed to address this issue in this work, and the effects of roughness and vegetation on the effective reflectivity of the Cyclone Global Navigation Satellite System (CyGNSS) are reconsidered. Ancillary data are provided by the SM Active Passive (SMAP) mission. The retrieved results from the training sets and test sets agree well with the referenced SMAP SM data. The correlation coefficient R is 0.93, the root mean square error (RMSE) is $0.058 \text{ cm}^3 \text{ cm}^{-3}$, the unbiased RMSE (ubRMSE) is $0.042 \text{ cm}^3 \text{ cm}^{-3}$, and the mean absolute error (MAE) is $0.040 \text{ cm}^3 \text{ cm}^{-3}$ in the training sets. For the test, the correlation coefficient is 0.91, the RMSE is $0.067 \text{ cm}^3 \text{ cm}^{-3}$, the ubRMSE is $0.051 \text{ cm}^3 \text{ cm}^{-3}$, and the MAE is $0.044 \text{ cm}^3 \text{ cm}^{-3}$. The proposed method has been evaluated using in situ measurements from the SMAP/in situ core validation site; in situ measurements and retrieval results exhibit good consistency with the ubRMSE value below $0.35 \text{ cm}^3 \text{ cm}^{-3}$. Moreover, the SM retrieval results using robust regression methods show better performance than CyGNSS official SM products that use linear regression. In addition, the land cover types significantly affect the accuracy of SM retrieval, and the incoherent scattering in densely vegetated areas (tropical forests) usually leads to more errors.

Keywords: soil moisture; CyGNSS; reflectivity; vegetation; roughness



Citation: Liu, Q.; Zhang, S.; Li, W.; Nan, Y.; Peng, J.; Ma, Z.; Zhou, X. Using Robust Regression to Retrieve Soil Moisture from CyGNSS Data. *Remote Sens.* **2023**, *15*, 3669. <https://doi.org/10.3390/rs15143669>

Academic Editors: Yajun Zhu, Jiyao Xu and Weihua Bai

Received: 8 June 2023

Revised: 15 July 2023

Accepted: 21 July 2023

Published: 23 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The crucial impact of soil moisture (SM) on the climate system, hydrologic processes, and vegetation growth has long been recognized and was named one of the essential climate variables (ECVs) in 2010 [1,2]. At present, two L-band passive microwave radiometers provide global SM measurements: the SM Active Passive (SMAP) mission and the SM and Ocean Salinity (SMOS) mission, both of which have a coarse spatial resolution of about 40 km and a revisit period of about 2–3 days, and provide SM retrieval at a depth of 5 cm on the soil surface [3]. SM can also be measured using synthetic aperture radar (SAR) systems, such as Sentinel-1 and the upcoming NASA-ISRO SAR (NISAR) mission [4], which can provide SM data with high spatial resolution. However, the temporal resolution is limited, and both are severely affected by vegetation structure and rough surfaces.

In recent decades, Global Navigation Satellite System Reflectometry (GNSS-R) has demonstrated significant potential for retrieving geophysical parameters. Although the

technique was originally designed to measure sea winds in tropical oceans [5], some recent studies and projects have demonstrated the sensitivity of reflected signals from the land surface to hydrological parameters, such as SM [6–8], mapping flood [9,10], wetlands [11,12], and vegetation [13,14]. Retrieving SM from spaceborne GNSS-R data has become a research hotspot due to its unique advantages (such as faster revisit time, all-weather operation, and lower payload costs). Existing SM estimation methods can be roughly divided into two categories: empirical models and machine learning (ML). ML methods mostly characterize the complex nonlinear relationship between GNSS-R observations (such as reflectivity and SNR) and reference SM through the inclusion of land surface characteristics, such as random forest [15,16], neural networks [17,18], and XGBoost approaches [19], and so on. ML requires a large number of GNSS-R observations and a great quality of ancillary data for a long period of training to obtain the inversion model, and it is difficult to explain the relationship between the involved variables. On the contrary, model-based SM estimation does not rely on a large amount of ancillary data and can clearly explain the relationship between GNSS-R observations and desired SM. Clarizia et al. [8] and Yan et al. [20] developed a trilinear regression model and multiple linear regression model to estimate SM, explaining the relationship between reflectivity, vegetation, and roughness. Based on the relationship between Cyclone GNSS (CyGNSS) reflectivity and SM using linear regression, Chew and Small [21] developed UCAR (University Corporation for Atmospheric Research) SM products (<https://cmr.earthdata.nasa.gov/virtual-directory/collections/C2205122332-POCLOUD>, accessed on 2 June 2023). However, the retrieval results using linear regression are affected by observation noise, such as the uncertainty in the antenna pattern correction (0.25 dB), GPS transmitter effective isotropic radiated power (EIRP) error (0.24 dB), noise floor uncertainty (~0.3 dB), and other noise [22]. Additionally, previous research on SM retrieval has not taken into account the fact that microwave-scale roughness (h) is insufficient to simulate the impact of roughness on reflectometry [23]. The vegetation optical depth (VOD) used for simulating vegetation attenuation effects needs to be reconsidered due to the wider observation angle of CyGNSS [24]. Al-Khalidi et al. presented a time-series model of SM obtained from the CyGNSS, and this method can offset vegetation attenuation by continuously measuring the ratio of CyGNSS DDM [25]. However, the roughness effect was solved using a mean square slope with a constant of 0.01 for all observations.

In this study, a new SM estimation method using robust regression is proposed to address the impact of any noise in CyGNSS observations on classical linear regression methods. Unlike studies [8,20] using a unified model to handle large-scale areas, the model in this work is parameterized with a pixel mode (36×36 km). In addition, this work has reconsidered the impact of surface roughness and vegetation attenuation on reflectometry. The remainder of this paper is organized as follows: Collected data are described in Section 2. The method and data process used in this study are shown in Section 3. Section 4 describes and discusses the experimental result. Section 5 discusses the results of the study, and Section 6 summarizes the work.

2. Datasets

In this section, the collection and quality control (QC) of CyGNSS data, as well as the use of referenced SMAP SM, in situ SM, and ancillary data are described.

2.1. CyGNSS Data

CyGNSS was launched on 15 December 2016. It is made up of eight microsatellites and is known for its nearly gapless earth coverage. An orbital inclination of about 35° from the equator can provide GNSS-R observations in the pantropical regions and allows CyGNSS to measure the reflected signals between the latitudes of about 38° N and 38° S, with a minimum spatial resolution of about $3.5 \text{ km} \times 0.5 \text{ km}$ and a revisit time of about several hours [26,27].

The reflected GPS signals from the land surface are recorded by a delay Doppler mapping instrument (DDMI) at CyGNSS satellites, and DDMs are processed by cross-correlating with a replica of the receivers and reflected GPS signals. CyGNSS bistatic radar cross section (BRCS) is a $17 \text{ Delay} \times 11 \text{ Doppler}$ array obtained by inverting the forward scattering model, and its peak value varies with SM, vegetation, roughness, etc. [15]. In addition to BRCS, the metadata related to CyGNSS DDM and the observational geometry are used here, such as incidence angles, signal-to-noise ratio (SNR), gain, geographic location of the specular point (SP), and the distance of the transmitter and receiver from the specular point. The CyGNSS Level 1 (L1) v3.0 datasets from 2019 to 2020 are used in this work (available at <https://cmr.earthdata.nasa.gov/virtual-directory/collections/C2205618435-POCLOUD>, accessed on 2 June 2023). For more variable descriptions of CyGNSS L1 V3.0 data, please see the CyGNSS User Guide [26] and Data Dictionary (https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-docs/cygnss/open/L1/docs/148-0346-8_L1_v3.0_netCDF_Data_Dictionary.xlsx, accessed on 2 June 2023). The data for 2019 are used to generate the roughness lookup table (LUT), and the data for 2020 are used to retrieve SM.

To filter low-quality DDM observations, similar procedures from previous studies [19–21] are referenced. CyGNSS data collected over land with SNR over 2 dB, antenna gain more than 0 dB (towards SP), and incidence angles lower than 65° are applied here. In addition, the quality flags of DDM are used, and the applied quality flags are listed in Table 1. Figure 1 shows an example of gridded mean reflectivity (January 2020) before and after QC. To avoid too many blanks on the map, the scope of maps was appropriately adjusted to longitude = $[-130, 160]$ and latitude = $[-38, 38]$. The remaining maps are processed in the same manner. Note that to obtain the roughness information at all grids, we did not apply QC to 2019 data.

Table 1. Applied quality flags in 2020 data processing.

Quality Flag Number	Quality Flag Name
2	S-Band Powered Up
4	Large Spacecraft Attitude Error
5	Blackbody DDM
6	DDMI Reconfigured
7	Space wire CRC Invalid
8	DDM is Test Patten
9	Channel Idle
16	Direct Signal in DDM
17	Low Confidence GPS EIRP Estimate
18	RFI Detected
22	GPS PVT sp3 error
23	SP Non-Existent Error
26	Blackbody Framing Error

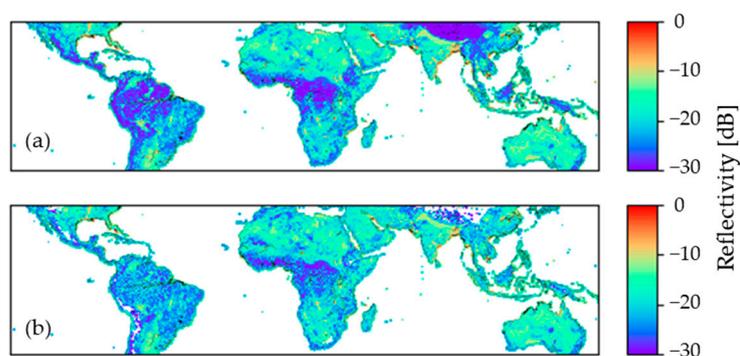


Figure 1. Example of CyGNSS gridded reflectivity (a) before and (b) after QC; data from January 2020 are used.

