remote sensing

MDPI

Article

A New Joint Retrieval of Soil Moisture and Vegetation Optical Depth from Spaceborne GNSS-R Observations

Mina Rahmani [1], Jamal Asgari [1] and Alireza Amiri-Simkooei [2,*]

[1] Department of Geomatics Engineering, Faculty of Civil Engineering and Transportation, University of Isfahan, Isfahan 81746-73441, Iran; mi.rahmani@trn.ui.ac.ir (M.R.); asgari@eng.ui.ac.ir (J.A.)
[2] Department of Control and Operations, Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands
* Correspondence: a.amirisimkooei@tudelft.nl

Highlights

What are the main findings?

- The trained ANN algorithm simultaneously retrieves soil moisture and vegetation optical depth from CYGNSS observations, showing strong agreement with reference satellite products (SMAP SM: R = 0.83, RMSE = 0.063 m^3/m^3; SMOS VOD: R = 0.89, RMSE = 0.088).
- ANN-derived VOD shows strong correlation with independent vegetation indicators—biomass (R~0.77), canopy height (R~0.95), Leaf Area Index (R = 0.96), and vegetation water content (R~0.90)—confirming reliable sensitivity to vegetation structure.

What are the implications of the main findings?

- The combination of GNSS-R data with environmental variables enables reliable dual retrieval of soil moisture and vegetation optical depth, serving as a cost-effective, higher-resolution alternative/complement to SMAP and SMOS.
- Joint retrieval of SM and VOD enables improved characterization of land–atmosphere interactions, supporting hydrological, ecological, and climate applications.

Abstract

Accurate estimation of soil moisture (SM) and vegetation optical depth (VOD) is essential for understanding land–atmosphere interactions, climate dynamics, and ecosystem processes. While passive microwave missions such as SMAP and SMOS provide reliable global SM and VOD products, they are limited by coarse spatial resolution and infrequent revisit times. Global Navigation Satellite System Reflectometry (GNSS-R) observations, particularly from the Cyclone GNSS (CYGNSS) mission, offer an improved spatiotemporal sampling rate. This study presents a deep learning framework based on an artificial neural network (ANN) for the simultaneous retrieval of SM and VOD from CYGNSS observations across the contiguous United States (CONUS). Ancillary input features, including specular point latitude and longitude (for spatial context), CYGNSS reflectivity and incidence angle (for surface signal characterization), total precipitation and soil temperature (for hydrological context), and soil clay content and surface roughness (for soil properties), are used to improve the estimates. Results demonstrate strong agreement between the predicted and reference values (SMAP SM and SMOS VOD), achieving correlation coefficients of R = 0.83 and 0.89 and RMSE values of 0.063 m^3/m^3 and 0.088 for SM and VOD, respectively. Temporal analyses show that the ANN accurately reproduces both seasonal and daily variations in SMAP SM and SMOS VOD (R ≈ 0.89). Moreover, the predicted SM and VOD maps show strong agreement with the reference SM and VOD maps (R ≈ 0.93).

Academic Editors: Frédéric Frappart, Hui Lu, Xiaojun Li and Lun Gao

Received: 2 December 2025
Revised: 15 January 2026
Accepted: 16 January 2026
Published: 20 January 2026

Additionally, ANN-derived VOD demonstrates strong consistency with above-ground biomass ($R \approx 0.77$), canopy height ($R \approx 0.95$), leaf area index ($R = 96$), and vegetation water content ($R \approx 0.90$). These results demonstrate the generalizability of the approach and its applicability to broader environmental sensing tasks.

Keywords: soil moisture; vegetation optical depth (VOD); GNSS reflectometry (GNSS-R); CYGNSS; artificial neural network (ANN)

1. Introduction

Accurate estimation of soil moisture (SM) and vegetation optical depth (VOD) is critical for understanding and monitoring climate change. SM, one of the essential climate variables (ECVs) according to the Global Climate Observing System (GCOS) [1], regulates the water cycle and weather patterns, while VOD, reflecting the water content in vegetation, serves as an indicator of vegetation health and carbon storage [2]. Changes in these variables reveal ecosystem responses to global warming, such as shifts in droughts, floods, and carbon dynamics. Passive microwave sensors include the Soil Moisture Active Passive (SMAP) mission [3] and the Soil Moisture Ocean Salinity (SMOS) mission [4], two well-known remote sensing missions that provide global SM and VOD measurements with a spatial resolution of approximately 40 km every 2–3 days.

Recently, Global Navigation Satellite System Reflectometry (GNSS-R) has been increasingly used for hydrological applications, including the monitoring of SM [5], VOD [6], and ice-covered lakes [7], as well as flood detection [8], due to its distinct advantages. Compared to traditional remote sensing missions, GNSS-R offers higher spatial resolution (down to 1–25 km) and more frequent temporal sampling (sub-daily). Unlike large and high-budget missions like SMAP and SMOS, GNSS-R relies on free GNSS signals reflected from the Earth's surface, measured by small, lightweight, and low-cost satellites [9]. These features make GNSS-R a promising alternative for hydrological applications including vegetation and soil moisture monitoring.

However, GNSS-R observations are influenced by multiple interacting factors, including soil moisture, vegetation attenuation, surface roughness, observation geometry and meteorological conditions [10–12]. The relationships among these factors are highly nonlinear and difficult to represent using physical or semi-empirical models alone. To address these challenges, deep learning methods are particularly well suited. Previous studies have demonstrated the effectiveness of supervised regression algorithms for hydrological applications [13]. For example, ref. [14] evaluated the performance of three Machine Learning (ML) models—namely, random forest (RF), artificial neural network (ANN), and support vector machine (SVM)—for soil moisture monitoring using NASA's GNSS-R mission, the Cyclone Global Navigation Satellite System (CYGNSS). CYGNSS is a constellation of eight microsatellites launched in late 2016 [9]. The RF, ANN, and SVM models achieved overall RMSE values of 0.052, 0.061, and 0.065 cm^3/cm^3, respectively, when compared to International Soil Moisture Network (ISMN) soil moisture data across the Contiguous United States (CONUS). Recently, ref. [15] developed a physics-informed deep learning framework (MSA-ViT) for retrieving soil moisture from CYGNSS observations and showed that it achieves strong performance, with correlations up to 0.75–0.85 and RMSE values of 0.03–0.05 m^3/m^3. While these studies demonstrate the strong potential of deep learning for GNSS-R-based SM estimation, they focus exclusively on SM and typically treat vegetation effects as external or prescribed inputs, thereby neglecting the coupled soil–vegetation signal interactions inherent in GNSS-R measurements.

Compared to SM retrieval, GNSS-R-based estimation of vegetation properties remains less explored. Ref. [6] developed a physics-based VOD product from GNSS-R observations and validated it at both global and regional scales using SMAP VOD. In this research, over three years of monthly average data, the CYGNSS-derived VOD shows good agreement with the SMAP-derived VOD. Recently, ref. [16] examined diurnal cycles of vegetation water content (VWC), which is linearly related to VOD, in the Amazon and their responses to water stress. Using high-frequency measurements from the CYGNSS constellation, the study identified notable diurnal variations in VWC, with higher levels in the morning than in the evening, reflecting water uptake and transpiration processes. Consequently, VWC derived from GNSS-R observations is suggested as a useful indicator of water stress, offering improved understanding of forest responses to increasing aridity under a changing climate. Ref. [17] developed a deep learning GNSS-R framework to retrieve vegetation water content (VWC), which is linearly related to VOD, by training on a triplet dataset from CYGNSS, GLDAS, and SMAP. The model achieves strong performance, with a minimum root mean square deviation of 1.10 kg/m^2 between the estimated VWC and the ground truth, demonstrating reliable large-scale VWC estimation. Despite these advances, existing GNSS-R vegetation studies generally focus on VOD or VWC as stand-alone products, are often limited in temporal resolution or spatial scope, and do not explicitly account for the interaction between vegetation and soil moisture signals.

Overall, most GNSS-R-based retrieval approaches estimate SM and VOD independently. In SM retrieval studies, vegetation effects are commonly represented using ancillary vegetation indices such as the normalized difference vegetation index (NDVI) or VWC, whereas SM is often prescribed or indirectly accounted for in VOD estimation. This separation limits the ability of existing models to capture the coupled soil–vegetation signal interactions present in GNSS-R observations and may lead to physically inconsistent estimates. Recently, ref. [18] proposed a stand-alone CYGNSS-based method for retrieving SM and VOD using two independently trained neural networks linked through an iterative search procedure. While this approach represents an important step toward joint SM–VOD retrieval, it relies exclusively on CYGNSS observations.

In contrast, the present study introduces a unified ANN-based GNSS-R framework that simultaneously retrieves SM and VOD by integrating CYGNSS observations with multiple ancillary datasets. This study provides the first daily time series of jointly estimated SM and VOD over the CONUS, as illustrated in Figure 1. Unlike previous GNSS-R studies that estimate SM and VOD separately or sequentially, the proposed framework jointly retrieves both variables, enabling the model to internally learn their coupled soil–vegetation signal interactions. Furthermore, vegetation effects are not prescribed through external vegetation indices; instead, the contribution of vegetation to GNSS-R observations is treated as an explicit signal and modeled through VOD jointly with SM.

ANNs are particularly well suited for this task because they can implicitly learn complex nonlinear relationships directly from data, without requiring explicit assumptions about scattering mechanisms or vegetation parameterization. Additionally, ANNs can handle high-dimensional inputs and integrate heterogeneous ancillary datasets, making them well suited for GNSS-R applications where multiple environmental variables jointly influence the reflected signal. Therefore, the proposed ANN-based dual-retrieval framework provides a flexible and scalable alternative to traditional retrieval methods and is particularly advantageous under complex surface and vegetation conditions.

The remainder of this paper is organized as follows: Section 2 describes the datasets utilized in the study and their corresponding sources. Section 3 elaborates on the methodology, including data preprocessing as well as the tuning of the ANN and its optimized hyperparameters. Section 4 discusses the results and validation of both soil moisture

and vegetation optical depth with respect to reference values (SMAP SM, in situ SM, and SMOS VOD). In Section 5 (Discussion), we investigate the sensitivity of the retrieved VOD to other vegetation indices—including the vegetation water content, canopy height, and above-ground biomass—as well as the contribution of each input feature to the model's performance. Section 6 presents the conclusions.

