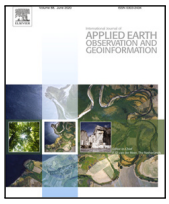
International Journal of Applied Earth Observation and Geoinformation 142 (2025) 104658

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Multimodal GNSS-R self-supervised learning as a generalist Earth surface monitor☆

Daixin Zhao [a,b,c], Konrad Heidler [a], Milad Asgarimehr [c], Conrad M. Albrecht [b], Jens Wickert [c,d], Xiao Xiang Zhu [a], Lichao Mou [a,*]

[a] Chair of Data Science in Earth Observation, Technical University of Munich, Munich, Germany
[b] German Aerospace Center (DLR), Oberpfaffenhofen, Germany
[c] Section 1.1 Space Geodetic Techniques, GFZ Helmholtz Centre for Geosciences, Potsdam, Germany
[d] Institute of Geodesy and Geoinformation Science, Technische Universität Berlin, Germany

ARTICLE INFO

Keywords:
Climate action
CYGNSS
Earth observation
Foundation model
GNSS reflectometry
Self-supervised learning

ABSTRACT

The increasing frequency of climate extremes and natural disasters demands rapid and scalable Earth surface scans for effective action. Emerging as a novel remote sensing technique, spaceborne global navigation satellite system reflectometry (GNSS-R) plays an increasingly vital role in monitoring Earth's surface parameters. Recent studies leverage the growing volume of GNSS-R measurements with data-driven approaches to enhance retrieval products over both ocean and land. Yet, these models are typically trained using supervised learning, which requires extensive feature engineering and application-specific annotations. To address these limitations, we propose the first GNSS-R self-supervised learning framework as a generalist Earth surface monitor (GEM). Our model is pretrained on multimodal observables, i.e., delay-Doppler maps (DDMs) and auxiliary parametric data, to learn cross-modal representations from GNSS-R data. To validate the effectiveness of the proposed approach, we fine-tune the pretrained model on various downstream retrieval tasks, including ocean wind speed retrieval, surface soil moisture estimation, and vegetation water content prediction. The results demonstrate that our framework generalizes well across these tasks, providing a versatile solution for GNSS-R-based Earth surface monitoring and facilitating further exploration of novel use cases.

Contents

☆ This article is part of a special issue entitled: 'Foundation models EO' published in International Journal of Applied Earth Observation and Geoinformation.

* Corresponding author.
E-mail address: lichao.mou@tum.de (L. Mou).

https://doi.org/10.1016/j.jag.2025.104658
Received 26 February 2025; Received in revised form 29 May 2025; Accepted 5 June 2025
Available online 21 June 2025

1. Introduction

Tropical cyclones, droughts, and wildfires constitute significant threats to society through severe damage to infrastructure, ecosystems, and human lives. With the increasing frequency and intensity of these extreme weather events, e.g., the devastating Hurricane Helene and the January 2025 wildfires in the suburbs of Los Angeles, monitoring Earth's surface parameters is critical for climate resilience (Balsamo et al., 2018). As these parameters are direct indicators of environmental changes and potential disasters, accurate and timely observations are essential to improve early warning systems and support mitigation strategies.

Spaceborne global navigation satellite system reflectometry (GNSS-R) has emerged as a promising remote sensing technique that offers new opportunities for Earth surface monitoring. Utilizing surface reflected signals of existing navigation satellites as signals-of-opportunity, GNSS-R receivers essentially function as L-band passive multi-bistatic radars (Pierdicca et al., 2022). One of the most important GNSS-R observables is the delay-Doppler map (DDM), a two-dimensional (2D) representation of the reflected signal power as a function of propagation delay and Doppler shift. The reflection pattern and intensity are dependent on the geophysical properties of the reflecting media, which can be extracted by analyzing these reflected signals. Compared to other remote sensing methods, GNSS-R provides unprecedented spatiotemporal coverage while being cost-effective, energy-efficient, and operable in all weather conditions around the clock (Winkelried et al., 2023). Consequently, spaceborne GNSS-R constellations are expected to provide global coverage and full temporal availability for Earth surface monitoring while simultaneously capturing observations along multiple tracks across wide areas. Several spaceborne missions, e.g., NASA's Cyclone GNSS (CYGNSS) (Ruf et al., 2017) and ESA's PRETTY mission (Dielacher et al., 2022), have demonstrated strong capabilities in retrieving surface parameters. The commercial sector has also made several investments in this technology, such as the Muon Space constellation (Masters et al., 2023) and the Spire constellation (Jales et al., 2020) with more than 40 operating satellites. The upcoming ESA's HydroGNSS mission (Unwin et al., 2021) with two satellites is expected to further enhance these capabilities.