2.2. SMAP Data

The SM derived from CyGNSS will be referenced and compared using the SMAP L3 Radiometer Global Daily 36 km equal-area scalable earth grid (EASE-Grid) version 8 SM data (SMAP_L3_SM_P). The data are freely available through the National Snow and Ice Data Center (NSIDC) at <https://nsidc.org/data/data-access-tool/SPL3SMP/versions/8/> (accessed on 2 June 2023). Although clay ratio (CR) and vegetation water content (VWC) from SMAP ancillary data were used (see Figure 2), CR and VWC data are independent of SMAP. Specifically, CR data (Soil texture) are assembled from an optimized combination of multiple global or regional soil databases, and VWC data are calculated based on the Normalized Difference Vegetation Index (NDVI) information provided by moderate-resolution imaging spectroradiometer (MODIS) [3]. The period of SMAP data used corresponds to the CyGNSS observations, i.e., 2019 to 2020. CyGNSS data are spatially averaged into the 36 km EASE-Grid used to facilitate the experimental evaluation in the following section.

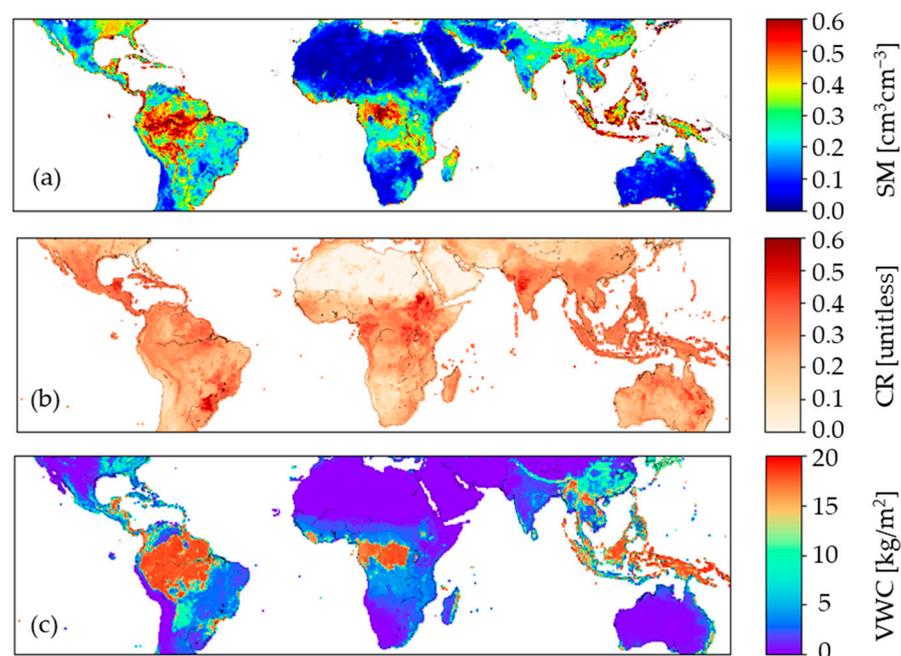


Figure 2. Example of 3 days (1–3 January 2020) aggregated (a) SM, (b) CR, and (c) VWC from SMAP ancillary data.

2.3. In Situ Data

To investigate and evaluate the accuracy of CyGNSS SM retrievals on the temporal scale and avoid the issue of mismatch in resolution between retrievals and in situ, the SMAP/in situ core validation site land surface parameters match-up datasets were used here [28]. Please visit the SMAP cal/val activities website (<https://smap.jpl.nasa.gov/science/validation/calvalpartners/>, accessed on 30 June 2023) for more details. The dataset is produced by matching SMAP SM products with in situ SM estimates from core validation sites. These data have been used to evaluate the performance of SMAP SM products in different regions, such as arid/semi-arid regions [29] and forested sites [30]. Thus, this dataset is reliable in validating CyGNSS SM retrievals. SMAP cal/val in situ data include not only the surface SM (0–5 cm) and soil temperature but also the verifiable grid scale of the station (3 km, 9 km, and 36 km). Based on the CyGNSS observation range, available periods, and the verifiable grid scale of sites (36 km), 4 in situ sites were chosen as validation data. Refer to Figure 3 for the distribution of sites and to Table 2 for detailed information on the sites. Note that soil temperature is an important dynamic auxiliary parameter because SM sensors do not provide reliable measurements when the soil temperature is below 4 degrees Celsius [31], so these data were removed during validation.

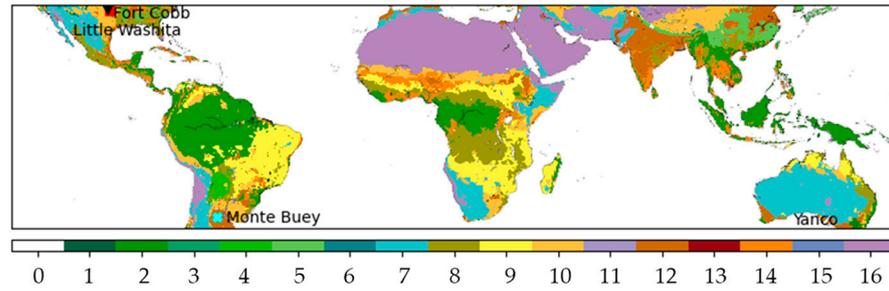


Figure 3. Distribution of SMAP cal/val sites used in this work. Different stations are represented in different colors and dot styles. The background map is the International Geosphere-Biosphere Programme (IGBP) land cover types with different colors and numbers. Specifically, 0: water; 1: evergreen needleleaf forest; 2: evergreen broadleaf forest; 3: deciduous needleleaf forest; 4: deciduous broadleaf forest; 5: mixed forest; 6: closed shrublands; 7: open shrublands; 8: woody savannas; 9: savannas; 10: grasslands; 11: permanent wetlands; 12: croplands; 13: urban and built-up; 14: cropland/natural vegetation mosaic; 15: snow and ice; 16: barren or sparsely vegetated.

Table 2. Details of in situ sites.

Site Name	Longitude, Latitude	IGBP Land Cover	Climate
Fort Cobb	35.36°N, 98.55°W	Grasslands	Temperate
Little Washita	34.97°N, 97.97°W	Grasslands	Temperate
Monte Buey	32.96°S, 62.52°W	Croplands	Arid
Yanco	34.8°S, 146.11°E	Croplands/Grasslands	Semi-arid

3. Methods

The section explains how to estimate and calibrate reflectivity from CyGNSS data, then goes into detail about the roughness LUT, and finally, describes a complete data processing flow.

3.1. SM Retrieval Method

The CyGNSS effective reflectivity is the observable used for SM retrieval. Assuming that the reflected signal over land is mostly determined by the coherent reflection [7,8,10], it is calculated as follows:

$$P_{coh} = \frac{\lambda^2 P_t G_t G_r}{(4\pi)^2 (r_{st} + r_{sr})^2} \Gamma_{eff}(\theta), \quad (1)$$

where θ is the incidence angle; P_{coh} is the peak value of the DDM of the analog scattered power; λ is the wavelength of GPS, about 19 cm; P_t is transmitter power; G_t is transmitter gain, and $P_t G_t$ is the transmitter equivalent isotropically radiated power (EIRP); G_r is CyGNSS antenna gain in the direction of SP; r_{st} and r_{sr} are the distances from SP to the transmitter and receiver, respectively.

The effective reflectivity can be modeled as

$$\Gamma_{eff}(\theta) = \mathfrak{R}_{RL}^2(\epsilon, \theta) \Gamma^2 \exp(-h^2 \cos^2 \theta) = \Gamma_{SM} \times \Gamma_{Veg} \times \Gamma_{Rou}, \quad (2)$$

where $\Gamma_{SM} = \mathfrak{R}_{RL}^2(\epsilon, \theta)$, $\Gamma_{Veg} = \Gamma^2$, and $\Gamma_{Rou} = \exp(-h^2 \cos^2 \theta)$ describe the effect of SM, vegetation, and roughness, respectively. \mathfrak{R}_{RL} is the surface's Fresnel reflection coefficient, and ϵ indicates the soil dielectric constant. Transmissivity Γ represents the signal attenuation as the signals propagate through the vegetation canopy, and is a function of vegetation optical depth (VOD) τ ,

$$\Gamma = \exp(-\tau) \quad (3)$$

VOD related to vegetation water content (VWC), vegetation attenuation parameter (b), and incidence angle,

$$\tau = b \times VWC / \cos \theta \quad (4)$$

A static global gridded map of b based on land cover types was successfully applied to the SMAP mission [32]. However, parameter b depends not only on the land cover types but also on the observation angle (GNSS-R satellites observed at a wider angle of about $0\sim 75^\circ$, SMAP observed at a fixed angle of around 40°). Parameter b estimated by [24] and VWC data are used here to calculate vegetation effects. The exponential term in Equation (2) accounts for the surface roughness effects, $h = 2ks$ denotes assumed linearly related to the root mean squared height (RMSH) s of the land surface, k is the wave number in free space. The influence of surface roughness on the reflectometer is much larger than that on passive microwave radiometry, and it is usually not appropriate to directly use the h parameter from SMAP for CyGNSS roughness modeling [23]. Here, a semi-empirical model is used to estimate the roughness LUT (Γ_{Rou} [dB]), as described in Section 3.2. Taking the logarithm in Equation (2), it is as follows:

$$\Gamma_{eff} = \Gamma_{SM} + \Gamma_{Veg} + \Gamma_{Rou}, \text{ unit in dB.} \quad (5)$$