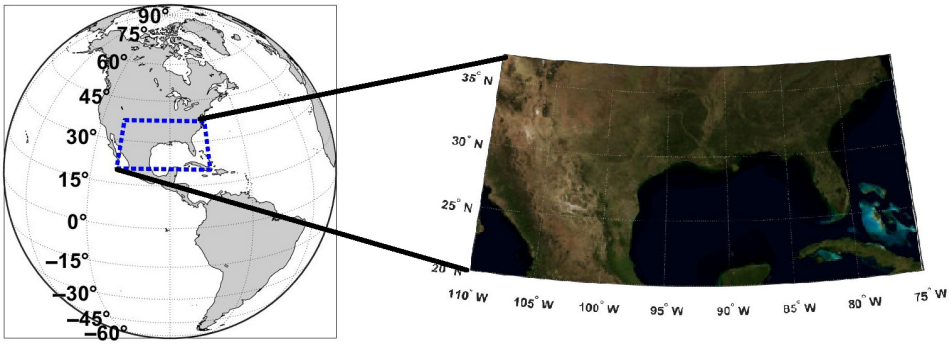

Figure 1. Geographic scope of the case study: CONUS. The left panel shows the global view with the study area highlighted, while the right panel presents a zoomed-in natural-color satellite image illustrating land, ocean, and vegetation.

2. Datasets

The data utilized in this study were gathered from six sources—SMAP, SMOS, CYGNSS, the Global Land Data Assimilation System (GLDAS), ERA5-Land, and [19]—between January 2020 and December 2021. Table 1 outlines the CYGNSS observations, ancillary datasets, and reference values, detailing their sources and characteristics as applied in this study. Soil surface temperature, soil clay content, soil surface roughness, and total precipitation were selected to model the influence of soil conditions—closely linked to soil moisture—on GNSS-R observations, similar to previous deep learning-based soil moisture retrieval efforts using GNSS-R data (e.g., [20,21]). SMAP (daily soil moisture at 9 km resolution) and SMOS vegetation optical depth are considered as the target output variables.

Table 1. Datasets used in this research and their characteristics.

Category	Variables	Source	Spatial Resolution	Temporal Sampling Rate	Unit	Reference
Ancillary datasets	Top 0–7 cm soil surface temperature	ERA5 [1]	0.1°	Hourly	Kelvin	[22]
	Total precipitation	ERA5	0.1°	Hourly	Meter	[22]
	Soil clay percentage	GLDAS [2]	0.25°	Yearly	Percent	[23]
	Soil surface roughness (Hr)	[19]	0.2°	Yearly	-	[19]
CYGNSS (Level 1)	CYGNSS DDM	PO.DAAC	0.5–25 km	Hourly to daily	Watt	[24]
	CYGNSS incidence angle	PO.DAAC [3]	0.5–25 km	Hourly to daily	Degree	[24]
	Specular point latitude	PO.DAAC	0.5–25 km	Hourly to daily	Degree	[24]
	Specular point longitude	PO.DAAC	0.5–25 km	Hourly to daily	Degree	[24]
	G_t, G_r	PO.DAAC	-	Hourly to daily	dBi	[24]
	R_{ts}, R_{sr}	PO.DAAC	-	Hourly to daily	Meter	[24]
	P_t	PO.DAAC	-	Hourly to daily	dBW	[24]
	Quality flags	PO.DAAC	-	Hourly to daily	-	[24]
Reference values	SMAP SM (SPL3SMPE)	SMAP [4]	9 km	1–2 days	m^3/m^3	[25]
	SMOS VOD (Level 2)	SMOS [5]	35–50 km	2–3 days	-	[26]

[1] https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land?tab=download (accessed on 1 November 2023); [2] https://ldas.gsfc.nasa.gov/ (accessed on 1 December 2023); [3] https://podaac.jpl.nasa.gov/dataset/CYGNSS_L1_FULL_DDM_V3.0 (accessed on 1 January 2024); [4] https://nsidc.org/data/spl3smp_e/versions/6 (accessed on 1 February 2024); [5] https://smos-diss.eo.esa.int/socat/SMOS_Open/search (accessed on 1 March 2024).

In our study, SMOS VOD was used instead of SMAP VOD. It is important to note that SMAP VOD is not a directly retrieved parameter; rather, it is derived from VWC, which, itself, is estimated using MODIS NDVI. In contrast, SMOS VOD is physically retrieved simultaneously with soil moisture from SMOS brightness temperature observations.

3. Methodology

This section presents the methodological framework used for the simultaneous retrieval of soil moisture and vegetation optical depth using an artificial neural network and CYGNSS observations. Section 3.1 details the data preprocessing steps, while Section 3.2 describes the training procedure and hyperparameter tuning of the ANN model. A flowchart summarizing the methodological steps followed in this study is shown in Figure 2.

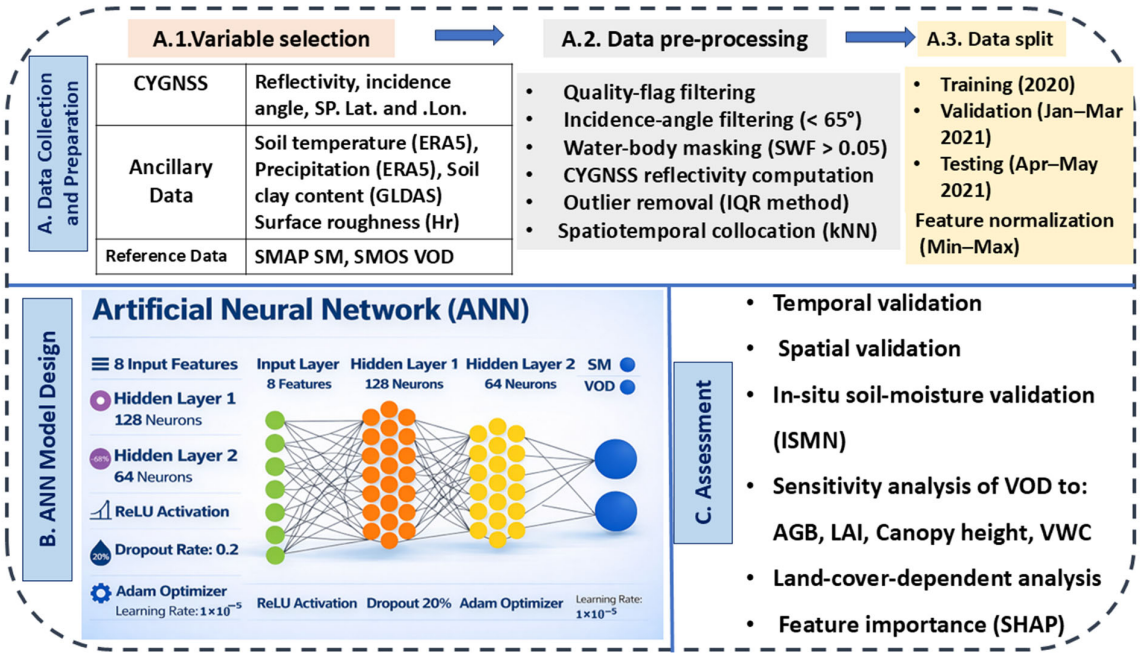

Figure 2. Overall workflow of the proposed ANN-based framework for soil moisture and vegetation optical depth retrieval from CYGNSS data. The main processing steps are highlighted in blue, while other colors are used to distinguish sub-steps within data collection and preprocessing, as well as the layers of the artificial neural network. The colors are used for illustrative purposes only and do not represent quantitative values.

3.1. Data Preprocessing

Before training the model, several preprocessing steps were applied to the data. First, following the methodology of [11,19], CYGNSS observations flagged for issues such as S-band activation, significant spacecraft attitude errors, blackbody DDM, DDM test patterns, or unreliable GPS EIRP estimates were removed. Observations with incidence angles above 65° were excluded because high angles can increase measurement noise, consistent with the methodology of [5]. Since forward-scattered signals over water surfaces have significantly higher power levels than those over soil due to strong coherence [27], regions with a static water fraction above 0.05, as provided by SMAP, were also excluded.

Next, CYGNSS reflectivity (Γ_{eff}) was calculated from Level 1 CYGNSS observations, following the methodology described in [19,28].

$$\Gamma_{eff}(\theta) = \frac{\tilde{P}_r^{S} (4\pi)^2 (R_{ts} + R_{sr})^2}{\lambda^2 \, P_t \;\; G_t \, G_r} \tag{1}$$

where P_t (dBW), $G_t(dBi)$, and G_r (dBi) represent the transmitted power, the gain of the GPS satellite antenna, and the gain of the CYGNSS antenna, respectively. The terms $R_{ts}(m)$ and $R_{sr}(m)$ denote the distances from the transmitter to the specular point (SP) and from the SP to the receiver. The specular point refers to the specific location on the Earth's surface where signals transmitted by GNSS satellites interact with the surface. The λ parameter represents the wavelength of the transmitted signals, which is 19.6 cm for the CYGNSS mission, as it exclusively detects GPS L-band signals [19].

The adjusted signal power is expressed as $\overset{\sim S}{\mathrm{P}}_r\,(\mathrm{watt}) = \mathrm{P_r} - \mathrm{P_N}$, where P_r denotes the intensity of the main signal within the CYGNSS Delay Doppler Map (DDM) and P_N represents the noise power calculated from the first two Doppler bins of the DDM. Generally, CYGNSS observations are presented as two-dimensional (2D) maps known as DDMs. These 2-D maps provide detailed physical information about the Earth's surface area surrounding the specular point [19].

Γ_{eff} values in decibels can be calculated based on Equation (1) as follows:

$$\Gamma_{eff} = 10log_{10}(\overset{\sim S}{P}_r) + 20log_{10}(4\pi) + 20log_{10}(R_{ts} + R_{sr}) - 20log_{10}(0.196) - P_t - G_t - G_r \quad (2)$$

The calculated CYGNSS reflectivity and incidence-angle values are considered representative of CYGNSS observations. Accordingly, the final set of input features includes the ancillary datasets listed in Table 1, along with the CYGNSS-derived reflectivity values, incidence angle, and latitude and longitude of the specular points. Based on the correlation matrix (refer to Figure 3), no significant relationships are observed among the input features, and the correlation values are generally low, indicating that multicollinearity is not a concern.