Owing to the improved capabilities of SmallSats platforms, GNSS-R constellations are generating growing data volumes on the order of millions of measurements per day. There is an emerging trend among researchers to utilize data-driven approaches in combination with the massive amount of GNSS-R measurements to improve existing products, develop novel applications, and achieve enhanced performance (Zhu et al., 2017; Yuan et al., 2020). Existing studies can be categorized into two main groups: (1) oceanic applications, such as wind speed retrieval; and (2) terrestrial applications, including soil moisture estimation and vegetation monitoring. The first group is established on the fact that GNSS-R observations are sensitive to surface roughness, which is directly influenced by wind-driven waves. Pioneering works use data-driven methods such as multilayer perceptrons (MLPs) (Liu et al., 2019; Asgarimehr et al., 2020; Reynolds et al., 2020; Li et al., 2021), convolutional neural networks (CNNs) (Asgarimehr et al., 2022; Guo et al., 2022; Liu et al., 2023; Xiao et al., 2024a), and Transformers (Zhao et al., 2023) to retrieve ocean wind speed. The second group is gaining increasing attention in the GNSS-R community with a broader range of applications, as soil and vegetation states contain key parameters for disaster monitoring, such as wildfires and drought events. Since soil water content affects the soil's dielectric constant and, in turn, the Fresnel reflection coefficients, GNSS-R observations can be used for surface soil moisture estimation (Nguyen et al., 2025). Several studies have demonstrated the effectiveness of applying random forests (Jia et al., 2020; Lei et al., 2022), MLPs (Eroglu et al., 2019; Senyurek et al., 2020b), and CNNs (Roberts et al., 2022; Nabi et al., 2023) for improving retrieval accuracy. Moreover, as GNSS signals traverse vegetation layers, their energy is attenuated during transmission to and reflection from the ground (Yueh et al., 2022). A few pioneering investigations have explored the synergy between abundant GNSS-R measurements and deep learning to prove the feasibility of vegetation water content (VWC) estimation (Zhao et al., 2024; Chen et al., 2024; Zhang et al., 2024a). Further, data-driven methods have been utilized for GNSS-R-based above-ground biomass retrieval (Santi et al., 2020; Chen et al., 2021; Pilikos et al., 2024), inland water body detection (Kossieris et al., 2023), and wildfire studies (Santi et al., 2022).

Despite the promising results demonstrated by previous supervised learning approaches in GNSS-R applications, these methods rely heavily on large quantities of annotated data. This dependency limits model scalability to novel applications and receiver configurations, particularly when labeled data is scarce. Moreover, supervised methods often involve task-specific feature engineering and model design, which constrain their transferability and extensibility. Recent remote sensing studies suggest that these limitations can be addressed by leveraging self-supervised learning (Wang et al., 2022; Jakubik et al., 2023; Zhu et al., 2024). Compared to supervised learning methods, self-supervised learning enables representation learning from abundant unlabeled data generated by existing and upcoming GNSS-R constellations. By capturing underlying patterns in GNSS-R observables that are relevant across different tasks, this approach reduces the costs of manual annotation and computational resources while improving the adaptability and generalizability of retrieval models. For instance, masked autoencoder (MAE) (He et al., 2022) has become one of the most widely adopted frameworks for self-supervised learning and foundation model development in remote sensing. It involves masking a portion of the input signals and training a network to reconstruct the original data from partial observations. This approach not only improves computational efficiency but also enhances the model's ability to learn robust feature representations without requiring data annotations. Cong et al. (2022) propose an MAE-based self-supervised learning framework to leverage temporal or multispectral information in satellite imagery. Experiment results yield strong improvements on downstream tasks, such as land cover classification and semantic segmentation. Another MAE-based framework, termed Scale-MAE, integrates scale information from imagery inputs through frequency-aware bandpass filters and achieves an average of 4% improvement across eight remote sensing datasets compared to competitor methods (Reed et al., 2023). Wang et al. (2025) introduce feature-guided MAE by setting up varied reconstruction targets for multispectral and SAR imagery, demonstrating both the

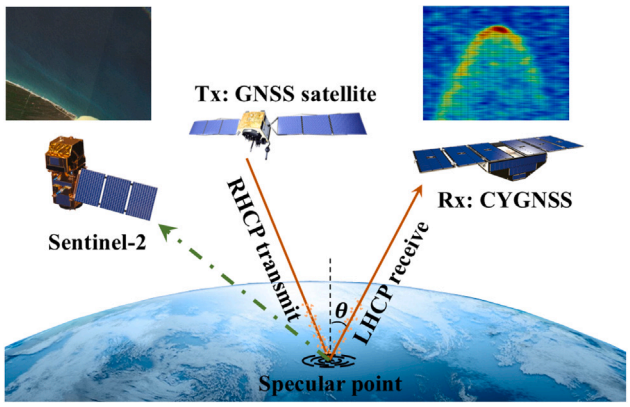


Fig. 1. Comparison of optical satellite imagery (Sentinel-2) and GNSS reflectometry satellite (CYGNSS) delay-Doppler maps (DDMs) from the same geolocation. A satellite imagery pixel measures surface reflectance for a specific wavelength band, while a DDM pixel represents reflected signal power with respect to propagation delay and Doppler shift.

effectiveness and scalability of MAE-based self-supervised pretraining. In addition, MAE-based frameworks that consider varying input size and resolution, sensor parameters, and spatiotemporal features are investigated by Hong et al. (2024), Prexl and Schmitt (2024), and Zhang et al. (2024b), respectively.

A key obstacle to applying pretrained remote sensing models to GNSS-R applications is the unique structural difference between GNSS-R observables and remote sensing imagery. As shown in Fig. 1, a multispectral image pixel measures the surface reflectance of a geolocation for a given band of wavelengths emitted by the sun. On the contrary, a pixel of the GNSS-R observable (i.e., DDMs) represents the reflected GNSS signal's power as a function of propagation delay and Doppler shift. In addition, GNSS-R measurements are typically accompanied by metadata related to measurement geometry, instrument calibration, and atmospheric effects, all of which are crucial for accurately deriving surface parameters. While recent works such as DDM-Former (Zhao et al., 2023) apply Transformer architectures to process DDMs, these approaches are limited to single modalities. The integration of auxiliary parameters into pretrained models remains an emerging field that has yet to be fully explored. To address these gaps, we propose a GNSS-R generalist Earth surface monitor (GEM) framework, the first multimodal GNSS-R self-supervised learning scheme for representation learning from DDMs and auxiliary parametric data. We demonstrate that an asymmetric encoder–decoder architecture with a feature fusion module can effectively capture cross-modal representations in multimodal GNSS-R data. Fine-tuning on various downstream tasks validates the effectiveness of our approach and highlights its potential for developing a scalable Earth surface monitoring system using GNSS-R observations.