By inverting Equation (5), the calibrated reflectivity can be obtained, that is, $\Gamma_{SM} = \Gamma_{eff} - \Gamma_{Veg} - \Gamma_{Rou}$. The calibrated reflectivity is converted to SM using the concept proposed in [21]. Therefore, a linear model between the variations of reflectivity ($\Delta\Gamma_{SM}$) and SM (ΔSM) is established and performed in each pixel manner. It can be expressed as

$$\Delta SM = f(\Delta\Gamma_{SM}) = \beta \times \Delta\Gamma_{SM} + \alpha, \quad (6)$$

$$SM_{cyg} = \Delta SM + \overline{SM}, \quad (7)$$

where β and α are coefficients that need to be determined by robust regression, SM_{cyg} is retrieved SM from CyGNSS, and \overline{SM} is mean SMAP SM in each grid. Considering the sample size (one grid), regression speed, and the ability to handle outliers, the Random Sample Consensus (RANSAC) regression is selected for parameter estimation. RANSAC is a robust model-fitting algorithm that aims to find a solution that is resistant to outliers in the data by iteratively fitting models to randomly sampled subsets of the input data. The RANSAC Regressor, which is interfaced via the “scikit-learn” library within the Python framework [33], provides a convenient solution and easy-to-use environment, which was used in this work.

The flowchart of the proposed method is shown in Figure 4. The 2020 CyGNSS data and corresponding auxiliary data (SMAP VWC and SM) were used for SM retrieval. In order to weaken the impact of seasonal changes on retrieval results, the 2020 CyGNSS data were randomly divided into two parts: one part was used for the training phase (estimation of coefficients β and α), and the remaining for testing.

3.2. Estimation of Roughness

The h parameter (microwave-scale surface roughness) of passive microwave radiometry is too small for the modeling of reflectometry; it may not apply to SM retrieval from GNSS-R [23]. Previous research on SM retrieval did not take that into account; here, we use the semi-empirical model proposed by [24] to estimate the roughness effect, which includes the combined effects of microroughness and terrain. The semi-empirical model is expressed as

$$\Gamma_{eff}(\theta) = \Gamma^2(\alpha + f_s S) \mathfrak{R}_{RL}^2(\epsilon, \theta) + \Gamma_{v0} S, \quad (8)$$

where $\alpha = \exp(-4k^2 s^2 \cos^2 \theta)$ represents the loss of coherent power due to microwave-scale roughness, and $f_s S$ accounts for the contributions of incoherent scattering, depending on

topography. S is the scale factor related to the distances from SP to the transmitter and receiver. It can be expressed as

$$S = A \frac{(r_{st} + r_{sr})^2}{4\pi r_{rs}^2 r_{ts}^2}. \tag{9}$$

where A is the effective reflection area related to the Woodward ambiguity function (WAF). Γ_{v0} is an empirical parameter, and the correction used here is $\Gamma_{v0} = -30.5$ dB for $VWC \geq 5$ kg/m² and $\Gamma_{v0} = 0$ for $VWC < 5$ kg/m².

Equation (8) can be inverted to

$$(\alpha + f_s S) = \frac{\Gamma_{eff}(\theta) - \Gamma_{v0} S}{\Gamma^2 \mathfrak{R}_{RL}^2(\epsilon, \theta)}, \tag{10}$$

and take the logarithm of its left and right sides to obtain the roughness LUT (Γ_{Rou}), which is

$$\Gamma_{Rou} = 10 \log(\alpha + f_s S) = 10 \log(\Gamma_{eff}(\theta) - \Gamma_{v0} S) - 20 \log(\Gamma) - 20 \log(\mathfrak{R}_{RL}(\epsilon, \theta)). \tag{11}$$

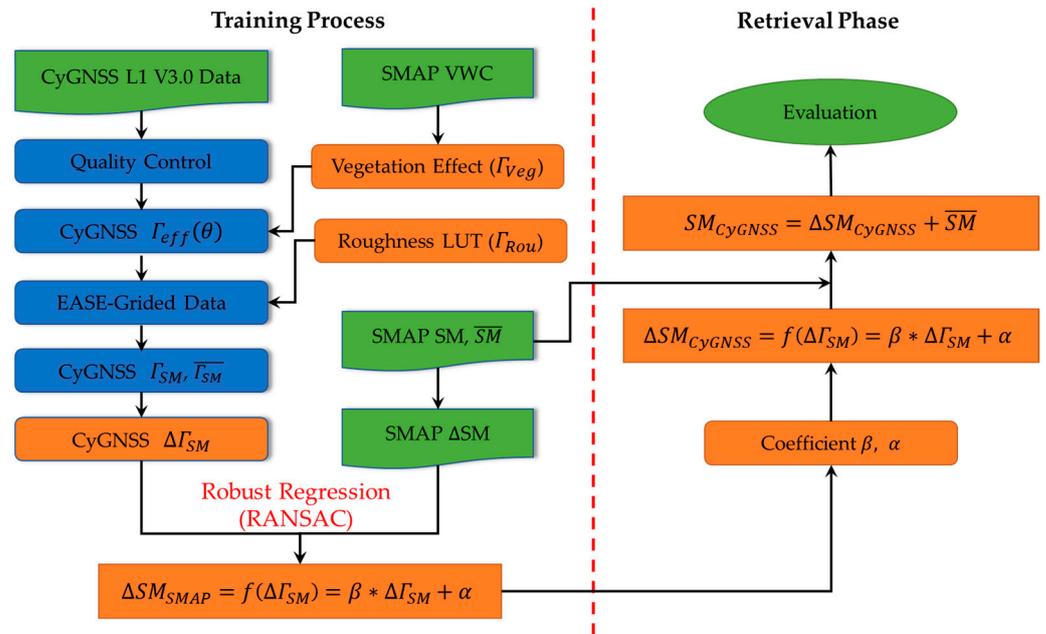


Figure 4. The flowchart of the proposed method.

The flowchart for estimating roughness LUT is presented in Figure 5. The SM and CR from SMAP data in 2019 are used as the input of the Mironov dielectric model to calculate the Fresnel coefficient (\mathfrak{R}_{RL}). Then, the SMAP VWC is input into the semi-empirical model to calculate the vegetation attenuation (Γ) and scattering ($\Gamma_{v0}S$) effects. As a result, a roughness LUT with EASE-Grid is created, assuming that it does not change with time. It should be noted that the one-day CyGNSS observations and the mean values of three-day SMAP data match because of the SMAP revisit time (2–3 days).

3.3. Validation Metrics

We will evaluate the accuracy of CyGNSS SM retrievals SM_{cyg} using reference SM SM_{ref} , including SMAP L3 SM (SMAP_L3_SM_P), in situ SM, and UCAR SM. The evaluation will involve four metrics: correlation coefficient (R), mean absolute error (MAE), root mean square error (RMSE), and unbiased root mean square error (ubRMSE). The correlation

coefficient R indicates the relative accuracy between the retrieved SM and reference values. The calculation of correlation coefficient R is as follows:

$$R = \frac{1}{N-1} \sum_{i=1}^N \left[\left(\frac{SM_{cyg} - \overline{SM_{cyg}}}{\sigma_{cyg}} \right) \left(\frac{SM_{ref} - \overline{SM_{ref}}}{\sigma_{ref}} \right) \right], \quad (12)$$

where N represents the total number of samples, $\overline{SM_{cyg}}$ and $\overline{SM_{ref}}$ are mean CyGNSS SM and mean reference SM, σ_{cyg} and σ_{ref} are the standard deviation of CyGNSS SM and reference SM, respectively.

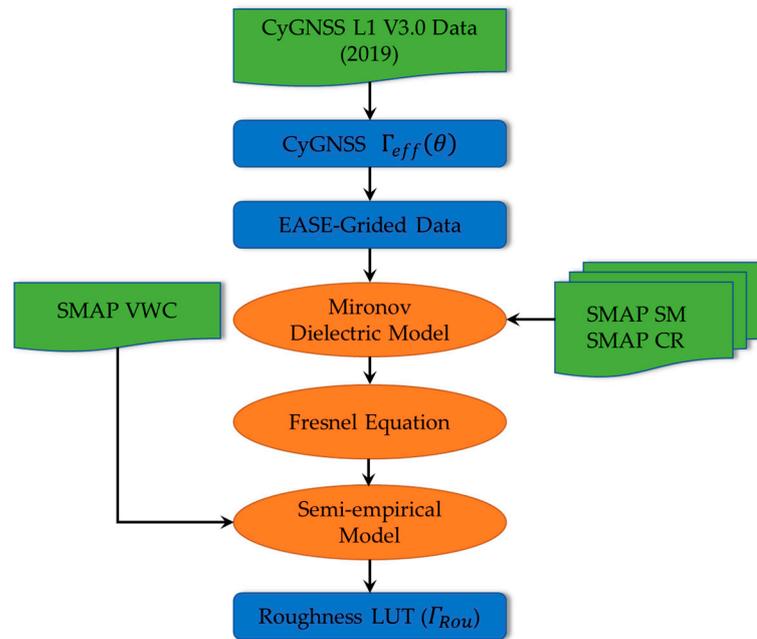


Figure 5. The flowchart for roughness LUT estimation, (dB).