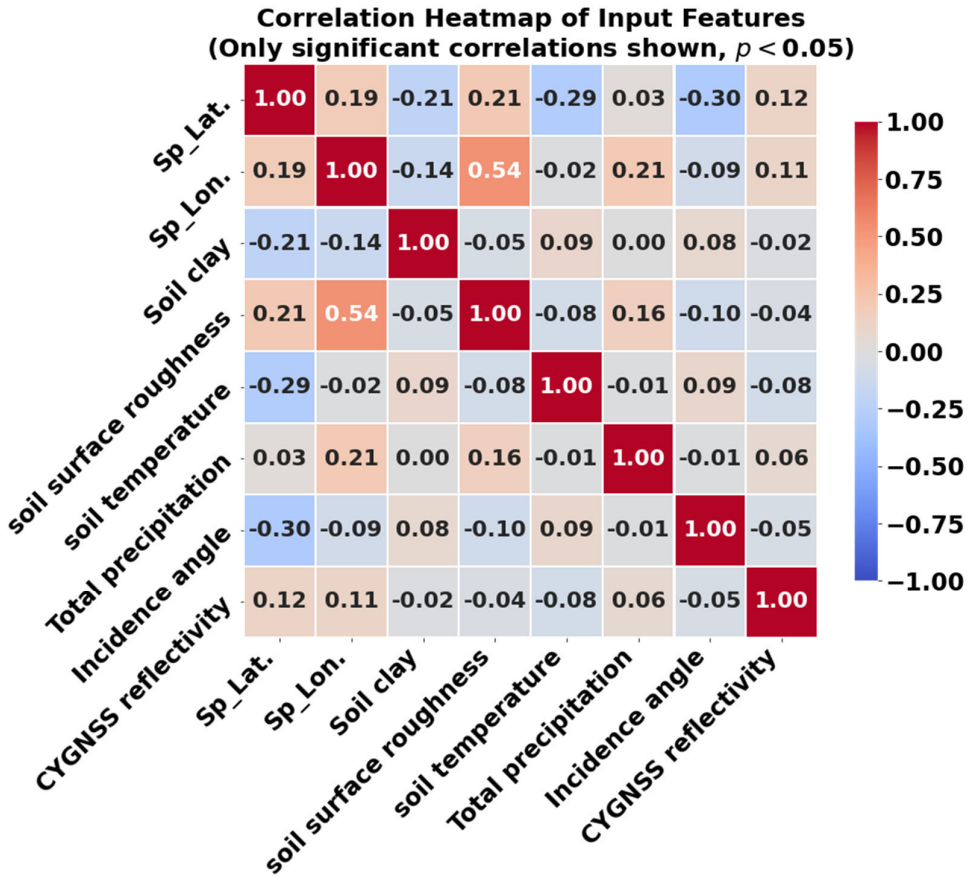

Figure 3. Interdependence of input features based on Pearson correlation.

Outliers—data points that deviate significantly from the majority due to errors or unusual measurements—can distort the dataset and negatively impact model performance. To ensure good generalization across all valid test inputs, we used Tukey's Interquartile Range (IQR) method [29] to detect and remove outliers in CYGNSS reflectivity, ancillary datasets, and reference values. Next, CYGNSS observations were collocated with ancillary and reference datasets using a spatiotemporal nearest-neighbor matching approach. For each CYGNSS observation on a given day, all available ancillary data acquired on the same day were first identified. Spatial matching was then performed by selecting the nearest ancillary pixel within a 0.2° radius using a k-nearest neighbor (kNN) search (k = 1), implemented with the knnsearch function in MATLAB R2021 a. No spatial interpolation or temporal averaging was applied, ensuring that each CYGNSS observation was paired with the closest available ancillary and reference values in space and time.

3.2. Tuning the ANN Model

We divided our dataset into three subsets: a training set (72%, covering all of 2020, with 5,387,341 observations), a validation set (16%, covering the first three months of 2021, with 1,248,147 observations), and a test set (12%, covering April and May of 2021, with 841,985 observations). This splitting ensures that the model is trained on a full annual cycle of data (2020), allowing it to learn seasonal variations effectively, which is crucial for SM and VOD prediction.

Many machine learning models achieve better performance when numerical features are normalized to a consistent range. Therefore, in this research, each dataset was normalized using the MinMaxScaler function implemented in the scikit-learn library (version 1.3.2) within a Jupyter Notebook environment. The MinMaxScaler is a normalization technique that rescales feature values to the range of [0, 1] [30]. Furthermore, the mean value of each dataset was employed to impute the missing data.

The ANN architecture (Table 2) was selected empirically through a systematic trial-and-error procedure, with a previously published GNSS-R–based ANN study [31] providing guidance for defining a reasonable range of hyperparameters. We first evaluated networks with a single hidden layer while gradually increasing the number of neurons and observed a progressive improvement in model performance up to an optimal size. Subsequently, a second hidden layer was introduced, which further improved the retrieval accuracy, particularly in capturing nonlinear relationships between inputs and outputs. Increasing the number of neurons in the two-layer configuration led to marginal gains until a stable optimum was reached. Additional experiments with a third hidden layer did not result in further performance improvement. Therefore, a two-hidden-layer architecture with the selected number of neurons was adopted as a trade-off between model complexity and performance.

Table 2. Optimal hyperparameters of the trained ANN model.

Hyperparameters	Value	Hyperparameters	Value
Number of hidden layers	2	Kernel initializer	normal
Number of neurons in the input layer	8	Batch size	128
Number of neurons in the output layer	2	Optimizer	Adam
Number of neurons in the first hidden layer	128	Learning rate	0.00001
Number of neurons in the second hidden layer	64	Activation function in the input and first and second hidden layers	ReLU
Dropout rate in the hidden layers	0.2		

Figure 4 shows the convergence of training and validation losses for soil moisture and vegetation optical depth predictions, indicating stable learning, with no signs of overfitting or underfitting.

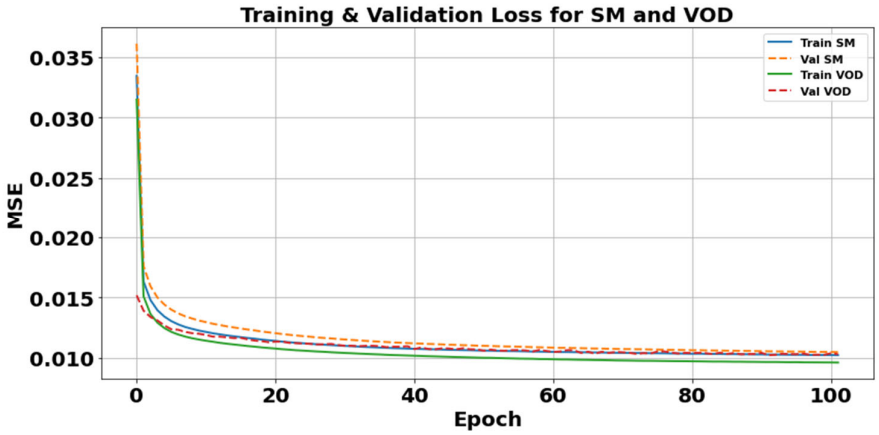

Figure 4. Learning curves of the ANN model for soil moisture and vegetation optical depth estimation.

The model's performance on unseen data and its generalization capability were assessed using the test dataset based on RMSE and the correlation coefficient. The strong correlation values (SM: R = 0.83; VOD: R = 0.89) and low RMSE scores (SM: 0.063 m^3/m^3; VOD: 0.088) demonstrate the effectiveness of the trained ANN model and its generality (see Figure 5). Additional statistical analysis, as presented in Table 3, further confirms the robustness of our model and its results.

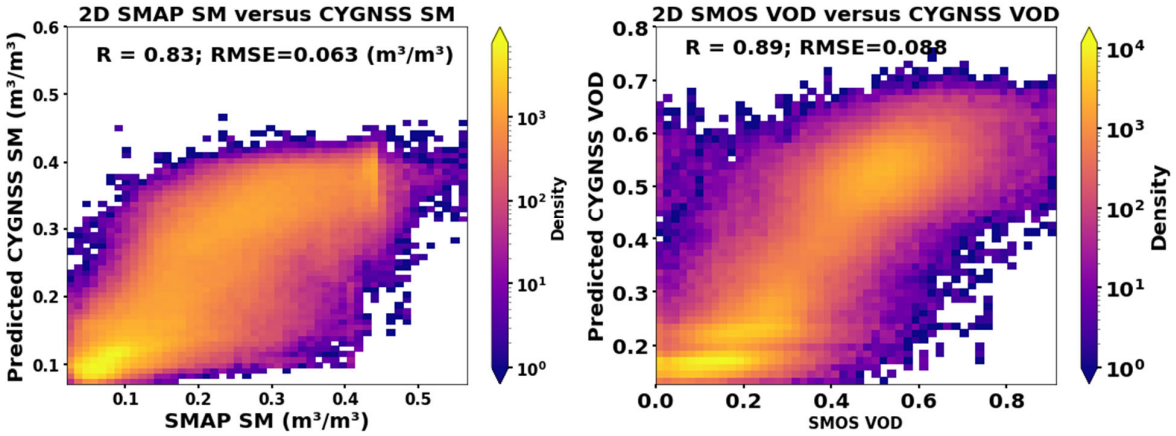

Figure 5. Predicted CYGNSS SM and VOD versus SMAP SM and SMOS VOD on the test dataset.

Table 3. ANN model performance for SM and VOD predictions.

	Mean Absolute Error (MAE)	R^2	Bias (MBE)
SM	0.0498 m^3/m^3	0.6588 (unitless)	0.0188 m^3/m^3
VOD	0.0678 (unitless)	0.7680 (unitless)	0.0195 (unitless)

4. Results

This section presents and validates the retrieved soil moisture and vegetation optical depth. First, Section 4.1 focuses on the temporal validation of the retrieved SM and VOD against SMAP SM and SMOS VOD. Second, Section 4.2 evaluates the spatial sensitivity of the retrieved SM and VOD to changes in SMAP SM and SMOS VOD. Third, Section 4.3

discusses the validation of the predicted SM against in situ soil moisture measurements from ground stations.

4.1. Temporal Validation of SM and VOD Retrievals

Figure 6 shows the monthly correlation coefficient (R) and RMSE between the ANN predictions and the reference datasets for 2021. For soil moisture (top panel), the model demonstrates strong agreement with SMAP SM (R~0.7–0.9) and low RMSE values ($\leq$0.04 m^3/m^3), showing the highest consistency during the validation period (Jan–Mar) and slightly reduced performance around mid-year and in October. For vegetation optical depth (bottom panel), the ANN predictions show good agreement with SMOS VOD throughout the year, with correlations generally between 0.7 and 0.9 and RMSE values $\leq$ 0.06. The performance is strongest during the validation period (Jan–Mar), becoming worse in July, when the vegetation density is highest. The reduced performance of soil moisture retrievals around mid-year and in October (Figure 6) could mainly be attributed to seasonal vegetation dynamics and precipitation-related effects. During mid-year (summer), vegetation reaches its peak growth across much of the CONUS, resulting in increased vegetation water content and canopy density, which enhance signal attenuation and scattering in GNSS-R observations and complicate the separation of soil and vegetation contributions [5,10,28]. In October, transitional conditions associated with vegetation senescence, cropland harvest activities, and episodic precipitation events introduce rapid changes in surface and canopy properties that are not always fully captured by the model, leading to increased uncertainty. Similar seasonal degradations in SM retrieval performance during peak vegetation periods have been widely reported in GNSS-R and passive microwave studies [11].