The remainder of the paper is organized as follows: Section 2 describes datasets, quality control, and collocation strategy used in this study. Section 3 details the proposed GEM pretraining, including components, downstream fine-tuning, and experimental setups. Section 4 presents numerical and geographical results, along with training efficiency, reconstruction visualizations, and current challenges. Finally, Section 5 provides the overall summary of this work and future perspectives.

2. Materials

2.1. Satellite data

2.1.1. CYGNSS

The CYGNSS is a NASA mission consisting of a constellation of eight small satellites that are first dedicated to GNSS-R applications (Ruf et al., 2018). Although the primary focus was on studying air-sea interactions across regions between 38° N and 38° S, recent research has demonstrated its capabilities for terrestrial applications. With a mean and a median revisit time of 7.2 h and 2.8 h, CYGNSS allows timely observations over the tropics and subtropics.

Our GEM framework is pretrained on CYGNSS L1 science data record version 3.1 (CYGNSS, 2021). We use four types of DDMs that include DDM bistatic radar cross section (BRCS), the corresponding effective scattering area, analog power, and raw counts. In terms of per-DDM auxiliary parameters, we specifically select fifteen variables based on their theoretical relevance to bistatic radar signal formation and empirical contributions validated through explainable AI techniques, specifically SHAP (SHapley Additive exPlanations) value analysis (Xiao et al., 2024b). SHAP values quantify the impact of each feature on model predictions, allowing for the identification of variables with consistent influence across diverse training tasks. All selected variables are divided into five categories: DDMs, map-related features, geometry-related parameters, receiver-related variables, and spacecraft identifiers. Table 1 lists each variable and describes its physical meaning. For example, we include variables crucial for deriving the total power of GNSS signal scattered near the specular direction with a coherent reflecting surface (Chew and Small, 2020):

$$P_r = \frac{P_t G_t}{4\pi (R_{ts} + R_{sr})^2} \frac{G_r \lambda^2}{4\pi} \Gamma_{\text{surface}}, \tag{1}$$

in which $P_t G_t$ is GPS effective isotropic radiated power (EIRP), G_r is the gain of transmitting antenna, λ is the wavelength of the GPS L1 signal, Γ_{surface} is CYGNSS derived surface reflectivity. R_{ts} and R_{sr} correspond to the distances between the transmitter and the specular point, and between the specular point and the receiver, respectively. Additionally, CYGNSS spacecraft numbers and GPS pseudorandom noise code are included to account for inter-satellite variances and GPS block type differences.

To consider seasonal variations while ensuring computational efficiency, we use four months of CYGNSS measurements with over three million samples, specifically from August and October 2019, and January and April 2020, for pretraining and downstream fine-tuning. To evaluate model performance, we select a test dataset with the same temporal distribution as the pretraining stage but without overlap, namely August and October 2020, and January and April 2021. For ocean wind speed retrieval, ground truth labels are obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis data (Hersbach et al., 2020), with further details available in Asgarimehr et al. (2022).

2.1.2. SMAP

Soil moisture active passive (SMAP) mission aims to measure and map Earth's surface soil moisture with global coverage every 2–3 days. In addition, VWC is an essential auxiliary variable from the data product. It is calculated with moderate resolution imaging spectroradiometer (MODIS)-derived normalized difference vegetation index (NDVI) values and stem factors (γ_s) associated with different land cover types (Chan et al., 2013), given by:

$$\text{VWC} = (1.9134 \times \text{NDVI}^2 - 0.3215 \times \text{NDVI}) + \gamma_s \times \frac{\text{NDVI}_{\max} - \text{NDVI}_{\min}}{1 - \text{NDVI}_{\min}}. \tag{2}$$

The NDVI derives from the red and near-infrared (NIR) spectral bands of the multispectral satellite mission as $\text{NDVI} = \frac{\text{NIR}-\text{Red}}{\text{NIR}+\text{Red}}$. For fine-tuning the proposed framework for terrestrial applications, we extract the ground truth variables, i.e., surface soil moisture and VWC, from SMAP enhanced L3 9 km equal-area scalable earth grid (EASE-Grid) version 5 product (O'Neill et al., 2021). The temporal distribution of pretraining and test datasets remains consistent with that of the CYGNSS data. For grid cells with multiple measurements due to SMAP ascending and descending overpasses, daily averaged values are assigned to the corresponding locations.

Table 1
DDMs and auxiliary parameters selected for model training, grouped by category and described by their physical meanings. For detailed definitions and data specifications, refer to the CYGNSS Handbook (Ruf et al., 2022).

Category	Variable	Description
Delay-Doppler Maps	raw_counts	DDM bin raw counts (uncalibrated power values)
	brcs	Bistatic radar cross-section of the DDM
	power_analog	Analog power corrected for quantization effects
	eff_scatter	DDM bin effective scattering area in m^2
Map-related	ddm_nbrcs	Normalized bistatic radar cross-section
	ddm_snr	Signal-to-noise ratio of the DDM
Geometry-related	sp_inc_angle	Incidence angle at the specular point
	sp_theta_body	Specular point body frame theta angle
	tx_pos_(x, y, z)	GPS spacecraft position X, Y, Z components
	tx_to_sp_range	Distance from transmitter to specular point
	rx_to_sp_range	Distance from receiver to specular point
Receiver-related	sp_rx_gain	Receiver antenna gain in specular direction
	gps_eirp	Effective isotropic radiated power of GPS signal
	gps_tx_power_db_w	Raw GPS transmit power in dBW
	gps_ant_gain_db_i	Gain of GPS antenna toward specular point
Spacecraft identifiers	prn_code	GPS spacecraft pseudorandom code index
	spacecraft_num	CYGNSS spacecraft number