The MAE, given by Equation (13), represents the average absolute difference between the CyGNSS SM and the reference data.

$$MAE = \frac{1}{N} \sum_{i=1}^N |SM_{cyg} - SM_{ref}| \quad (13)$$

The RMSE is a measure of random error and represents the accuracy between the retrieved SM and the reference SM. The calculation of RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (SM_{cyg} - SM_{ref})^2} \quad (14)$$

But, RMSE essentially depends on the bias. To obtain a more reliable estimate of RMSE, the unbiased metric ubRMSE is used to evaluate the reliability of CyGNSS retrievals [34,35].

$$ubRMSE = \sqrt{RMSE^2 - (\overline{SM_{cyg}} - \overline{SM_{ref}})^2} \quad (15)$$

4. Results

The roughness LUT and coefficient estimation process are presented in this section, and then the CyGNSS SM retrievals are evaluated and validated for comparison with SMAP, in situ measurements, and UCAR SM products. The influence of land cover on CyGNSS SM retrieval is also discussed.

4.1. Calibration of Reflectivity

As explained in Section 3.1, the daily reflectivity obtained from CyGNSS is gridded using the approach of 36 km EASE-Grid, which means one sample presents one pixel with 36×36 km resolution. Figure 6a shows the annual mean effective reflectivity that was estimated from CyGNSS data in 2020; Figure 6b, roughness LUT; Figure 6c, vegetation effect; and Figure 6d, calibrated reflectivity. In other words, Figure 6 shows the product of the whole 2020 data processing, which is consistent with Equation (5), namely, Figure 6a = Figure 6b + Figure 6c + Figure 6d, (dB).

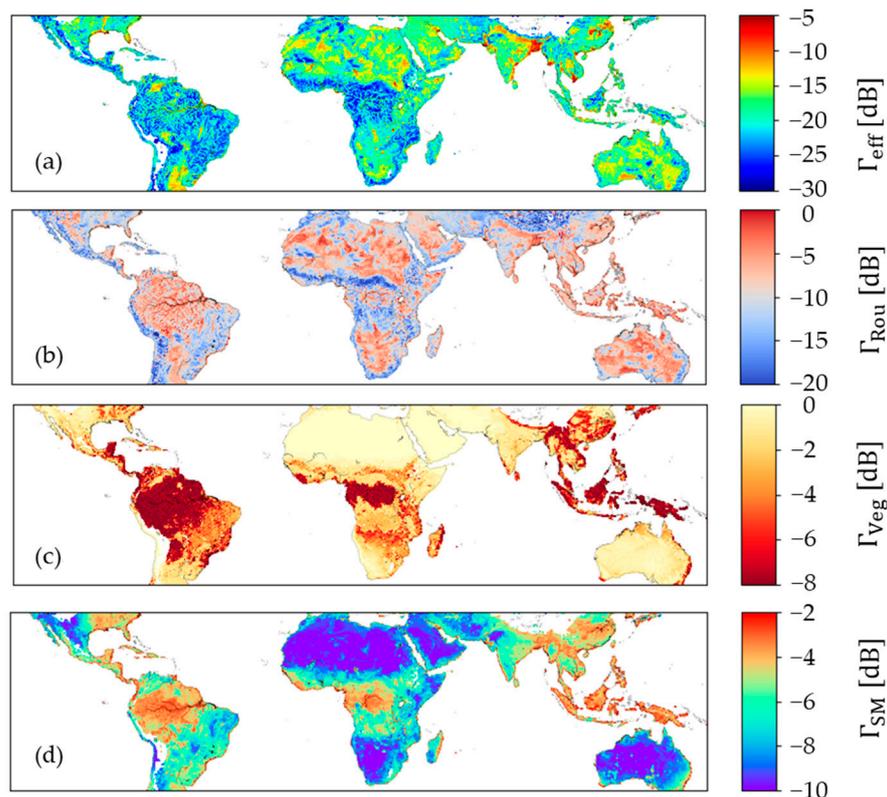


Figure 6. Calibrated CyGNSS reflectivity for 2020 is based on Equation (5). (a) The effective reflectivity (Γ_{eff}); (b) roughness LUT (Γ_{Rou}), assuming it does not change over time; (c) vegetation effect (Γ_{Veg}); (d) calibrated reflectivity (Γ_{SM}).

The Sahara Desert, the Arabian Desert, the Indian croplands, and most of Australia have higher reflectivity in Figure 6a, which ranges from -15 to -5 dB. These regions have limited vegetation and dry soil; thus, the primary source of reflectivity loss is surface roughness. In contrast, the reflectivity of CyGNSS is centered at -30 to -20 dB in the Amazon, Congo, and Southeast Asia. The vegetation is dense in these areas, and VWC is greater than 15 kg/m^2 (as shown in Figure 2c), indicating that the vegetation causes a loss of reflectivity. The roughness LUT is displayed in Figure 6b. The roughness effects depend on the position, whereby smoother surfaces lead to signals being closer to coherent reflection, while rougher surfaces cause stronger scattering. The high values (> -10 dB) are mostly located over deserts and croplands with lower surface roughness. The low values (< -10 dB) are mostly located over areas with complex topography, such as the Rocky Mountains, Andes Mountains, Tibetan plateau, etc. This demonstrates the influence of topography, that is, using only small-scale surface roughness (e.g., SMAP h -parameter) from a radiometer is insufficient to simulate surface scattering. Figure 6c shows the two-way attenuation of vegetation. Figure 6d displays the calibrated reflectivity (Γ_{SM}) for the SM retrieval using robust regression in the next step.

Figure 7 illustrates the annual average differences in CyGNSS reflectivity before and after calibration. Smaller improvements originate from sparsely vegetated or bare soil regions, such as the Sahara Desert, Arabia, South Africa, and most parts of Australia, with the improvement centered around 0–10 dB. The primary contribution to these differences is the roughness calibration. Larger differences arise from densely vegetated areas (Amazon, Congo, and Southeast Asia) or complex terrain regions (eastern Australia, western Americas), with the improvements concentrated in the range of 10–20 dB. The primary contribution is vegetation calibration or the combined effect of vegetation and surface roughness.

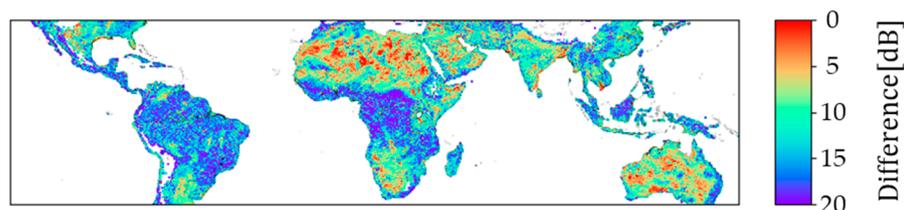


Figure 7. The difference in CyGNSS reflectivity before and after the calibration, (dB).

4.2. Estimation of Coefficients

A well-known robust estimator of the slope is the RANSAC estimator, which was first proposed by Fischler and Bolles [36]. RANSAC iteratively computes the residuals of random subsets within each grid cell to classify all data as either signal or noise, and subsequently uses the signal to estimate the parameters of the model. Its advantages over linear regression include a greater capacity for dealing with noise and outliers in the data and a greater robustness to perturbations. Figure 8 shows an example to illustrate the advantages of robust regression. It can be found that robust regression is more consistent with the trend of data, whereas the results of linear regression are biased by outliers. This sets it apart from other similar studies [21]. Another distinction is that this method of parameter estimation is conducted within each grid cell, as opposed to employing a single parameter across a broad scale [8,20].

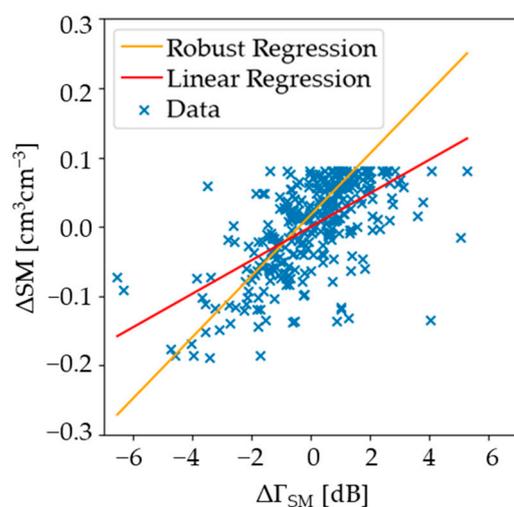


Figure 8. Examples of data fitting based on robust regression and linear regression. The data come from an EASE-Grid (36.3758°N, −76.7427°W).

Robust regression was used to estimate the coefficients in each separate grid. It should be pointed out that when there were fewer than 10 collocated observations in a grid, results of robust regression were removed in this work. To confirm the contribution of reflectivity to SM retrieval and the effectiveness of the method used, we calculated the correlation between $\Delta\Gamma_{SM}$ and ΔSM , as shown in Figure 9, from which strong positive correlations can

be obtained for most regions. However, most wet areas have poor correlation, such as the Amazon, Congo, and Southeast Asia. Coefficients β of the robust regression between ΔI_{SM} and ΔSM represent the sensitivity of CyGNSS to SM, as illustrated in Figure 10. Lower β have been found in dry areas with almost constant SM, e.g., North Africa and Western Australia. Humid areas with a lower correlation also showed lower β . Although some data in these areas have been filtered out after applying QC (SNR > 2 dB), the remaining data may not be reliable for SM retrieval. The IGBP land cover type of each EASE-Grid cell is shown in Figure 3. The land classification map is based on the most dominant land types with more than 50% coverage in each EASE-Grid. Average coefficients β over different land cover types, VWC, and SM are listed in Table 3. Coefficient β increases with the increase in mean VWC and SM; however, the land type of forests exhibit inconsistency with coefficient β decreasing as VWC and SM increase.