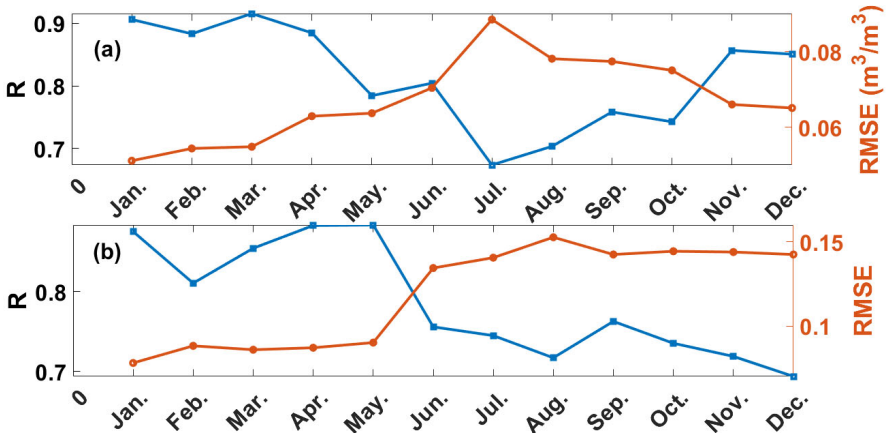

Figure 6. Correlation coefficient (R) and RMSE (**a**) between the predicted SM and SMAP SM and (**b**) between the predicted VOD and SMOS VOD for different months of the year 2021. The blue lines represent R, whereas the orange lines represent the RMSE values.

Analysis of the 2021 time series indicates that the ANN is able to reliably track the temporal variations in soil moisture and vegetation optical depth, closely matching the patterns observed in the reference datasets, SMAP SM and SMOS VOD. In Figures 7 and 8, the top panel presents the ANN-derived estimates; the middle panel shows the corresponding SMAP (for SM) and SMOS (for VOD) reference values; and the bottom panel depicts the daily changes, calculated as the differences between successive days, for both datasets. The fourth row displays the daily total precipitation.

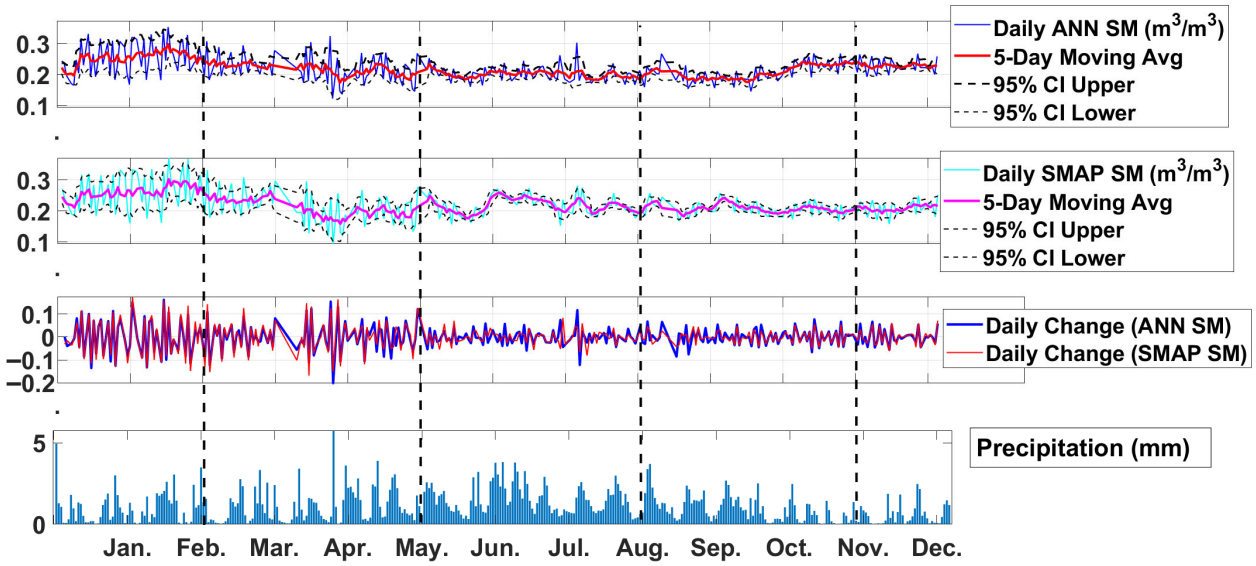

Figure 7. Time-series comparison of ANN-predicted soil moisture (SM; top row) with SMAP observations (second row) for 2021. The panels display daily values, 5-day moving averages with 95% confidence intervals, and day-to-day changes (third row). Total precipitation is shown in the fourth row. Dashed lines are used to distinguish the seasons (spring: March–May; summer: June–August; fall: September–November; winter: December–February).

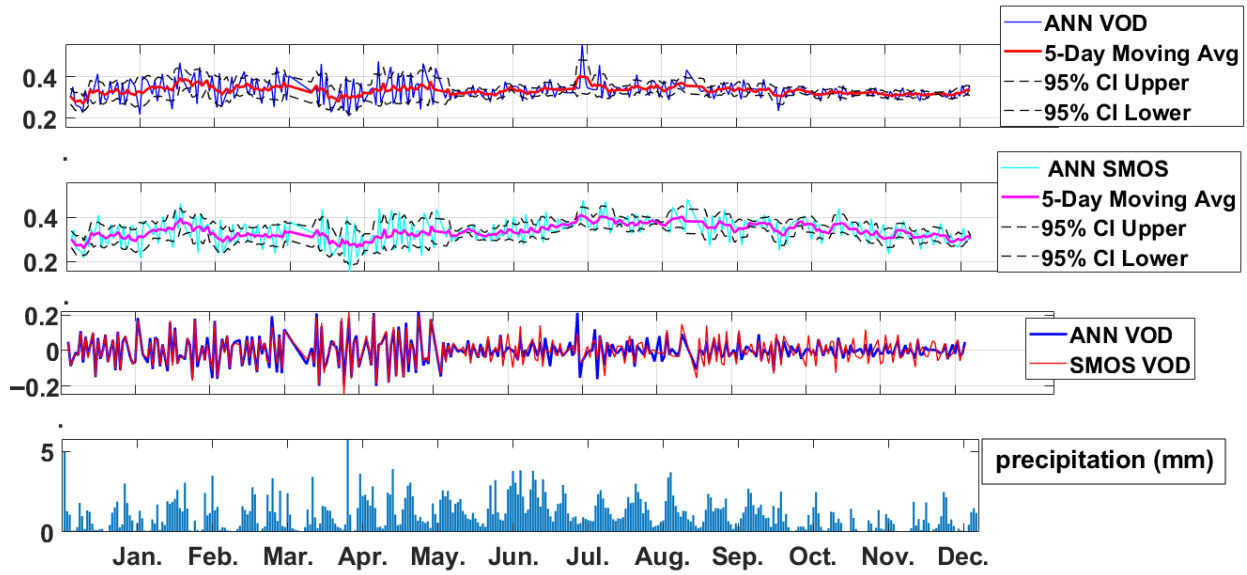

Figure 8. Time-series comparison of ANN-predicted VOD (top row) with SMOS VOD (second row) in 2021. The panels show daily values, 5-day moving averages with 95% confidence intervals, and day-to-day changes (third row). Total precipitation is shown in the fourth row.

For SM, the ANN estimates closely follow SMAP values with high correlation (R = 0.78) and low error (RMSE = 0.0215 $\frac{m^3}{m^3}$), capturing both short-term variability and seasonal trends within the confidence intervals.

Similarly, for VOD, the ANN predictions align well with SMOS observations (R = 0.7562; RMSE = 0.0241), successfully reproducing seasonal dynamics of vegetation. The comparison of day-to-day changes highlights the ability of the ANN in capturing the dynamic variations in both variables. The daily difference series shows very strong correlations between the predicted and reference values (R = 0.93 for SM and R = 0.87 for VOD). These results confirm the robustness of the ANN in generalizing from the training year (2020) to the independent test dataset.

Figure 9 presents two histograms comparing the ANN-predicted SM and VOD with satellite observations. Panels (a) and (b) show that the distributions of predicted SM and SMAP SM, as well as predicted VOD and SMOS VOD, are closely aligned, indicating that the model successfully reproduces the overall variability of both variables.

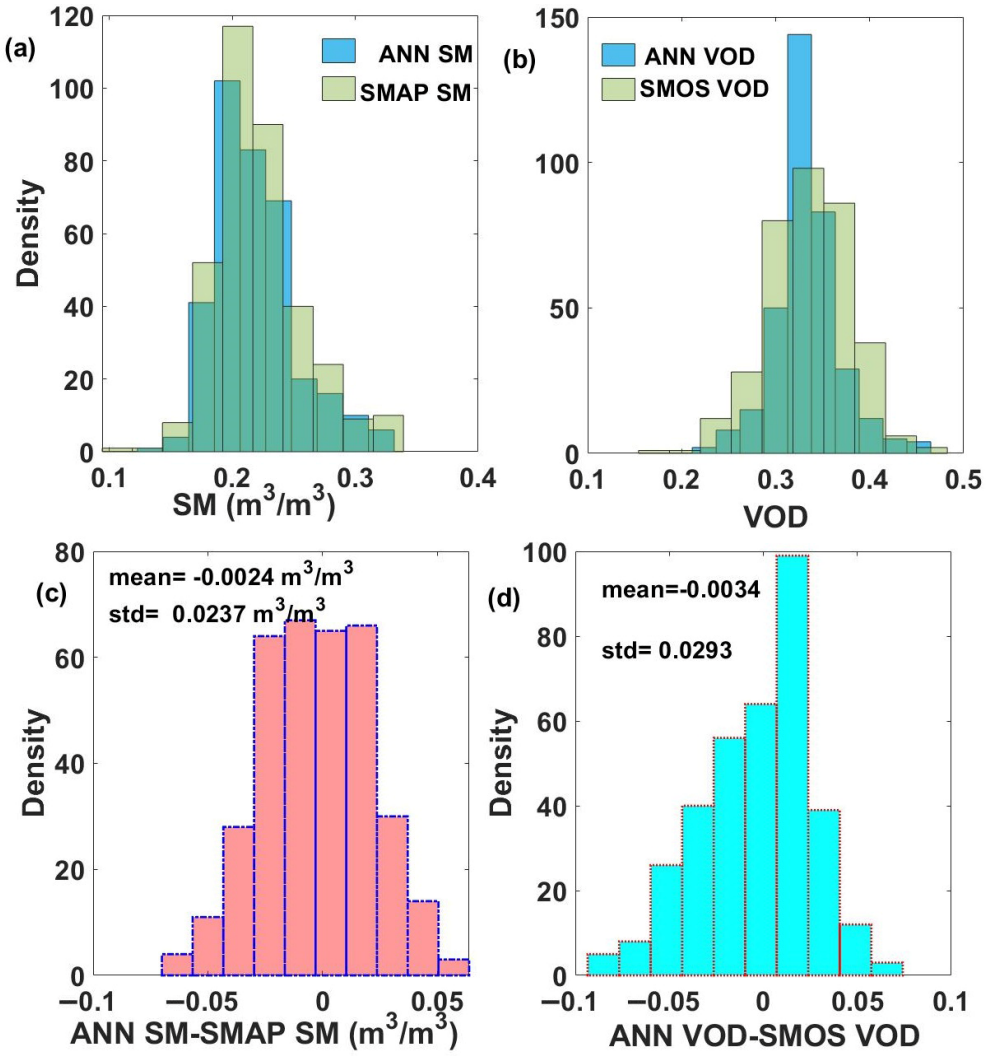

Figure 9. Histograms of (**a**) the predicted SM and SMAP SM, (**b**) the predicted VOD and SMOS VOD, (**c**) the difference between the predicted SM and SMAP SM (shown in pink), and (**d**) the difference between the predicted VOD and SMOS VOD (shown in cyan).