2.2. Quality control

As self-supervised learning frameworks are data-driven, quality control is essential to minimize the propagation of erroneous signals and to ensure that the learned representations reflect geophysical variability rather than sensor artifacts or signal contamination. For CYGNSS data, observations are first filtered using empirical thresholds. We remove samples with an incidence angle larger than 70 degrees, receive antenna gain in the direction of the specular point and a direct signal-to-noise ratio (SNR) that are smaller than 0 dB. As the incidence angle increases, Fresnel reflectivity decreases, especially for land surfaces, thereby reducing the SNR and leading to higher retrieval uncertainty (Rodriguez-Alvarez et al., 2019). In addition, we discard measurements with DDM peak values falling outside a range of delay bins between 5 and 11, as such peaks typically correspond to high-altitude specular points where GNSS-R performance is known to degrade. Further quality control is applied using L1 product quality flags to exclude low-quality data, including "black-body DDM", "DDM in test pattern", "radio frequency interference (RFI) detected", "low confidence in GPS EIRP estimation", "S-band transmitter powered up", and "incorrect ddmi antenna selection". These conditions are known to introduce ambiguous signal characteristics and lead to unreliable training samples. For SMAP data, we remove samples flagged as "soil moisture retrieval was not successful" and "freeze/thaw state retrieval was not successful", as these indicate retrieval failures due to RFI, anomalous surface conditions, or unfavorable observation geometry to ensure high-quality inputs for representation learning (Senyurek et al., 2020a). Corroborated by previous studies (Zhao et al., 2023), quality control measures improve the reliability of input data for both pretraining and fine-tuning. By filtering low-quality or outlier data, we reduce noise in a reconstruction loss during pretraining and avoid label uncertainty in downstream retrieval tasks.

2.3. Collocation

After data cleaning, we prepare the datasets for pretraining and downstream fine-tuning with spatiotemporal alignment. Concretely, for each CYGNSS sample, daily averaged SMAP soil moisture and VWC values are collocated within a search window of 0.1 degrees in both longitude and latitude. This balances spatial proximity with the need for a sufficient amount of daily matched samples, given the variations in the CYGNSS glistening zone over land and ocean surfaces, where the spatial resolution is estimated to range from 3.5 km to 40 km. If multiple potential matches exist, the CYGNSS sample is assigned with the nearest neighboring SMAP measurement. Finally, we normalize the DDMs and auxiliary parameters to have zero mean and unit variance to stabilize gradients and facilitate model convergence during both pretraining and fine-tuning phases.

3. Methodology

3.1. Overall workflow

Existing GNSS-R deep learning models, e.g., CyGNSSnet (Asgarimehr et al., 2022) and DDM-Former, are mainly based on supervised learning with task-specific objectives. These methods require large labeled datasets for supervised training or are limited to single-modality inputs. In contrast, the proposed GEM framework aims to overcome these limitations by employing a self-supervised learning scheme to map multimodal GNSS-R data to a diverse set of Earth surface parameters. By leveraging the synergy between advanced representation learning and multimodal GNSS-R measurements, we aim to mitigate key bottlenecks, such as high annotation cost across various use cases, while enhancing existing products and catalyzing the development of further GNSS-R applications. Fig. 2 depicts the overall workflow of our approach, which consists of two stages: GEM pretraining and GEM fine-tuning.

During the pretraining stage, multimodal GNSS-R observables, i.e., DDMs as images and auxiliary data in tabular form, are simultaneously processed for representation learning on GNSS-R data. Specifically, four types of DDMs and essential auxiliary parameters are masked randomly and embedded through a set of encoders. Given that the feature representations of multimodal observables should be complementary to each other, the feature fusion module concerns the integration of different modalities into a joint feature space. The model is then trained to reconstruct the missing components through lightweight decoders. This allows the model to develop a meaningful understanding of the underlying distribution and correlations between different GNSS-R modalities.

During the fine-tuning stage, the pretrained model is fine-tuned on a limited amount of labeled GNSS-R data to adapt learned representations for various downstream tasks. Inputs and label pairs are carefully curated through quality control and collocation to ensure accurate and representative correlations between GNSS-R observables and the target retrieval parameters. While keeping the pretrained dual-branch encoder and feature fusion module frozen, the model's decoders are replaced with different sets of regression heads. At this stage, unmasked input DDMs and auxiliary parameters are fed into the model, enabling it to effectively map inputs to target parameters by leveraging prior knowledge from the learned representations.

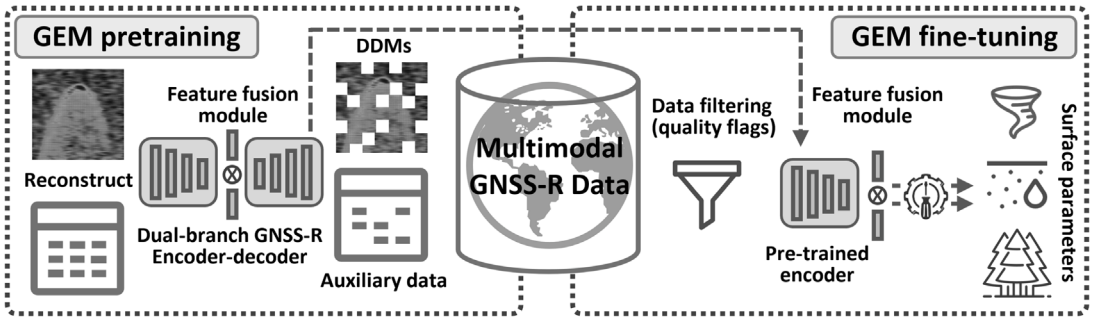


Fig. 2. Schematic view of the proposed GNSS-R generalist Earth surface monitor (GEM) framework. The framework consists of two stages: (1) self-supervised pretraining on multimodal GNSS-R data using a dual-branch encoder–decoder and feature fusion module, and (2) fine-tuning for downstream tasks such as ocean wind speed, soil moisture, and vegetation parameter retrieval.