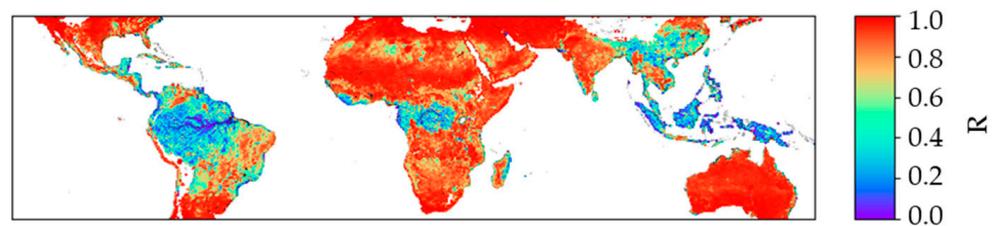


Figure 9. The correlation coefficient between ΔI_{SM} and ΔSM .

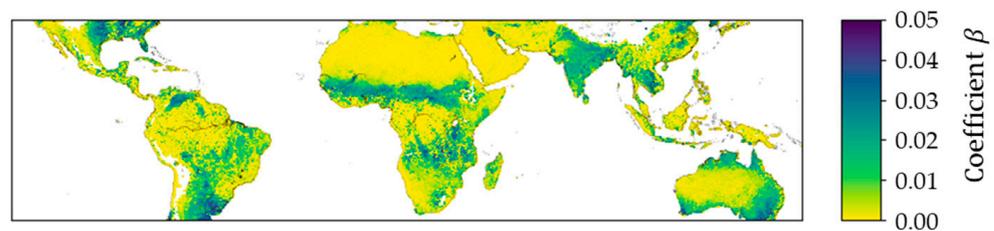


Figure 10. Determined coefficients β .

Table 3. Mean values of VWC, SM, coefficients β , and statistical error over different land types.

Land Cover Type	Mean VWC (kg/m ²)	Mean SM (cm ³ /cm ³)	Mean β	RMSE (cm ³ cm ⁻³)	ubRMSE (cm ³ cm ⁻³)	MAE (cm ³ cm ⁻³)
Barren/Sparsely Vegetated	0.01	0.0708	0.0010	0.021	0.016	0.014
Open Shrublands	0.44	0.1047	0.0045	0.037	0.031	0.026
Grasslands	1.01	0.1628	0.0121	0.050	0.043	0.040
Croplands	2.00	0.2478	0.0150	0.051	0.045	0.039
Savannas	2.77	0.1731	0.0113	0.054	0.048	0.045
Cropland/Natural Vegetation	3.57	0.2358	0.0118	0.056	0.051	0.046
Woody Savannas	4.28	0.2653	0.0117	0.059	0.050	0.049
Deciduous	9.31	0.2472	0.0104	0.066	0.056	0.056
Broadleaf Forest	9.49	0.3809	0.0054	0.072	0.059	0.058
Mixed Forest	15.99	0.4257	0.0024	0.079	0.062	0.066
Evergreen Broadleaf Forest						

4.3. Evaluation of the Retrieved SM

Assessing the accuracy of SM retrieved from the CyGNSS satellite and understanding its spatial distribution of uncertainty is crucial for the development of retrieval algorithms. The CyGNSS data from 2020 were split into two random subsets to mitigate the impact of seasonal factors on parameter estimation. One subset was employed for training purposes,

whereas the remaining subset was utilized for testing, and both the training and testing sets consist of approximately 12.82 million samples. This indicates that the training and testing datasets are independent in time. They are also spatially independent, owing to the pseudo-random distribution of CyGNSS SP. A good agreement between the retrieved SM and SMAP SM was obtained by examining the training and test data. The density diagram comparing the retrieved and SMAP SM is shown in Figure 11. For the training data (half of 2020 data), the correlation coefficient R reaches 0.93, the RMSE is $0.058 \text{ cm}^3 \text{ cm}^{-3}$, the ubRMSE is $0.042 \text{ cm}^3 \text{ cm}^{-3}$, and the MAE is $0.040 \text{ cm}^3 \text{ cm}^{-3}$. For the test (remaining data in 2020), the correlation coefficient is 0.91, the RMSE is $0.067 \text{ cm}^3 \text{ cm}^{-3}$, the ubRMSE is $0.051 \text{ cm}^3 \text{ cm}^{-3}$, and the MAE is $0.044 \text{ cm}^3 \text{ cm}^{-3}$. The generalizability of this strategy was demonstrated by very small accuracy loss between the test dataset and the training dataset.

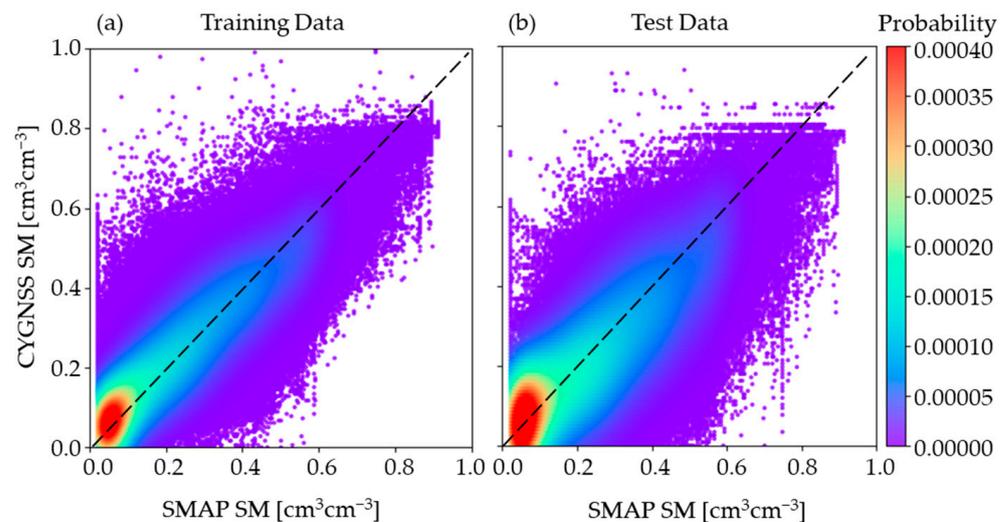


Figure 11. Density plot comparing the CyGNSS SM and SMAP SM. (a) Training and (b) test data in 2020. The black dotted line is a 1:1 reference line.

We assessed the correlation and error between SM derived from CyGNSS and SMAP to further validate the capability of CyGNSS to retrieve SM in each EASE-Grid cell. Figure 12 displays the statistical maps that were computed using the daily CyGNSS SM and SMAP SM, (a) correlation, (b) RMSE, (c) ubRMSE, and (d) MAE. The majority of pantropical areas have a good correlation (>0.8), low RMSE ($<0.06 \text{ cm}^3 \text{ cm}^{-3}$), ubRMSE ($<0.06 \text{ cm}^3 \text{ cm}^{-3}$), and MAE ($<0.06 \text{ cm}^3 \text{ cm}^{-3}$) overall. In contrast, the results in the Amazon, the Congo, and Southeast Asia are inconsistent, with low correlation (<0.6) and significant error ($>0.08 \text{ cm}^3 \text{ cm}^{-3}$). The SMAP SM data in these regions are always masked as not recommended for retrieval, since these areas are covered by forests with a high VWC of 15 kg/m^2 . These reference SMAP data were kept for the investigation, and the low-quality reference data may be the reason for the inconsistency [20].

In addition, retrieval errors for different land types were statistically analyzed, as listed in Table 3. It can be noted that mean values of VWC, SM, and retrieval errors are roughly positively correlated and that the RMSE, ubRMSE, and MAE of the retrieved results increase with the increase in SM and VWC. The error in densely forested areas, especially evergreen broadleaf forests, is usually greater than in areas with sparse vegetation, designated as Barren or Sparsely Vegetated or Shrublands. Furthermore, the correlation and errors rely on the land types when comparing Figures 10 and 12 visually for resemblance. Broadleaf and mixed forests are primary locations with weak correlation and significant errors in CyGNSS daily SM retrievals.