Panels (c) and (d) illustrate the corresponding bias distributions, calculated as the predicted values minus the satellite-derived values. The bias in SM has a mean of -0.0024 m^3/m^3 and a standard deviation of 0.0237 m^3/m^3, indicating minimal systematic error and a narrow spread around zero. Similarly, the VOD bias has a mean of -0.0034 and a standard deviation of 0.0293, confirming that the predicted VOD values are consistent with SMOS data. Together, these results demonstrate that the ANN model provides reliable estimates of both SM and VOD, with only small deviations from satellite-based reference products.

4.2. Spatial Validation of SM and VOD Retrievals

Figure 10a shows the spatial distribution of the number of CYGNSS observations aggregated within a regular 0.2° grid, as defined in [19], for the year 2021. Observation density is highest over the eastern and central United States, particularly along the Gulf

Coast and adjacent areas, whereas the southern and southwestern regions exhibit relatively lower densities. This uneven distribution highlights the spatial sampling characteristics of CYGNSS over land surfaces.

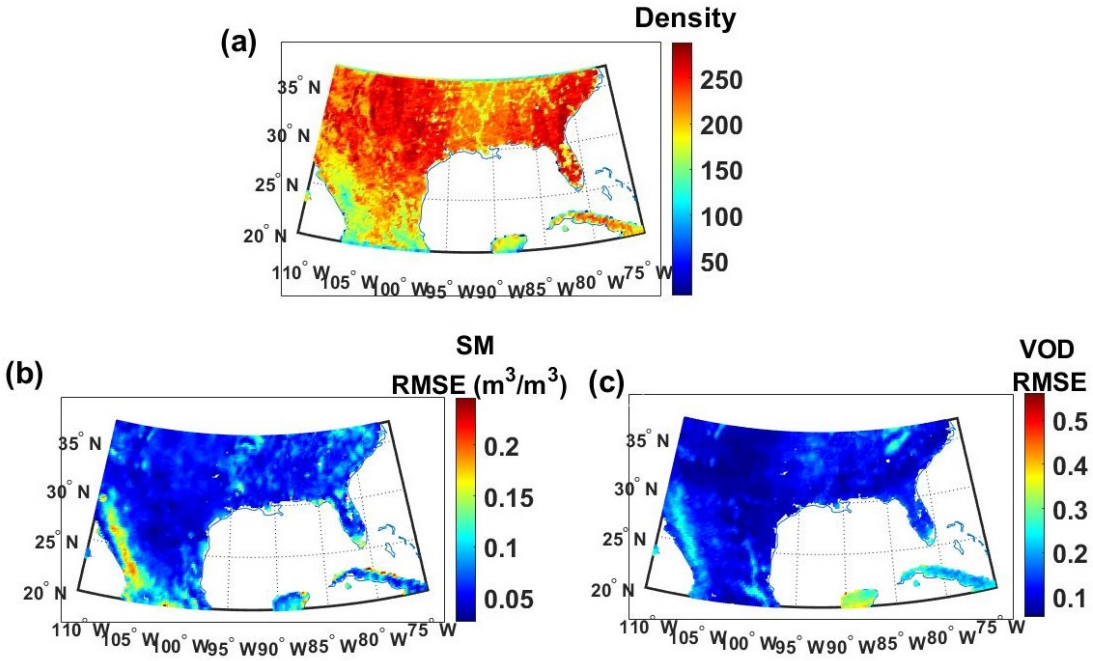

Figure 10. (**a**) The spatial distribution of the number of CYGNSS observations aggregated onto a 0.2° grid over the CONUS for the year 2021. (**b**) Prediction errors (RMSE) of SM compared with SMAP SM. (**c**) Prediction errors (RMSE) of VOD compared with SMOS VOD.

Figure 10b,c show the spatial distribution of the RMSE between the predicted SM and VOD and the corresponding satellite observations (SMAP SM and SMOS VOD). This metric is computed from the time series of collocated values within each 0.2° grid cell shown in Figure 10a for the year 2021. The RMSE of SM remains largely below 0.1 m^3/m^3 across the region, with higher errors concentrated along coastal areas and in regions with sparse observations (see Figure 10a,b). Similarly, the RMSE of VOD is generally below 0.2, with the highest errors also occurring mainly over grids with sparse observations and coastal regions (see Figure 10a,c). These results demonstrate that the model provides reliable predictions for SM and VOD throughout 2021, with performance variations tied to observation density. In other words, our analysis shows that the model generalizes more effectively in regions with dense CYGNSS sampling, whereas its performance decreases in areas with sparse observational coverage.

Figure 11 presents the mean SM and VOD values predicted in this research, alongside the corresponding values provided by SMAP and SMOS, respectively, over the first six months of 2021. For both SM and VOD, the spatial patterns produced by the ANN closely match those of the reference products, successfully capturing the spatial variability present in the SM and VOD fields. The predicted SM exhibits a strong correlation with the reference data ($R = 0.93$) and an RMSE of 0.037 m^3/m^3. Similarly, the predicted VOD shows a high correlation ($R \approx 0.93$), with an RMSE of 0.059. Overall, these findings indicate strong spatial agreement between the ANN-derived maps and the corresponding satellite observations.

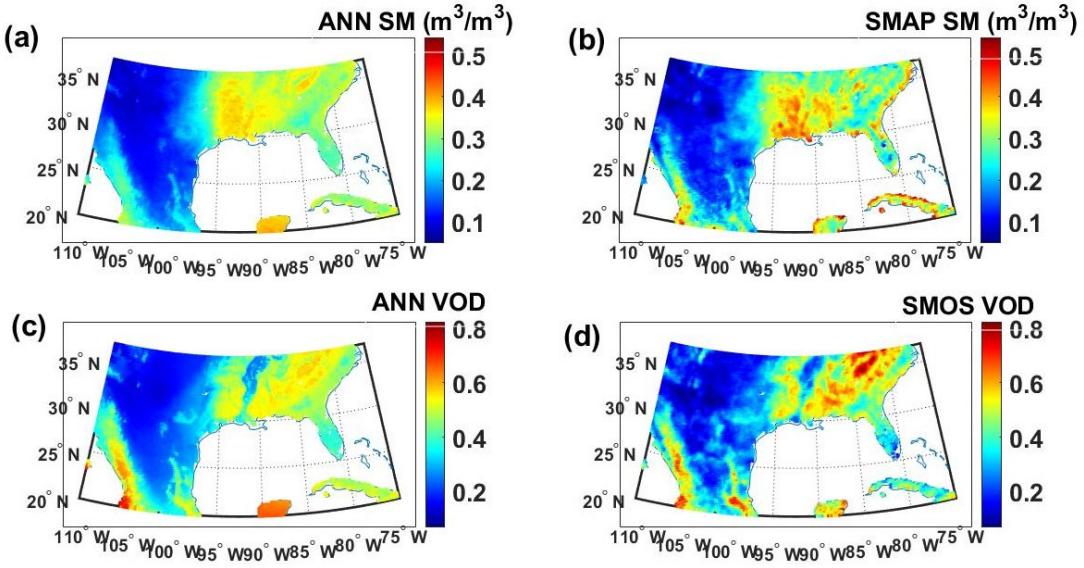

Figure 11. Maps of the mean values of (**a**) predicted SM and (**b**) SMAP SM, as well sa the (**c**) predicted VOD and (**d**) SMOS VOD for the first six months of 2021 (January–June).

4.3. Validation of Predicted SM Values Against In Situ SM Measurements

Figure 12 shows the evaluation of ANN-derived soil moisture estimates from CYGNSS observations against in situ measurements across the CONUS for 2021. The in situ dataset was obtained from the International Soil Moisture Network (ISMN) (https://ismn.earth/en/dataviewer/ (accessed on 1 May 2025)), and only measurements flagged as good quality ("G") were included. A total of 41 distinct stations were analyzed.

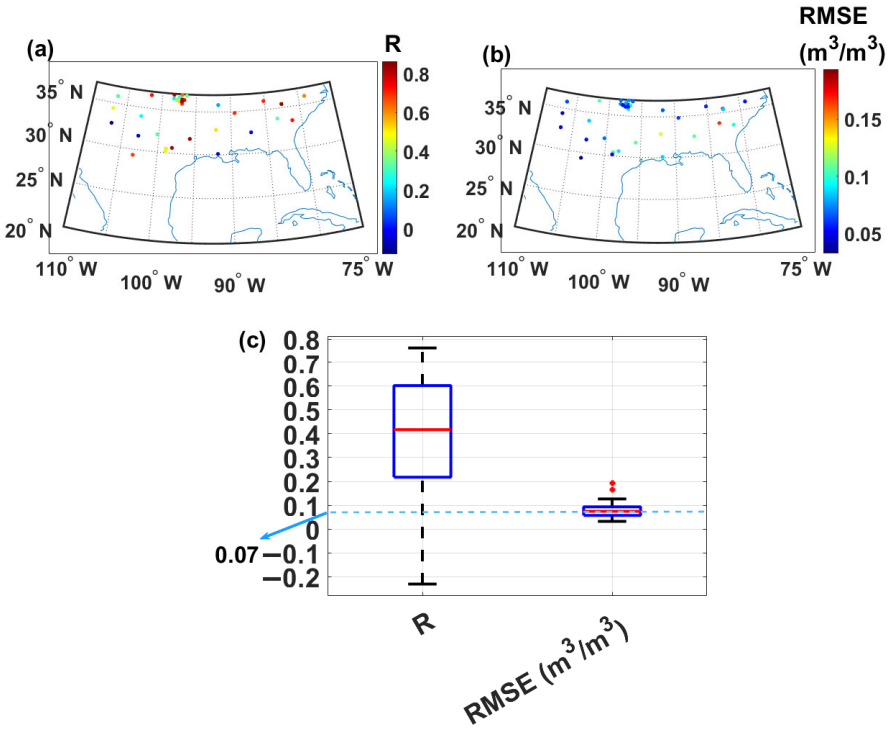

Figure 12. Distribution of (**a**) correlation coefficient and (**b**) RMSE between the ANN-derived SM estimates from CYGNSS and in situ soil moisture, as well as (**c**) their corresponding boxplots. In panel (**c**), black dashed lines represent the full data range excluding outliers, blue boxes indicate the interquartile range (25th–75th percentiles), and red lines denote the median value of each dataset.