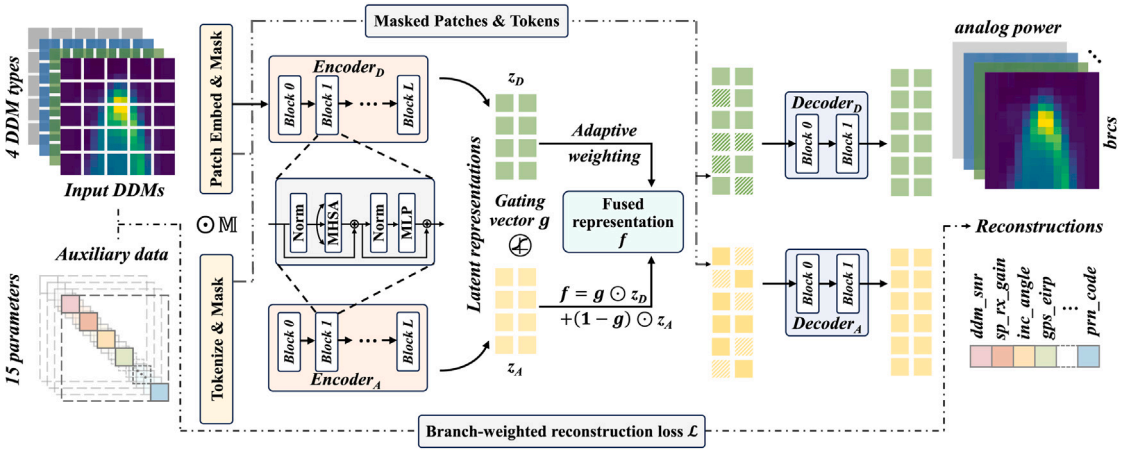


Fig. 3. Schematic of the proposed GEM pretraining for multimodal GNSS-R representation learning. Input GNSS-R observables are randomly masked and processed by dual-branch modality-specific encoders. The resulting latent representations are adaptively fused and used to reconstruct the masked inputs through lightweight Transformer-based decoders. The pretraining loss is computed as a branch-weighted sum of reconstruction errors.

3.2. GEM pretraining

In self-supervised learning for remote sensing, contrastive learning is a widely used approach that learns meaningful representations by maximizing the similarity between similar samples while pushing apart representations of different ones. However, its effectiveness is limited for multimodal GNSS-R data due to the highly ill-posed nature of GNSS-R measurements over ocean and land, where complex signal scattering and interactions among multiple factors introduce challenges. Moreover, contrastive learning methods are highly dependent on augmentations, but applying them to DDMs is challenging since traditional transformations do not align with the physical properties of GNSS-R signals. To overcome these issues, we propose GEM pretraining, a novel architecture for multimodal GNSS-R self-supervised learning that learns meaningful representations in a generative manner. Fig. 3 provides a schematic of the proposed pretraining process, which will be detailed in the following sections.

3.2.1. Patchify and masking

Let $\mathcal{X}^d$ and $\mathcal{X}^a$ be two archives associated with two different GNSS-R modalities, namely, DDMs and auxiliary data. Each archive $\mathcal{X}^j = \{\mathbf{x}_i^j\}_{i=1}^{N_j}\ \forall j \in \{d, a\}$ include N_j samples, in which $\boldsymbol{x}_i^j$ is the ith model input in the jth archive, and $(\boldsymbol{x}_i^d, \boldsymbol{x}_i^a)$ is the ith input pair that includes multimodal data from the same specular point. Given an input DDM $\boldsymbol{x}^d \in \mathbb{R}^{H\times W\times C}$, where H, W, and C represent height, width, and number of DDM types, respectively, we reshape and partition it into non-overlapping 3D tensor patches along delay and Doppler dimensions. Auxiliary data vector $\boldsymbol{x}^a \in \mathbb{R}^A$ with A as the count of auxiliary parameters is transformed to a diagonal matrix and aligned with the input DDMs to maintain balanced weights between two branches. Next, a fraction of patches are removed from both modalities by a random masking operation, e.g., for $\boldsymbol{x}_{\text{vis}}^d$ and $\boldsymbol{x}_{\text{mask}}^d$:

$$[\boldsymbol{x}_{\text{vis}}^d, \boldsymbol{x}_{\text{mask}}^d] = \mathbb{M} \odot \boldsymbol{x}^d\,, \tag{3}$$

in which $\mathbb{M} \in \{0,1\}^{\frac{H}{p}\times\frac{W}{p}\times C}$ with p as patch size is a set of indices denoting the masked patches. Given an unlabeled pretraining set $\mathcal{T} = \{(\boldsymbol{x}_i^d, \boldsymbol{x}_i^a)\}_{i=1}^N$, which include N input pairs, only the visible patches from both modalities are sent into the encoders for further processing.

3.2.2. Dual-branch encoder

In order to derive representative features from both modalities that contain complex correlations, representations from masked DDMs $\boldsymbol{x}_{\text{vis}}^d$ and auxiliary data $\boldsymbol{x}_{\text{vis}}^a$ are processed by modality-specific encoders $\boldsymbol{E}_d(\cdot)$ and $\boldsymbol{E}_a(\cdot)$. The produced latent representations are given by:

$$\boldsymbol{z}_i^d = \boldsymbol{E}_d(\boldsymbol{x}_{\text{vis}}^d)\,, \qquad \boldsymbol{z}_i^a = \boldsymbol{E}_a(\boldsymbol{x}_{\text{vis}}^a)\,. \tag{4}$$

Following the setup of MAE, for $\boldsymbol{x}_i^j$, the encoders embed only on the unmasked patches with positional embeddings, then processed via L layers of Transformer blocks. Given the relatively small size of DDMs (17×11 pixels) compared to natural images, we use $L = 4$ Transformer encoder layers. Each layer has a hidden embedding dimension of 128, with 4 attention heads and a feed-forward expansion ratio of 2 (i.e., intermediate feed-forward layers have a hidden size of 256). Layer normalization is applied before each multi-head self-attention and MLP sub-block. Since the encoders process only unmasked input patches, this enables training with large batch sizes and leads to faster convergence across multimodal GNSS-R inputs. These configurations allow

for effective modality-specific feature extraction while maintaining a balance between model complexity and training efficiency.