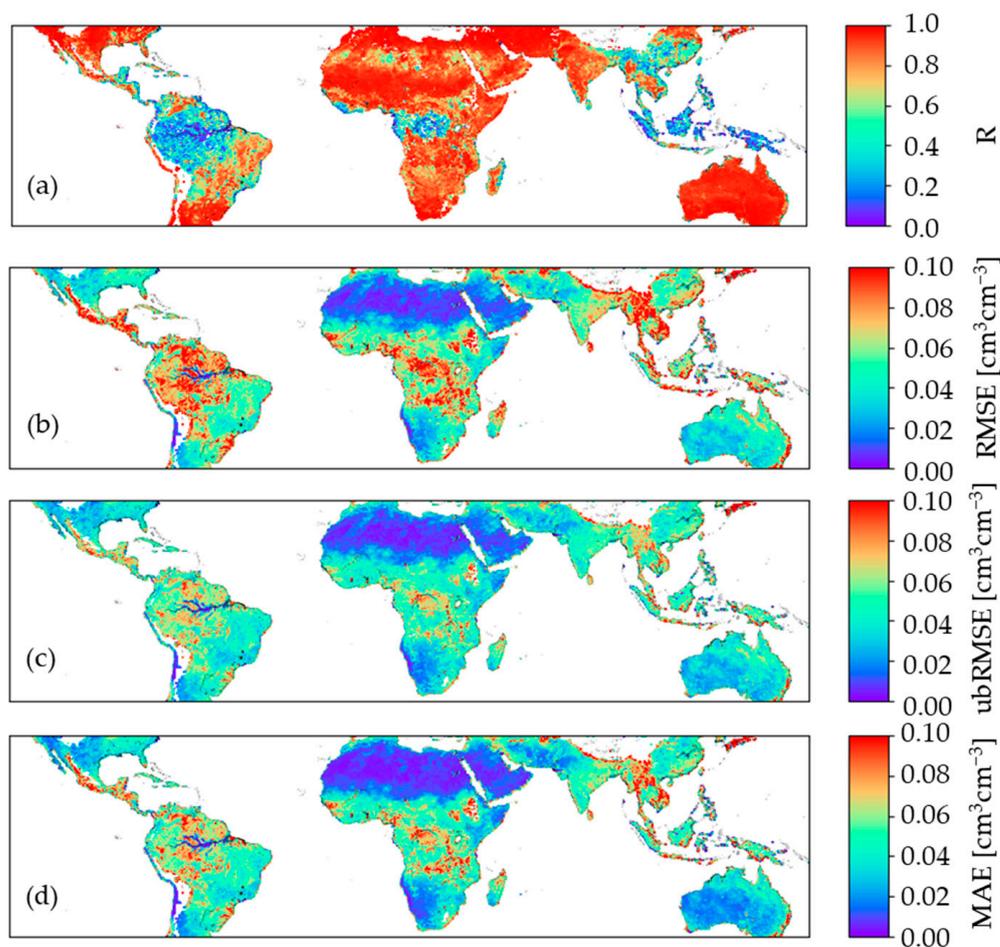


Figure 12. Statistical maps for the CyGNSS daily SM against the SMAP daily SM in 2020. (a) Correlation, (b) RMSE, (c) ubRMSE, and (d) MAE of each EASE-Grid.

4.4. Validation Using the In Situ Measurements

Due to the errors of CyGNSS and SMAP SM being expected to be correlated, the proposed method was evaluated with 4 in situ SM stations from the SMAP core validation/calibration sites. Figure 3 represents the sensor distribution. The EASE-Grid information and the geographical position of the ground sites are used to match the SM time-series data of the in situ sites, CyGNSS, and SMAP. Figure 13 shows the SM time series of Fort Cobb (USA), Little Washita (USA), Monte Buey (Argentina), and Yanco (Australia) from top to bottom. Statistical results have been listed in the title of Figure 13. It can be observed that the variations in the three datasets exhibit good consistency (ubRMSE is less than $0.35 \text{ cm}^3 \text{ cm}^{-3}$). It is important to note that under soil freezing conditions (the soil temperature is below 4 degrees Celsius), in situ SM records are not taken into account. Furthermore, we have included rainfall data from Global Precipitation Measurement (GPM, the spatial resolution is $0.1^\circ \times 0.1^\circ$) mission as a reference. Although the differences in the resolution of rainfall data and retrieved SM may lead to the inability to capture the spatial variability of rainfall in the study, overall, it shows a sharp increase in SM after rainfall events. The difference between CyGNSS SM and ground observations seems to be seasonally dependent. In the rainy season, the differences are relatively significant, whereas the disparities are minimal during the period of drought, which may be attributed to changes in surface conditions (e.g., vegetation).

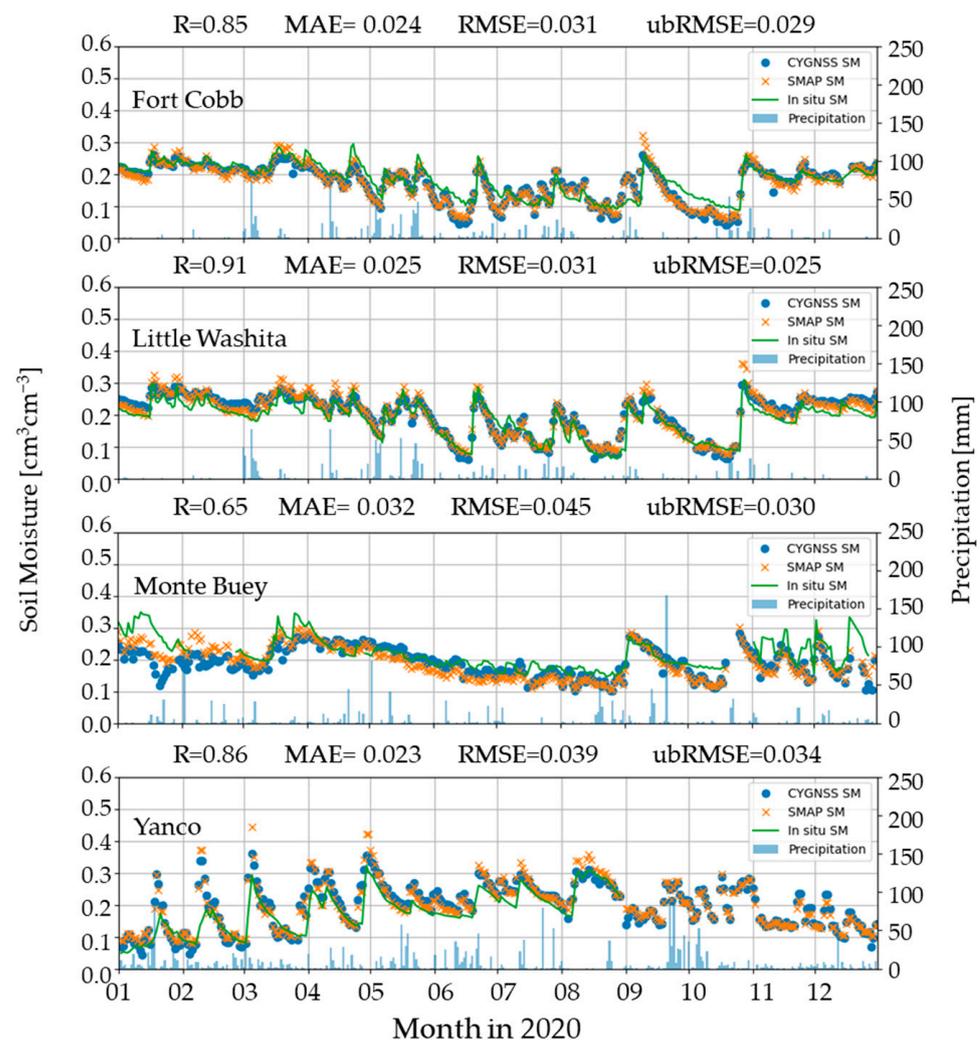


Figure 13. The SM time series derived from CyGNSS (blue dot), SMAP (orange cross), and in situ (green line); precipitation (mm) is displayed with a blue bar chart.

4.5. Comparison of UCAR SM Products

Here, the proposed robust regression method is compared with UCAR SM Products. It has been selected because it is a classic method of using linear regression to retrieve SM. Figure 14 shows the SMs from (a) CyGNSS, (b) UCAR, and (c) SMAP on 1 January 2020, and the error of (d) CyGNSS and (e) UCAR obtained concerning SMAP. Similar to previous discussions, significant errors mainly focus on forest-covered areas. We calculated the correlation and error between UCAR and SMAP SM: $R = 0.88$, $RMSE = 0.074 \text{ cm}^3\text{cm}^{-3}$, $ubRMSE = 0.053 \text{ cm}^3\text{cm}^{-3}$, $MAE = 0.052 \text{ cm}^3\text{cm}^{-3}$. Figure 15 shows the density map for the error of CyGNSS SM and UCAR SM compared with SMAP for the entire year of 2020, with a standard deviation of $0.064 \text{ cm}^3\text{cm}^{-3}$ for retrievals and a slightly larger standard deviation of $0.072 \text{ cm}^3\text{cm}^{-3}$ for UCAR products. This indicates that the retrieval results obtained using the present method are superior to those obtained using classical linear regression.