The spatial distribution of the correlation coefficient (R; Figure 12a) and its corresponding boxplot (Figure 12c) show that most stations exhibit positive and reasonable correlation values, with some exceeding 0.6. This indicates the model's ability to capture temporal variations in soil moisture. Meanwhile, RMSE values remain below 0.07 m^3/m^3 at the majority of stations (Figure 12 b,c), reflecting relatively low prediction errors.

Overall, these results demonstrate that the ANN driven by CYGNSS observations provides reliable soil moisture estimates when compared with in situ measurements.

5. Discussion

To better understand the strengths and limitations of the proposed retrieval approach, the discussion is organized into four parts. First, Section 5.1 evaluates how ANN-derived VOD responds to vegetation structure. Next, Section 5.2 examines how prediction performance varies across land cover types. Section 5.3 presents an assessment of the influence of each input feature on model performance based on SHAP analysis. Finally, Section 5.4 compares our results with those of previous CYGNSS-based retrieval studies.

5.1. Sensitivity of ANN-Derived VOD to Vegetation Parameters (AGB, LAI, Canopy Height, and VWC)

According to [32,33], VOD correlates with above-ground biomass (AGB), the Leaf Area Index (LAI), and canopy height (CH). Therefore, we used CH data from [34], ERA5 LAI, and the annual AGB map provided by the Climate Change Initiative (CCI) [35] to further evaluate the predicted VOD values.

The CCI AGB data are generated from various Earth observation sources. These sources include the Copernicus Sentinel-1 mission, the ASAR (Advanced Synthetic Aperture Radar) instrument aboard Envisat, and the Japan Aerospace Exploration Agency's (JAXA) Advanced Land Observing Satellites (ALOS-1 and ALOS-2), along with additional Earth observation data [35].

Figure 13 compares VOD estimates from SMOS (red) and the CYGNSS-driven ANN (blue) against CCI AGB, LAI, and canopy height. In all panels, the ANN-derived VOD exhibits a clear positive relationship with vegetation structure, increasing with LAI (R = 0.96), AGB (R = 0.77), and canopy height (R = 0.95). SMOS VOD shows similar trends (R = 0.95 with LAI, R = 0.73 with AGB, and R = 0.98 with canopy height). These results highlight the strong consistency of ANN-derived VOD with independent vegetation indicators, supporting its reliability for biomass and canopy characterization.

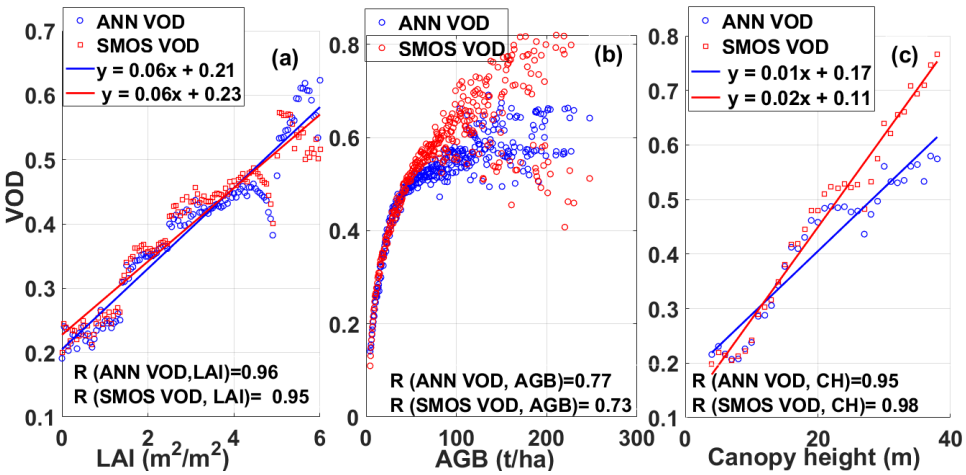

Figure 13. Comparison of VOD estimates from SMOS and ANN against (**a**) LAI, (**b**) AGB, and (**c**) canopy height.

A more detailed analysis is provided in Table 4. For AGB values below 100 t/ha, both ANN-derived VOD (R = 0.8731) and SMOS VOD (R = 0.9291) show strong positive correlations with biomass, indicating that both datasets retain high sensitivity to vegetation density in low-to-moderate biomass conditions. However, when AGB exceeds 100 t/ha, the correlation of ANN VOD decreases substantially (R = 0.3630), while SMOS VOD shows no meaningful relationship with biomass (R = −0.0458). This clear reduction in sensitivity suggests the presence of a saturation effect at higher biomass levels, which is particularly pronounced in SMOS VOD.

Table 4. Correlations between VOD estimates and AGB under different biomass levels.

	ANN VOD	SMOS VOD
AGB $<$ 100	0.8731	0.9291
AGB $>$ 100	0.3630	−0.0458

This behavior is consistent with previous findings reported for other VOD products (e.g., [36]), where microwave signals become less responsive once vegetation reaches high structural density. In our study, we observe similar limitations for GNSS-R-derived VOD in dense forest regions, likely due to signal attenuation, reduced penetration depth, and increased canopy scattering at high biomass levels. These factors collectively reduce the ability of VOD to discriminate biomass beyond a certain threshold.

Figure 14a illustrates the relationship between CYGNSS-ANN-derived VOD and SMAP VWC for different land cover classes identified in Figure 14b. Each point corresponds to a land cover type, with colors representing the mean canopy height. A strong positive linear relationship is observed (R = 0.94), indicating that ANN-derived VOD increases consistently with VWC. The color gradient shows that classes with taller canopies, such as forests, generally have higher VOD values, while classes with shorter vegetation, such as grasslands and shrublands, exhibit lower values. This highlights the strong sensitivity of ANN-derived VOD to both VWC and canopy structure across diverse land cover types.

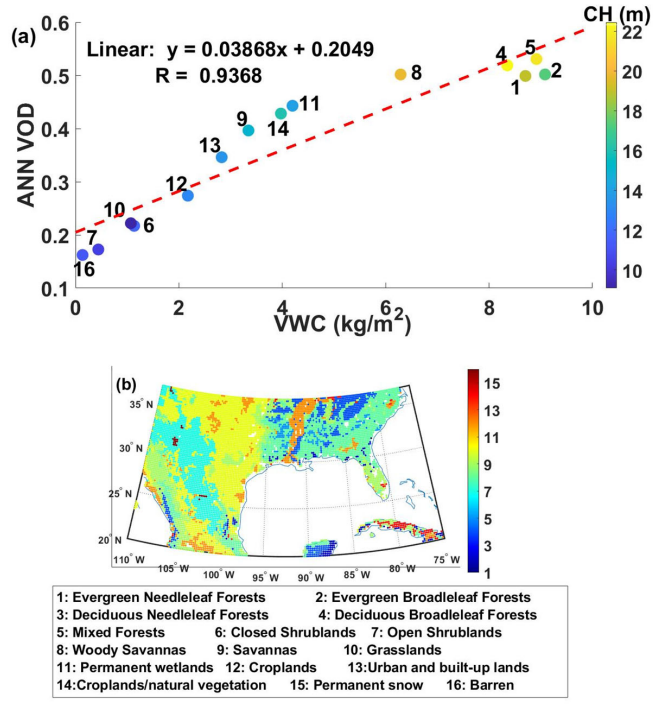

Figure 14. (**a**) Relationship between ANN-derived VOD and VWC across land cover types. (**b**) Yearly land cover classification provided by MODIS (MCD12C1) [37] across the CONUS.

5.2. Influence of Land Cover Classification on Prediction Performance

The observed differences in model performance across land cover types (Figure 15) highlight the varying influences of vegetation structure and surface roughness on GNSS-R observations. Over croplands and grasslands (land cover classes 10, 12, and 14), both soil moisture and vegetation optical depth estimates exhibit relatively high correlations with SMAP SM (up to ~0.93) and SMOS VOD (up to ~0.80). These areas are characterized by moderate vegetation density, seasonal dynamics, and relatively homogeneous surface conditions, which facilitate a more robust separation of soil and vegetation contributions.

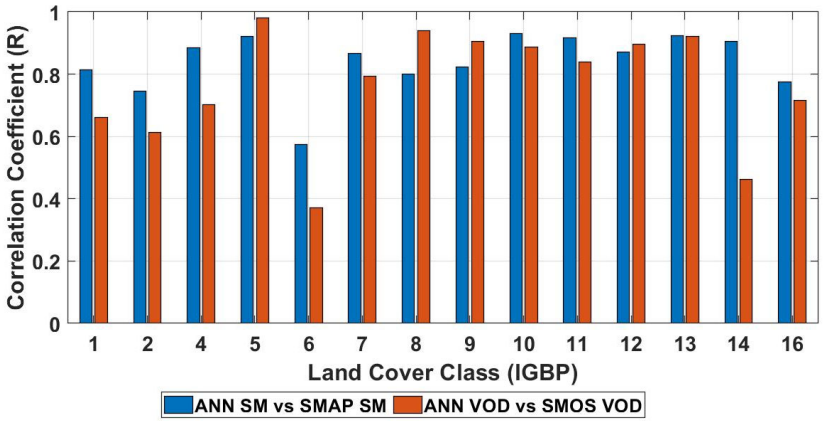

Figure 15. Correlation coefficients between ANN-predicted soil moisture and SMAP soil moisture and between ANN-derived VOD and SMOS VOD across different land cover classifications for the year 2021.

In contrast, forested regions (evergreen and deciduous needleleaf and broadleaf forests; classes 1, 2, 4, and 5) show lower correlations, particularly for VOD. Dense canopy cover, complex vertical vegetation structure, and higher biomass increase signal attenuation and multiple scattering effects, which are not fully represented in the current model parameterization, leading to increased uncertainty in both SM and VOD retrievals, especially in evergreen forests, where vegetation conditions remain relatively stable throughout the year.