3.2.3. DA-fuse module

Our DA-Fuse module is designed to integrate multimodal latent representations from DDMs and auxiliary data through an adaptive weighted fusion mechanism. Instead of using fixed weights or simple concatenation, it learns to assign modality-specific importance to each branch based on its latent representations. The assumption is to enhance cross-modal interactions by ensuring that features from one modality complement those from the other while preserving modality-specific characteristics. Given the encoded latent representations from both modalities, $\boldsymbol{z}_i^d$ and $\boldsymbol{z}_i^a$, these are first concatenated and passed through a dense layer with sigmoid activation to generate a gating vector $\mathbf{g} \in [0, 1]$:

$$\mathbf{g} = \sigma\left(\mathbf{W}[\boldsymbol{z}_i^d; \boldsymbol{z}_i^a] + \mathbf{b}\right), \tag{5}$$

where $\sigma(\cdot)$ is the element-wise sigmoid function, and $\mathbf{W}$, $\mathbf{b}$ are learnable parameters. The fused representation $\boldsymbol{f}_i$ is computed as a convex combination of the two latent representations:

$$\boldsymbol{f}_i = \mathbf{g} \odot \boldsymbol{z}_i^d + (1 - \mathbf{g}) \odot \boldsymbol{z}_i^a. \tag{6}$$

This adaptive weighting allows the model to dynamically adjust the contribution of each modality, allowing their features to interact while keeping the dimensionality unchanged for decoders. As a result, the model can learn to prioritize modality-specific information and enhance cross-modal representation learning.

3.2.4. Multimodal decoders

The multimodal decoders are lightweight Transformer-based modules that operate only during the pretraining stage to reconstruct the masked patches. Each module consists of 2 Transformer decoder layers with 64 embedding dimension and 4 attention heads. These decoders process a complete set of patches that includes the fused latent representations and mask patches from each branch. Input DDMs and auxiliary data are reconstructed separately, given by:

$$\begin{aligned} &\hat{\boldsymbol{x}}_i^d = \boldsymbol{D}_d(\boldsymbol{f}_i), \qquad \hat{\boldsymbol{x}}_i^a = \boldsymbol{D}_a(\boldsymbol{f}_i), \\ &\widehat{\mathcal{T}} = \left\{(\hat{\boldsymbol{x}}_i^d, \hat{\boldsymbol{x}}_i^a)\right\}_{i=1}^N. \end{aligned} \tag{7}$$

where $\hat{\boldsymbol{x}}_i^d$ and $\hat{\boldsymbol{x}}_i^a$ are reconstructions of both modality, and $\widehat{\mathcal{T}}$ represents the reconstructed pretraining set.

3.2.5. Network learning

The pretraining stage aims to learn the representation of x_i^j by reconstructing its masked patches by minimizing the objective loss. Given the pretraining set $\mathcal{T}$ and its reconstruction $\widehat{\mathcal{T}}$, the loss function of GEM pretraining is calculated with a branch-weighted sum of reconstruction errors, given by:

$$\mathcal{L}_{\text{pretraining}} = \frac{1}{N} \sum_{i=1}^{N} \left[\lambda\, \mathcal{R}(\boldsymbol{x}_i^d, \hat{\boldsymbol{x}}_i^d) + (1 - \lambda)\, \mathcal{R}(\boldsymbol{x}_i^a, \hat{\boldsymbol{x}}_i^a)\right], \tag{8}$$

where $\mathcal{R}$ is the branch-wise mean squared error (MSE), computed only over masked regions. Furthermore, $\lambda \in [0, 1]$ regulates the significance of weights attributed to two branches, balancing the reconstruction contributions from the DDMs and auxiliary data. This learning strategy allows modality-specific processing while enabling interactions between modalities at the representation level. In addition, performing reconstructions only on masked patches forces the model to infer missing signals from the entire DDMs, in contrast to theoretical approaches that typically focus on peak values within the "3 × 5" grid set around the specular point DDM bin. Consequently, the model is capable of learning task-agnostic latent features that support flexible adaptation to downstream tasks without overfitting to a specific retrieval target.

3.3. Downstream fine-tuning

During fine-tuning on downstream tasks, the pretrained encoders and DA-Fuse module are frozen to preserve the learned cross-modal representations. For initial evaluation, we replace the decoders with a single fully connected layer and perform linear probing (LP) on different tasks. In addition, we employ partial fine-tuning (FT) by replacing the decoders with MLP sublayers as regression heads. Concretely, multimodal GNSS-R inputs are processed through the dual-branch encoder, integrated into joint latent representations, and then passed through regression heads to generate predictions for different tasks. These fine-tuning strategies assess the model's generalizability across various GNSS-R applications.

3.4. Implementation details and evaluation

For both pretraining and fine-tuning stages, each mini-batch contains 1024 training samples. During pretraining, random masks are generated on the fly of each batch with a mask ratio of 0.3. The weighting parameter λ is set at 0.5 by default to balance contributions from both modalities. AdamW optimizer is used with early-stop patience as 3 epochs. The learning rate is established with warmup cosine with a warmup epoch percentage of 15% at $2{\times}10^{-3}$ and $6{\times}10^{-4}$ for pretraining and fine-tuning, respectively. We use the Tensorflow platform (Abadi et al., 2015) to implement the proposed framework and train it on an NVIDIA RTX 3090 GPU. Evaluation is conducted on the curated test sets for three downstream tasks using root mean square error (RMSE), bias, mean absolute percentage error (MAPE), and the coefficient of determination (R^2 score) as evaluation metrics.