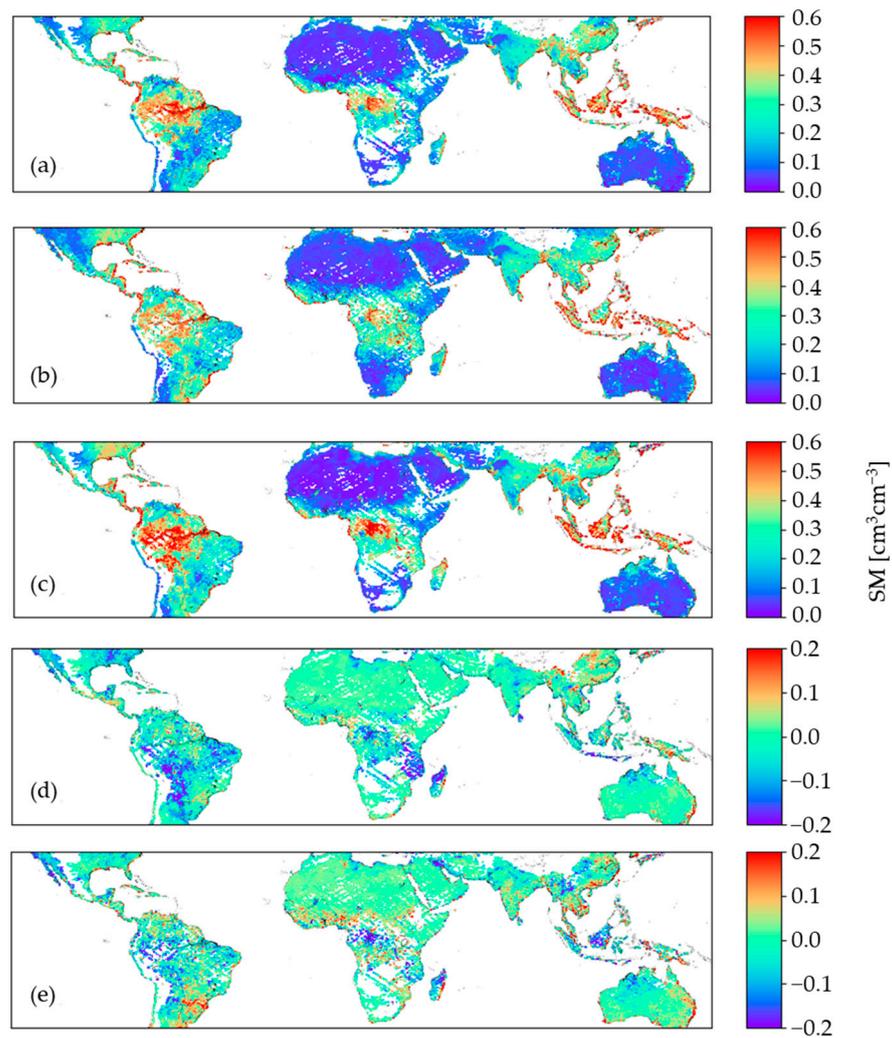


Figure 14. Comparison of SM values on 1 January 2020. (a) CyGNSS SM; (b) UCAR SM; (c) SMAP SM; (d) CyGNSS SM minus SMAP SM; (e) UCAR SM minus SMAP SM.

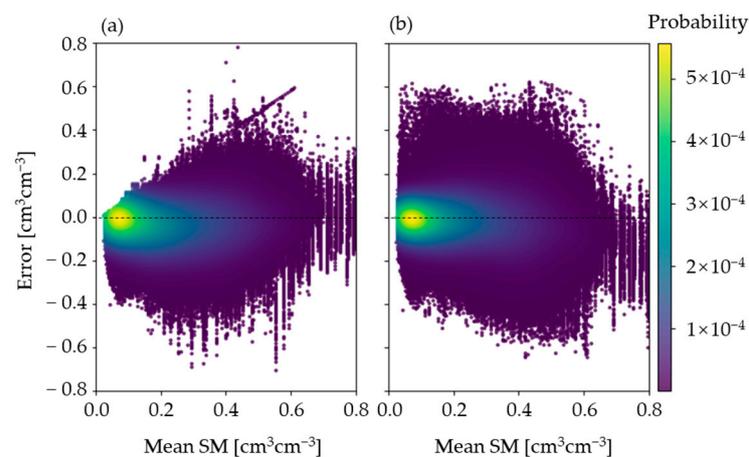


Figure 15. Density plot of SM error: (a) the error between CyGNSS SM and SMAP SM; (b) the error between UCAR SM and SMAP SM.

5. Discussions

Regarding the impact of observation noise on classical linear regression, a new method based on robust regression for retrieving SM from CyGNSS data is developed, and the

impact of roughness and vegetation is reconsidered. Similar to most remote sensing technologies, the proposed retrieval algorithm has limitations: (1) It may be incorrect to assume that the sensitivity of reflectivity to SM does not change over time in certain areas, such as agricultural areas, although this possibility is currently ignored. (2) SMAP SM errors may propagate to CyGNSS, especially for SMAP data that are not recommended for retrieval. Using other methods or models may reduce the dependence on SMAP SM information.

The retrieval capacity of CyGNSS is limited in densely vegetated areas. First, the density of canopies governs the feasibility of GNSS signals reaching the ground and being received by CyGNSS, making a precise characterization of vegetation coverage and its impact on microwave signals indispensable for SM retrieval. Second, like most existing methods, retrieving SM using reflectometers is performed under the assumption of coherent reflection. However, the incoherent scattering is the main contribution to the reflected signal over the tropical forest. Several recent studies [18,37–39] on the use of reflectometers to estimate above-ground biomass (AGB) demonstrated that the sensitivity of reflectivity to forest canopies improves when the incoherent scattering term is dominant on the reflected signals and different tropical forests have different AGB saturation. Last, in these regions, the influence of seasonal floods on CyGNSS reflectivity may overwhelm the SM signals, resulting in a decline in retrieval capability [21].

The results using ground observations for validation are site-specific and depend on sensors' locations, land cover, climate, and freeze-thaw conditions. Through the utilization of ground observations, CyGNSS showcases its ability to capture SM trends. It exhibits good performance during dry periods but shows a slight decrease during periods of increased rainfall.

As droughts and floods occur more frequently, the demand for water information is growing. In the future, GNSS-R SM products with high temporal resolution could be utilized as inputs for continental-scale hydrological models, such as the GEOGloWS ECMWF streamflow service [40] and the National Oceanic and Atmospheric Administration (NOAA) Water Model [41], enhancing streamflow prediction through SM data assimilation. Additionally, such products can play a crucial role in regions with often ineffective or limited hydrological services.

6. Conclusions

In this research, we present a robust regression-based approach for SM retrieval from CyGNSS reflectivity data, incorporating calibration for vegetation and roughness impacts on reflectivity. The proposed method's efficacy is demonstrated by SMAP, in situ measurements, and UCAR SM products, and satisfactory consistency is attained.

Spatially, CyGNSS retrieval yields satisfactory results in most of pantropical regions. However, the accuracy of CyGNSS is decreased in densely vegetated regions. This feature may belong to the more complex structure of the canopy layer and surface characteristics. We should be careful when using SM data over dense forests. In terms of time-series analysis, the consistency between CyGNSS retrieval and ground observations has reached a reasonable level. In the future, investigating and quantifying the impact of additional vegetation parameters on SM retrieval may be of interest.

Author Contributions: Conceptualization, Q.L. and S.Z.; methodology, Q.L. and W.L.; software, Q.L.; validation, Q.L., S.Z. and Y.N.; formal analysis, Z.M.; resources, S.Z. and J.P.; data curation, X.Z.; writing—original draft preparation, Q.L.; writing—review and editing, S.Z.; visualization, Q.L.; supervision, S.Z.; project administration, S.Z.; funding acquisition, S.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China Projects (Grant No. 42074041; 42127802); The National Key Research and Development Program of China (Grant No. 2020YFC1512000; 2019YFC1509802); State Key Laboratory of Geo-Information Engineering (Grant No. SKLGIE2022-ZZ2-07), Shaanxi Natural Science Research Program (Grant No. 2020JM-227). This research was also supported in part by the Fundamental Research Funds for the Central Universities, Chang'an University, (Grant No. 300102260301, 300102262401), by the Shaanxi Province

Science and Technology Innovation Team (Grant No. 2021 TD-51), and by the European Space Agency through the ESA-MOST DRAGON-5 Project (Grant No. 59339). This work of Weiqiang Li is partially supported by Grant RYC2019-027000-I funded by MCIN/AEI/<http://dx.doi.org/10.13039/501100011033> and by “European Union Next Generation EU/PRTR,” as is also supported by Grant 20215AT007 funded by Spanish National Research Council.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to NASA for providing CyGNSS and SMAP data. The authors are also grateful for the in situ data from soil moisture sensors from the SMAP/in situ core validation site. The authors thank all anonymous reviewers and editors for their constructive review of this manuscript.