Over shrublands and woody savannas (classes 6–8), the performance is intermediate. These regions present mixed surface conditions, with sparse to moderate vegetation and heterogeneous roughness, resulting in variable sensitivity of GNSS-R reflectivity to soil moisture and vegetation changes. Finally, barren and sparsely vegetated areas (class 16) show reduced performance compared to croplands (classes 12), likely due to strong surface roughness effects and the limited vegetation signal, which can amplify noise and model errors.

Overall, these results highlight that the proposed model performs best over land cover types with moderate vegetation density and relatively simple surface conditions, while dense forests and highly heterogeneous or rough surfaces remain more challenging. This land cover-dependent behavior is consistent with the physical sensitivity of GNSS-R observations and underscores the importance of incorporating land cover-specific parameterizations in future model improvements.

It should be noted that this study is limited to the CONUS region; therefore, the model performance and land cover-dependent behavior reported here may not be directly transferable to other climate regimes or surface conditions, and the extension and validation of the approach at the global scale require further investigation.

Figure 16 shows the correlation between VOD (from our method) and SMOS VOD with three vegetation indicators—canopy height, AGB, and LAI—across four major land cover types: forests (classes 1–5), shrublands (classes 6,7), savannas (classes 8,9), and

grassland and croplands (classes 10,12,14). Overall, both VOD datasets exhibit strong correlations with CH and AGB across most land cover types, while correlations with LAI are generally lower, particularly in shrublands. The comparison highlights that our VOD closely follows SMOS VOD patterns, demonstrating its reliability across diverse vegetation types. Our VOD demonstrates improved correlation compared to SMOS VOD in several cases, particularly with LAI across all land cover types, as well as with AGB in forests and canopy height in grassland and croplands. This suggests that our method captures vegetation structure more effectively under these conditions.

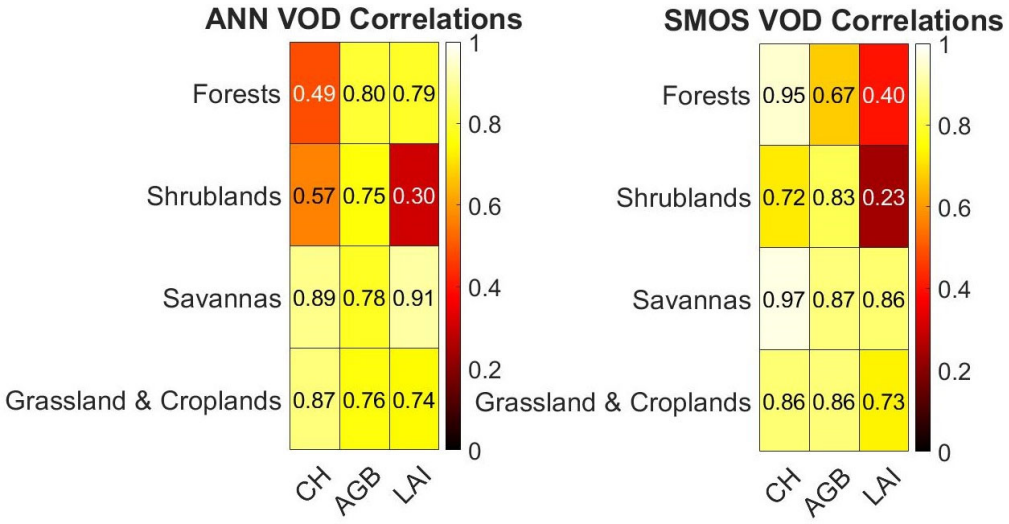

Figure 16. Comparison of ANN-derived VOD and SMOS VOD correlations with vegetation indicators across vegetation classes.

5.3. Contribution of Each Input Feature to the Prediction Outcomes

To enhance model interpretability, a feature importance analysis was conducted using SHAP (SHapley Additive exPlanations) values. Figure 17a illustrates the mean absolute SHAP values for soil moisture retrieval, quantifying the relative contribution of each input variable to the ANN predictions. The results indicate that land surface descriptors and GNSS-R-related parameters (e.g., surface roughness and reflectivity) play a dominant role in the retrieval. In addition, meteorological variables—particularly, soil surface temperature and precipitation—also play a noticeable role in soil moisture prediction by capturing short-term hydrological and energy-balance processes. In contrast, ancillary variables such as soil texture and incidence angle exhibit a comparatively smaller influence on the model outcomes.

Figure 17b shows that the VOD retrieval is predominantly governed by spatial inputs (longitude and latitude) and surface roughness, underscoring the strong influence of geographic location and land surface structure on vegetation optical depth. The strong influence of spatial coordinates on VOD retrievals is consistent with previous GNSS-R studies, which reported a dominant spatial dependency in VWC estimation from CYGNSS observations [17], reflecting large-scale climatic and ecological gradients rather than model bias. Among the remaining variables, GNSS-R reflectivity exhibits a stronger contribution than precipitation, indicating that microwave scattering information provides more direct sensitivity to vegetation conditions than purely meteorological inputs. Soil temperature plays a moderate role, while soil texture (clay content) and incidence angle have a comparatively minor influence on the model predictions.

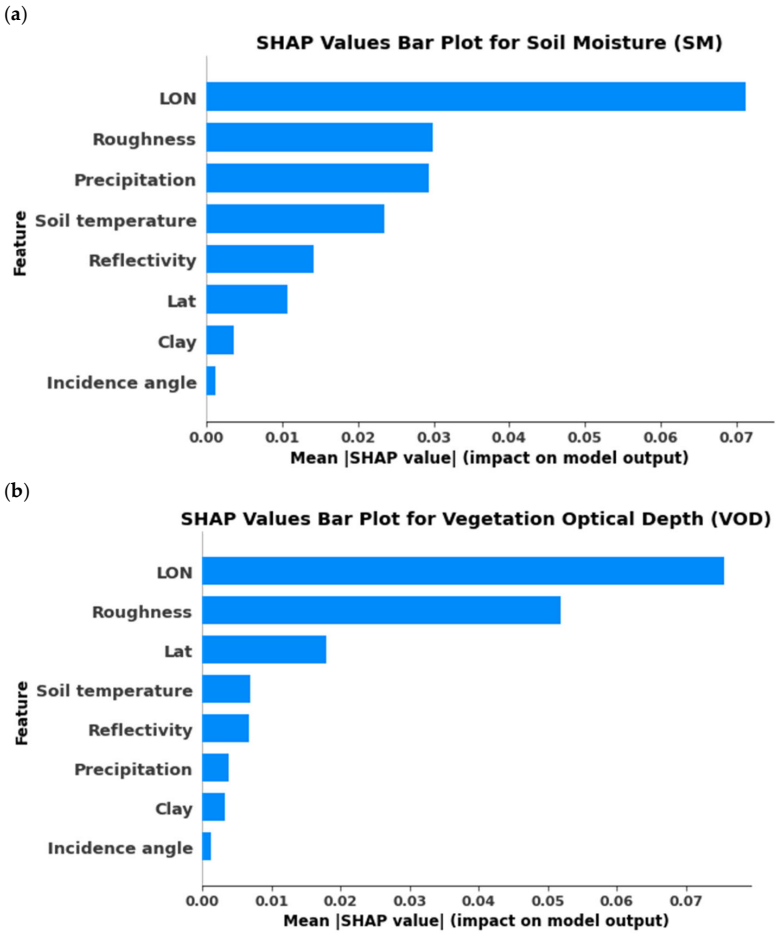

Figure 17. Mean absolute SHAP values of the input features for (**a**) SM prediction and (**b**) VOD prediction.

5.4. Comparisons with the Literature

Compared with [18], which employs a stand-alone CYGNSS approach using two separately trained neural networks and an iterative exhaustive search to retrieve SM and VOD, our method introduces several key advances. We integrate CYGNSS reflectivity with ancillary geophysical and meteorological variables to retrieve SM and VOD simultaneously using a single multi-output ANN, thereby ensuring internal consistency between the two variables. In addition, we provide focused, high-resolution validation over the CONUS, including detailed spatial and temporal comparisons against independent reference datasets, demonstrating improved robustness relative to the approach reported in [18].

While ref. [18] reported RMSE values of approximately 0.07 and 0.06 m^3/m^3 for VOD and SM, respectively, relative to SMAP-derived VOD and SM, we observed RMSE values of 0.037 m^3/m^3 for SM and 0.059 for VOD for the year 2021. These results demonstrate a notable improvement in SM and VOD prediction accuracy compared with the previous study, highlighting the benefits of incorporating ancillary datasets.

6. Conclusions

This study presented a novel deep learning framework for the simultaneous retrieval of soil moisture and vegetation optical depth from spaceborne GNSS-R observations acquired by the CYGNSS mission. By fusing CYGNSS reflectivity and observation geometry with ancillary geophysical and meteorological datasets within a multi-output ANN, the proposed approach effectively captures the coupled soil–vegetation signal and addresses key sources of uncertainty, including surface roughness, vegetation attenuation, and precipitation effects.

Comprehensive temporal and spatial validations over the CONUS demonstrated strong agreement between the ANN-derived products and independent reference datasets. The retrieved soil moisture showed high consistency with SMAP SM (R = 0.83, RMSE = 0.063 m^3/m^3), while the predicted VOD closely matched SMOS VOD (R = 0.89, RMSE = 0.088). Additional validation against in situ soil moisture measurements confirmed the model's ability to reproduce observed temporal variability with low error across most stations. The ANN-derived VOD further exhibited robust correlations with independent vegetation indicators, including above-ground biomass, canopy height, leaf area index, and vegetation water content, highlighting its sensitivity to vegetation structure and water status.

The land cover-dependent analysis revealed that the proposed framework performs best over regions with moderate vegetation density, such as croplands and grasslands, while denser forested areas remain more challenging due to signal attenuation and saturation effects.

Overall, this work demonstrates that GNSS-R observations, when combined with deep learning and ancillary datasets, provide a robust and scalable solution for joint SM and VOD retrieval, offering a cost-effective complement to traditional passive microwave missions such as SMAP and SMOS. The high spatiotemporal sampling resolution of CYGNSS further highlights the potential of GNSS-R for the monitoring of land surface hydrological and vegetation dynamics.

Looking ahead, the proposed framework is highly relevant to upcoming missions, such as HydroGNSS, that are expected to provide enhanced GNSS-R observations optimized for hydrological applications. The methodology developed in this study can be directly extended to HydroGNSS data to support global-scale, high-resolution monitoring of soil moisture and vegetation properties. Future work could also explore the fusion of multi-source remote sensing data, including SAR observations, with GNSS-R measurements to further improve sensitivity under dense vegetation and complex surface conditions. In addition, advanced machine learning approaches, including random forests, XGBoost, and emerging spatiotemporal graph neural networks, offer promising opportunities to better model the nonlinear spatial and temporal dependencies inherent in GNSS-R DDMs, potentially leading to further improvements in retrieval accuracy, robustness, and transferability across regions and climate regimes.