4. Results and discussion

The primary goal of this work is to provide pretrained models that efficiently enhance existing GNSS-R products and reduce the exploration cost for novel applications. In order to confirm the hypothesis that our framework is well-suited for GNSS-R Earth surface monitoring, we evaluate both the GEM (LP) and GEM (FT) against competitor models on several retrieval tasks. We present quantitative and qualitative analyses for global estimations, DDM reconstructions, as well as comparisons of training efficiency, label efficiency, and fusion strategies. Finally, we briefly discuss the remaining challenges of the current approach.

4.1. Performance comparison

To evaluate the effectiveness of the proposed GEM framework, we compare its performance against several widely adopted baseline models that have demonstrated strong performance across a diverse set of GNSS-R retrieval tasks. These baselines are selected to represent the most widely applied and recent deep learning-based approaches in GNSS-R remote sensing for both oceanic and terrestrial applications, including MLP, ResNet, ConvLSTM, CyGNSSnet, and Transformer models. MLPs provide lightweight yet competitive baselines for GNSS-R-based Earth surface parameter retrieval, as shown in prior studies (Chen et al., 2024; Pilikos et al., 2024; Li et al., 2024). ResNet and CyGNSSnet are well-established CNN-based backbones capable of capturing fine-grained delay-Doppler correlations by treating DDMs as 2D image inputs and have been successfully applied to various retrieval tasks (Xiao et al., 2024a; Du et al., 2024; Song et al., 2025). Moreover, Transformer-based model and hybrid deep learning methods such as ConvLSTM are included to extend the comparison (Zhao et al., 2023; Wang et al., 2024). While all models use DDMs as the primary input, the inclusion of auxiliary parameters differs significantly across studies. To ensure a fair comparison, we adopt standardized backbone configurations for each method and evaluate their performance using multimodal GNSS-R inputs across three representative downstream

Table 2
Root mean square error (RMSE) and bias comparison of different methods for ocean wind speed retrieval, surface soil moisture estimation, and vegetation water content prediction.

Method	Ocean wind speed		Surface soil moisture		Vegetation water content	
	RMSE (m/s)	Bias (m/s)	RMSE (m^3/m^3)	Bias (m^3/m^3)	RMSE (kg/m^2)	Bias (kg/m^2)
MLP	1.62	−0.38	0.090	0.005	2.95	0.08
ResNet	1.65	−0.27	0.105	**0.002**	3.06	0.14
ConvLSTM	1.74	−0.38	0.135	0.009	3.10	0.17
CyGNSSnet	**1.53**	−0.15	0.101	0.003	2.98	0.12
Transformer	1.83	−0.55	0.141	0.007	3.14	0.14
GEM (LP)	1.81	−0.23	0.147	0.010	3.18	0.16
GEM (FT)	1.55	**−0.10**	**0.088**	0.003	**2.91**	**0.06**

Table 3
Mean absolute percentage error (MAPE) and R^2 score comparison of different methods for ocean wind speed retrieval, surface soil moisture estimation, and vegetation water content prediction.

Method	Ocean wind speed		Surface soil moisture		Vegetation water content	
	MAPE (%)	R^2 score	MAPE (%)	R^2 score	MAPE (%)	R^2 score
MLP	18.0	0.57	21.2	**0.52**	26.3	0.65
ResNet	19.5	0.52	24.3	0.51	26.1	0.64
ConvLSTM	23.5	0.45	24.7	0.48	26.5	0.61
CyGNSSnet	17.2	**0.62**	23.6	**0.52**	26.2	0.66
Transformer	19.5	0.52	25.2	0.49	27.5	0.58
GEM (LP)	20.4	0.49	25.0	0.48	27.2	0.57
GEM (FT)	**16.8**	0.60	**19.3**	**0.52**	**25.5**	**0.67**

tasks: ocean wind speed retrieval, surface soil moisture estimation, and VWC prediction.

Retrieved results from the test data using different models are compared against their corresponding label sources. The RMSE and bias for all samples across four seasons are reported in Table 2. Utilizing DDMs and auxiliary parameters as model inputs, deep learning-based models achieve satisfactory performance for all three tasks, in particular for ocean wind speed retrieval. With minimal effort in partial fine-tuning, our GEM (FT) method demonstrates robust performance across both oceanic and terrestrial retrieval tasks and validates its cross-task transferability. For ocean wind speed retrieval, despite the strong performance of the task-specific model CyGNSSnet, the proposed method attains comparable RMSE with a lower bias. Given the long-tailed distribution of global wind speeds, where most winds fall within 4–10 m/s, models inevitably tend to fit moderate wind regimes and underestimate higher wind speeds. A lower bias in GEM (FT) suggests a better understanding of the underlying mapping from measurements to ground truth labels, attributed to effective representation learning during the pretraining stage.

For terrestrial applications, the proposed method also exhibits improved performance compared to other models. With the lowest RMSE of $0.088\,m^3/m^3$ for surface soil moisture estimation, and an RMSE of $2.91\,kg/m^2$ with minimal bias of $0.06\,kg/m^2$ for VWC prediction, our results confirm that a pretrained model incorporating prior knowledge of data distribution and cross-modal relationships can better capture correlations across diverse tasks. While recent deep learning-based terrestrial applications often require additional input sources to account for the complex signal attenuation, our framework achieves considerable improvements using only GNSS-R inputs. It is noteworthy that fine-tuning with MLP sublayers as regression heads is essential for learning the complex nonlinearities within the data and allowing accurate prediction of continuous values. Although GEM (LP) does not lead to strong performance, it still benefits from representation learning in the pretraining stage and achieves reasonably satisfactory results. Furthermore, given the Transformer-based method processes only DDMs in its original setup, its performance degradation suggests that the naive inclusion of auxiliary parameters is less effective for representation learning. The performance advantages of the GEM (FT) further demonstrate that careful design of dual-branch encoder and the feature fusion module can improve model's retrieval accuracy.