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

References

1. Bennett, A.C.; Penman, T.D.; Arndt, S.K.; Roxburgh, S.H.; Bennett, L.T. Climate more important than soils for predicting forest biomass at the continental scale. *Ecography* **2020**, *43*, 1692–1705. [[CrossRef](#)]
2. Bojinski, S.; Verstraete, M.; Peterson, T.C.; Richter, C.; Simmons, A.; Zemp, M. The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 1431–1443. [[CrossRef](#)]
3. Entekhabi, D.; Das, N.; Njoku, E.; Johnson, J.; Shi, J. *SMAP L3 Radar/Radiometer Global Daily 9 km EASE-Grid Soil Moisture, Version 3*; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2016. [[CrossRef](#)]
4. Paul, A.R.; Raj, K. NASA-ISRO SAR (NISAR) Mission Status. In Proceedings of the 2021 IEEE Radar Conference (RadarConf21), Atlanta, GA, USA, 7–14 May 2021; pp. 1–6. [[CrossRef](#)]
5. Ruf, C.S.; Atlas, R.; Chang, P.S.; Clarizia, M.P.; Garrison, J.L.; Gleason, S.; Katzberg, S.J.; Jelenak, Z.; Johnson, J.T.; Majumdar, S.J.; et al. New Ocean Winds Satellite Mission to Probe Hurricanes and Tropical Convection. *Bull. Amer. Meteor. Soc.* **2016**, *97*, 385–395. [[CrossRef](#)]
6. Camps, A.; Park, H.; Pablos, M.; Foti, G.; Gommenginger, C.P.; Liu, P.; Judge, J. Sensitivity of GNSS-R Spaceborne Observations to Soil Moisture and Vegetation. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2016**, *9*, 4730–4742. [[CrossRef](#)]
7. Chew, C.; Shah, R.; Zuffada, C.; Hajj, G.; Masters, D.; Mannucci, A.J. Demonstrating soil moisture remote sensing with observations from the UK TechDemoSat-1 satellite mission. *Geophys. Res. Lett.* **2016**, *43*, 3317–3324. [[CrossRef](#)]
8. Clarizia, M.P.; Pierdicca, N.; Costantini, F.; Floury, N. Analysis of CyGNSS data for soil moisture retrieval. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2019**, *12*, 2227–2235. [[CrossRef](#)]
9. Zhang, S.; Ma, Z.; Li, Z.; Zhang, P.; Liu, Q.; Nan, Y.; Zhang, J.; Hu, S.; Feng, Y.; Zhao, H. Using CyGNSS Data to Map Flood Inundation during the 2021 Extreme Precipitation in Henan Province, China. *Remote Sens.* **2021**, *13*, 5181. [[CrossRef](#)]
10. Chew, C.; Reager, J.T.; Small, E. CyGNSS data map flood inundation during the 2017 Atlantic hurricane season. *Sci. Rep.* **2018**, *8*, 9336. [[CrossRef](#)]
11. Chew, C.; Small, E. Estimating inundation extent using CyGNSS data: A conceptual modeling study. *Remote Sens. Environ.* **2020**, *246*, 111869. [[CrossRef](#)]
12. Molina, I.; Calabia, A.; Jin, S.; Edokossi, K.; Wu, X. Calibration and Validation of CyGNSS Reflectivity through Wetlands’ and Deserts’ Dielectric Permittivity. *Remote Sens.* **2022**, *14*, 3262. [[CrossRef](#)]
13. Jia, Y.; Savi, P. Sensing soil moisture and vegetation using GNSS-R polarimetric measurement. *Adv. Space Res.* **2017**, *59*, 858–869. [[CrossRef](#)]
14. Wu, X.; Guo, P.; Sun, Y.; Liang, H.; Zhang, X.; Bai, W. Recent Progress on Vegetation Remote Sensing Using Spaceborne GNSS-Reflectometry. *Remote Sens.* **2021**, *13*, 4244. [[CrossRef](#)]
15. Lei, F.; Senyurek, V.Y.; Kurum, M.; Gurbuz, A.C.; Boyd, D.R.; Moorhead, R.J.; Crow, W.; Eroglu, O. Quasi-global machine learning-based soil moisture estimates at high spatio-temporal scales using CyGNSS and SMAP observations. *Remote Sens. Environ.* **2022**, *276*, 113041. [[CrossRef](#)]
16. Senyurek, V.; Lei, F.; Boyd, D.; Kurum, M.; Gurbuz, A.C.; Moorhead, R. Machine Learning-Based CyGNSS Soil Moisture Estimates over ISMN sites in CONUS. *Remote Sens.* **2020**, *12*, 1168. [[CrossRef](#)]
17. Eroglu, O.; Kurum, M.; Boyd, D.; Gurbuz, A.C. High Spatio-Temporal Resolution CyGNSS Soil Moisture Estimates Using Artificial Neural Networks. *Remote Sens.* **2019**, *11*, 2272. [[CrossRef](#)]
18. Santi, E.; Clarizia, M.P.; Comite, D.; Dente, L.; Guerriero, L.; Pierdicca, N.; Floury, N. Combining CyGNSS and Machine Learning for Soil Moisture and Forest Biomass Retrieval in View of the ESA Scout Hydrognss Mission. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 7433–7436. [[CrossRef](#)]
19. Yan, J.; Jin, S.; Chen, H.; Yan, Q.; Patrizia, S.; Jin, Y.; Yuan, Y. Temporal-Spatial Soil Moisture Estimation from CyGNSS Using Machine Learning Regression with a Preclassification Approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4879–4893. [[CrossRef](#)]
20. Yan, Q.; Huang, W.; Jin, S.; Jia, Y. Pan-tropical soil moisture mapping based on a three-layer model from CyGNSS GNSS-R data. *Remote Sens. Environ.* **2020**, *247*, 111944. [[CrossRef](#)]

21. Chew, C.; Small, E. Description of the UCAR/CU Soil Moisture Product. *Remote Sens.* **2020**, *12*, 1558. [[CrossRef](#)]
22. Gleason, S.; Ruf, C.S.; O'Brien, A.J.; McKague, D.S. The CYGNSS Level 1 Calibration Algorithm and Error Analysis Based on On-Orbit Measurements. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *12*, 37–49. [[CrossRef](#)]
23. Ren, B.; Zhu, J.; Tsang, L.; Xu, H. Analytical Kirchhoff Solutions (AKS) and Numerical Kirchhoff Approach (NKA) for First-Principle Calculations of Coherent Waves and Incoherent Waves at P Band and L Band in Signals of Opportunity (SoOp). *Prog. Electromagn. Res.* **2021**, *171*, 35–73. [[CrossRef](#)]
24. Yueh, S.; Shah, R.; Chaubell, M.J.; Hayashi, A.; Xu, X.; Colliander, A. A semi-empirical modeling of soil moisture, vegetation, and surface roughness impact on CyGNSS reflectometry data. *IEEE Trans. Geosci. Remote Sens.* **2020**, *60*, 5800117. [[CrossRef](#)]
25. Al-Khaldi, M.M.; Johnson, J.T.; O'Brien, A.J.; Balenzano, A.; Mattia, F. Time-Series Retrieval of Soil Moisture Using CyGNSS. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4322–4331. [[CrossRef](#)]
26. University of Michigan. *CyGNSS Handbook*; Michigan Publishing: Ann Arbor, MI, USA, 2016; ISBN 978-1-60785-380-0. Available online: https://cygnss.engin.umich.edu/wp-content/uploads/sites/534/2021/06/CyGNSS_Handbook_April2016.pdf (accessed on 2 June 2023).
27. CyGNSS. CyGNSS Level 1 Science Data Record Version 3.1. Ver. 3.1. PO.DAAC, CA, USA, 2021. Available online: https://podaac.jpl.nasa.gov/dataset/CYGNSS_L1_V3.1 (accessed on 2 June 2023).
28. Colliander, A.; Asanuma, J.; Berg, A.; Bongiovanni, T.; Bosch, D.; Caldwell, T.; Holifield-Collins, C.; Jensen, K.; Livingston, S.; Lopez-Baeza, E.; et al. *SMAP/In Situ Core Validation Site Land Surface Parameters Match-Up Data, Version 1*; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2020. [[CrossRef](#)]
29. AlJassar, H.; Temimi, M.; Abdelkader, M.; Petrov, P.; Kokkalis, P.; AlSarraf, H.; Roshni, N.; Hendi, H.A. Validation of NASA SMAP Satellite Soil Moisture Products over the Desert of Kuwait. *Remote Sens.* **2022**, *14*, 3328. [[CrossRef](#)]
30. Colliander, A.; Cosh, M.H.; Misra, S.; Bourgeau-Chavez, L.; Kelly, V.; Siqueira, P.; Roy, A.; Lakhankar, T.; Kraatz, S.; Konings, A.; et al. SMAP Validation Experiment 2019–2022 (SMAPVEX19-22): Detection of soil moisture under temperate forest canopy. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium, Brussels, Belgium, 11–16 July 2021. [[CrossRef](#)]
31. Abdelkader, M.; Temimi, M.; Colliander, A.; Cosh, M.H.; Kelly, V.R.; Lakhankar, T.; Fares, A. Assessing the Spatiotemporal Variability of SMAP Soil Moisture Accuracy in a Deciduous Forest Region. *Remote Sens.* **2022**, *14*, 3329. [[CrossRef](#)]
32. Das, N.; Entekhabi, D.; Dunbar, R.S.; Kim, S.; Yueh, S.; Colliander, A.; O'Neill, P.E.; Jackson, T.; Jagdhuber, T.; Chen, F.; et al. *SMAP/Sentinel-1 L2 Radiometer/Radar 30-Second Scene 3 km EASE-Grid Soil Moisture, Version 3*; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2020; Available online: https://nsidc.org/sites/default/files/spl2_smap_s-v003-userguide_0.pdf (accessed on 2 June 2023).
33. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
34. Zhang, X.; Zhang, T.; Zhou, P.; Shao, Y.; Gao, S. Validation Analysis of SMAP and AMSR2 Soil Moisture Products over the United States Using Ground-Based Measurements. *Remote Sens.* **2017**, *9*, 104. [[CrossRef](#)]
35. Walker, V.A.; Hornbuckle, B.K.; Cosh, M.H.; Prueger, J.H. Seasonal Evaluation of SMAP Soil Moisture in the U.S. Corn Belt. *Remote Sens.* **2019**, *11*, 2488. [[CrossRef](#)]
36. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
37. Carreno-Luengo, H.; Luzi, G.; Crosetto, M. Above-Ground Biomass Retrieval over Tropical Forests: A Novel GNSS-R Approach with CyGNSS. *Remote Sens.* **2020**, *12*, 1368. [[CrossRef](#)]
38. Pettinato, S.; Paloscia, S.; Clarizia, M.P.; Dente, L.; Guerriero, L.; Guerriero, L.; Pierdicca, N. Soil Moisture and Forest Biomass retrieval on a global scale by using CyGNSS data and Artificial Neural Networks. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 5905–5908. [[CrossRef](#)]
39. Zribi, M.; Dehaye, V.; Dassas, K.; Fanise, P.; Page, M.L.; Laluet, P.; Boone, A. Airborne GNSS-R Polarimetric Multiincidence Data Analysis for Surface Soil Moisture Estimation over an Agricultural Site. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8432–8441. [[CrossRef](#)]
40. Sanchez Lozano, J.; Romero Bustamante, G.; Hales, R.C.; Nelson, E.J.; Williams, G.P.; Ames, D.P.; Jones, N.L. A Streamflow Bias Correction and Performance Evaluation Web Application for GEOGloWS ECMWF Streamflow Services. *Hydrology* **2021**, *8*, 71. [[CrossRef](#)]
41. National Oceanic and Atmospheric Administration (NOAA). NOAA Launches America's First National Water Forecast Model | National Oceanic and Atmospheric Administration. Available online: <https://www.noaa.gov/media-release/noaa-launches-america-s-first-national-water-forecast-model> (accessed on 15 July 2023).

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