Author Contributions: Conceptualization, M.R. and J.A.; methodology, M.R., J.A., and A.A.-S.; software, M.R.; validation, M.R.; formal analysis, M.R., J.A., and A.A.-S.; investigation, M.R.; data curation, M.R.; writing—original draft preparation, M.R.; writing—review and editing, J.A. and A.A.-S.; visualization, M.R.; supervision, J.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding and The APC was funded by Delft University of Technology.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

Acknowledgments: We acknowledge the use of multiple satellite, reanalysis, and in situ datasets that were fundamental to this study. Specifically, we utilized SMAP soil moisture (SM) and vegetation water content (VWC), SMOS vegetation optical depth (VOD), and CYGNSS Level-1 observations. We also benefited from ERA-5 reanalysis products, including top 0–7 cm soil surface temperature and total precipitation, as well as GLDAS soil clay content. In situ soil moisture data from the International Soil Moisture Network (ISMN) and CCI above-ground biomass (AGB) were also incorporated. We gratefully acknowledge the mission teams and data providers for making these datasets available, which enabled the analyses and findings presented in this work.

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

References

1. Palfreyman, A.; Curial, A.d.S.; Rawlinson, J.; Unwin, M.; Sweeting, M.; Donati, S.; Nazzareno, P.; Estel, C.; Leila, G.; Giuseppe, F.; et al. Scouting for climate variable with small satellites. In Proceedings of the AIAA/USU Conference on Small Satellites, Weekday Session V, SSC22-V-05, Logan, UT, USA, 6–11 August 2022.
2. Frappart, F.; Wigneron, J.-P.; Li, X.; Liu, X.; Al-Yaari, A.; Fan, L.; Mengjia, W.; Christophe, M.; Erwan, L.M.; Zacharie, A.L.; et al. Global monitoring of the vegetation dynamics from the vegetation optical depth (VOD): A review. *Remote Sens.* **2020**, *12*, 2915. [CrossRef]
3. Entekhabi, D.; Yueh, S.; O'Neill, P.E.; Kellogg, K.H.; Allen, A.; Bindlish, R.; Molly, B.; Steven, C.; Andreas, C.; Wade, T.C.; et al. *SMAP Handbook—Soil Moisture Active Passive: Mapping Soil Moisture and Freeze/Thaw from Space*; NASA JPL Publication: Pasadena, CA, USA, 2014.
4. Kerr, Y.H.; Waldteufel, P.; Richaume, P.; Wigneron, J.P.; Ferrazzoli, P.; Mahmoodi, A.; Al Bitar, A.; Cabot, F.; Gruhier, C.; Juglea, S.E. The SMOS soil moisture retrieval algorithm. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1384–1403. [CrossRef]
5. 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]
6. Xu, X.; Yueh, S.; Shah, R.; Hayashi, A. Vegetation optical depth retrieval from CYGNSS data. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Brussels, Belgium, 11–16 July 2021. [CrossRef]
7. Ghiasi, Y.; Duguay, C.R.; Murfitt, J.; Asgarimehr, M.; Wu, Y. Potential of GNSS-R for the monitoring of lake ice phenology. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2023**, *17*, 660–673. [CrossRef]
8. Rajabi, M.; Nahavandchi, H.; Hoseini, M. Evaluation of CYGNSS observations for flood detection and mapping during Sistan and Baluchestan torrential rain in 2020. *Water* **2020**, *12*, 2047. [CrossRef]
9. Rahmani, M.; Asgari, J.; Asgarimehr, M. Soil moisture retrieval using space-borne GNSS reflectometry: A comprehensive review. *Int. J. Remote Sens.* **2022**, *43*, 5173–5203. [CrossRef]
10. Camps, A.; Park, H.; Pablos, M.; Foti, G.; Gommenginger, C.P.; Liu, P.-W.; 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]
11. Chew, C.; Small, E. Description of the UCAR/CU soil moisture product. *Remote Sens.* **2020**, *12*, 1558. [CrossRef]
12. Carreno-Luengo, H.; Luzi, G.; Crosetto, M. Impact of the elevation angle on CYGNSS GNSS-R bistatic reflectivity as a function of effective surface roughness over land surfaces. *Remote Sens.* **2018**, *10*, 1749. [CrossRef]
13. Wilson, M.; Datta, R.; Savarimuthu, S.; Moller, D.; Ruf, C. Prediction of Soil Moisture from Near-Global CYGNSS GNSS-Reflectometry Using a Random Forest Machine Learning Model. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Athens, Greece, 7–12 July 2024. [CrossRef]
14. 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]
15. Yan, Q.; Chen, Y.; Pan, Y.; Jin, S.; Huang, W. A modified hierarchical vision transformer for soil moisture retrieval from CYGNSS data. *Water Resour. Res.* **2026**, *62*, e2024WR039476. [CrossRef]
16. Asgarimehr, M.; Entekhabi, D.; Camps, A. Diurnal vegetation moisture cycle in the Amazon and response to water stress. *Geophys. Res. Lett.* **2024**, *51*, e2024GL111462. [CrossRef]
17. Zhao, D.; Asgarimehr, M.; Heidler, K.; Wickert, J.; Zhu, X.X.; Mou, L. Deep learning-based GNSS-R global vegetation water content: Dataset, estimation, and uncertainty. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2025**, *18*, 17386–17404. [CrossRef]
18. Yan, Q.; Jin, S.; Huang, W.; Jia, Y. Stand-Alone Retrievals of Soil Moisture and Vegetation Opacity Using the CYGNSS Data. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Brussels, Belgium, 11–16 July 2021. [CrossRef]
19. Rahmani, M.; Asgari, J.; Asgarimehr, M.; Wickert, J. Investigation of the global influence of surface roughness on space-borne GNSS-R observations. *J. Geophys. Res. Biogeosci.* **2025**, *130*, e2024JG008243. [CrossRef]
20. Yan, Q.; Gong, S.; Jin, S.; Huang, W.; Zhang, C. Near real-time soil moisture in China retrieved from CYGNSS reflectivity. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 1–5. [CrossRef]
21. Yang, T.; Wan, W.; Sun, Z.; Liu, B.; Li, S.; Chen, X. Comprehensive evaluation of using TechDemoSat-1 and CYGNSS data to estimate soil moisture over mainland China. *Remote Sens.* **2020**, *12*, 1699. [CrossRef]
22. Copernicus Climate Change Service (C3S). (2019): ERA5-Land Hourly Data from 1950 to Present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Available online: https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land?tab=overview (accessed on 1 November 2023).
23. Rodell, M.; Houser, P.; Jambor, U.; Gottschalck, J.; Mitchell, K.; Meng, C.-J.; Arsenault, K.; Cosgrove, B.; Radakovich, J.; Bosilovich, M.; et al. The global land data assimilation system. *Bull. Am. Meteorol. Soc.* **2004**, *85*, 381–394. [CrossRef]
24. CYGNSS. 2020. CYGNSS Level 1 Full Delay Doppler Map Data Record Version 3.0. Ver. 3.0. PO.DAAC, CA, USA. Dataset. Available online: https://podaac.jpl.nasa.gov/dataset/CYGNSS_L1_FULL_DDM_V3.0 (accessed on 1 January 2024).

25. O'Neill, P.E.; Chan, S.; Njoku, E.G.; Jackson, T.; Bindlish, R. *SMAP Enhanced L3 Radiometer Global Daily 9 km EASE-Grid Soil Moisture (SPL3SMP_E, Version 2)*; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2018.
26. Kerr, Y.H.; Waldteufel, P.; Wigneron, J.-P.; Delwart, S.; Cabot, F.; Boutin, J.; Mecklenburg, S. The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proc. IEEE* **2010**, *98*, 666–687. [CrossRef]
27. Balakhder, A.M.; Al-Khaldi, M.M.; Johnson, J.T. On the coherency of ocean and land surface specular scattering for GNSS-R and signals of opportunity systems. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 10426–10436. [CrossRef]
28. Yueh, S.H.; Shah, R.; Chaubell, M.J.; Hayashi, A.; Xu, X.; Colliander, A. A semiempirical modeling of soil moisture, vegetation, and surface roughness impact on CYGNSS reflectometry data. *IEEE Trans. Geosci. Remote Sens.* **2020**, *60*, 1–17. [CrossRef]
29. Schwertman, N.C.; Owens, M.A.; Adnan, R. A simple more general boxplot method for identifying outliers. *Comput. Stat. Data Anal.* **2004**, *47*, 165–174. [CrossRef]
30. Thakker, Z.L.; Buch, S.H. Effect of feature scaling pre-processing techniques on machine learning algorithms to predict particulate matter concentration for Gandhinagar, Gujarat, India. *Int. J. Sci. Res. Sci. Technol.* **2024**, *11*, 410–419. [CrossRef]
31. Gavili Kilane, N. Application of Physics-Informed Neural Network Approach in Soil Moisture Retrieval Using GNSS Reflectometry. Ph.D. Thesis, York University, Toronto, Canada, December 2023.
32. Chen, F.; Guo, F.; Liu, L.; Nan, Y. An improved method for pan-tropical above-ground biomass and canopy height retrieval using CYGNSS. *Remote Sens.* **2021**, *13*, 2491. [CrossRef]
33. 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]
34. Simard, M.; Pinto, N.; Fisher, J.B.; Baccini, A. Mapping forest canopy height globally with spaceborne lidar. *J. Geophys. Res. Biogeosc.* **2011**, *116*. [CrossRef]
35. Santoro, M.; Cartus, O. *ESA Biomass Climate Change Initiative (Biomass_cci): Global Datasets of Forest Above-Ground Biomass for the Years 2010, 2017, 2018, 2019 and 2020, v4*; NERC EDS Centre for Environmental Data Analysis: Chilton, UK, 2023.
36. Chang, Z.; Fan, L.; Wigneron, J.-P.; Wang, Y.-P.; Li, X.; Wang, M.; Yan, J. Evaluation of optical and microwave-derived vegetation indices for monitoring aboveground biomass over China. *Geo-Spat. Inf. Sci.* **2025**, *28*, 421–436. [CrossRef]
37. Friedl, M.; Sulla-Menashe, D. MCD12C1 MODIS/Terra+ Aqua Land Cover Type Yearly L3 Global 0.05 Deg CMG V006. 2015. [Data Set]. NASA Land Processes Distributed Active Archive Center. Available online: https://www.earthdata.nasa.gov/data/catalog/lpcloud-mcd12c1-006 (accessed on 15 March 2024).