The statistical results for each model, including MAPE and R^2 score, are reported in Table 3. Consistent with the findings in Table 2, all evaluated metrics indicate a steady improvement in performance across different tasks for the proposed method. Essentially, GEM (FT) achieves an improved MAPE with a large margin, in particular for ocean wind speed retrieval and surface soil moisture estimation. These results highlight the flexibility of the GEM framework, which uses a unified architecture pretrained in a self-supervised manner to capture geophysically relevant patterns across oceanic and terrestrial surfaces. By jointly learning from DDMs and auxiliary GNSS-R parameters, the model enables robust cross-task generalization without retraining, which is well-suited for data-scarce conditions or rapidly changing environmental events.

One bottleneck of exploiting novel GNSS-R applications with deep learning is the dependence on large amounts of labeled data required by task-specific supervised learning approaches. Collecting and curating such data can be costly and time-consuming, particularly when scaling to global applications. To specifically assess the label efficiency of our proposed framework, we compare it against supervised learning models trained on limited labeled data across different proportions of annotations. Fig. 4 illustrates RMSE as a function of labeled sample count for different downstream tasks. The number of labeled samples, shown on a logarithmic scale, depicts model performance from low-resource scenarios ($\sim 10^3$ samples) to large-scale datasets ($\sim 10^6$ samples). Across all three tasks, our GEM framework consistently achieves lower RMSE regardless of labeled data sizes. In low-data regimes, supervised models struggle with high estimation errors, whereas the proposed method significantly outperforms them even when only limited labels are available. As the number of labeled samples increases, RMSE decreases for all models, but GEM (FT) maintains a clear performance advantage. These results emphasize the importance of GNSS-R pretraining for efficiently reducing label dependency, which is of great interest for exploring novel applications in domains where labeled data is limited.

To complement the assessment of label efficiency, Fig. 5 examines training efficiency by illustrating the relationship between the number of training steps and RMSE values across different tasks. Given a fixed number of training samples, GEM (FT) demonstrates rapid convergence to lower RMSE values, especially observed in the early training stages. While all models improve as training proceeds, the proposed framework requires significantly fewer steps to reach optimal performance. These results support the hypothesis that a model pretrained on a large volume of unlabeled GNSS-R observations retains useful prior knowledge, which reduces the optimization burden during downstream

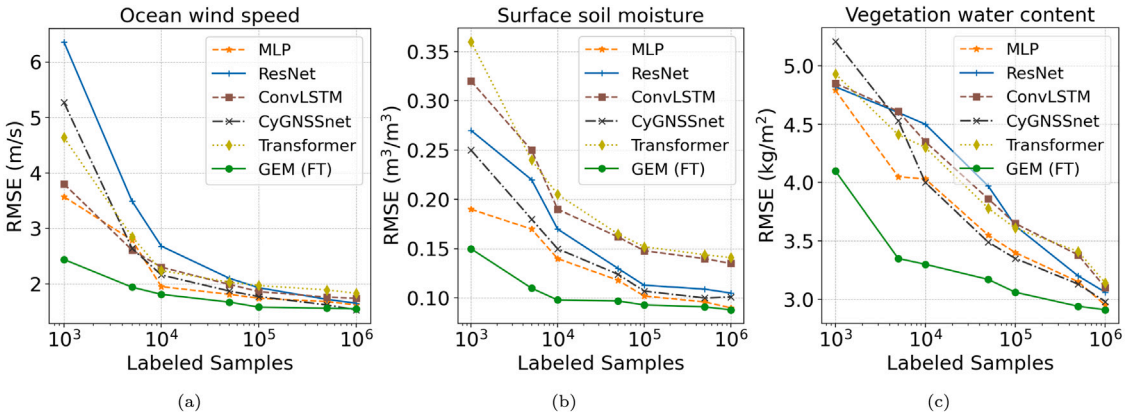


Fig. 4. Label efficiency comparison across three GNSS-R downstream tasks. The proposed GEM (FT) framework outperforms competing models and achieves better label efficiency across different proportions of annotations, especially in low-data regimes.

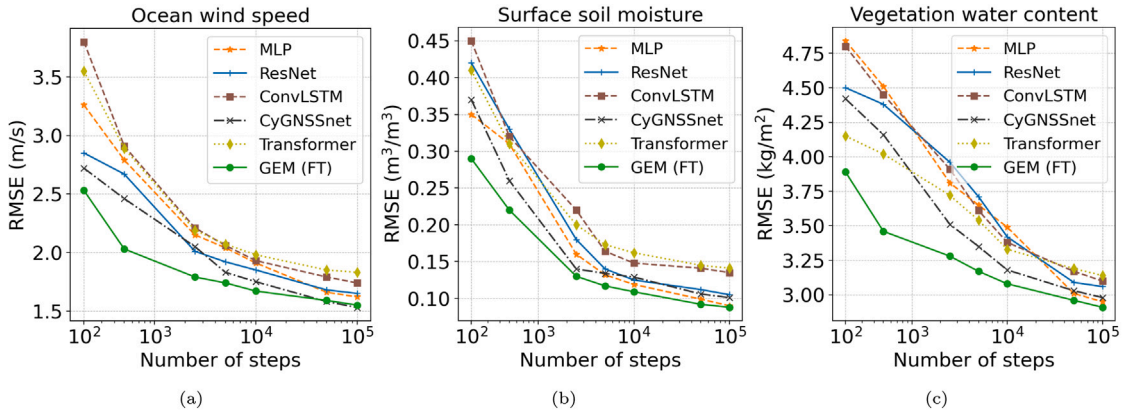


Fig. 5. Training efficiency comparison across three GNSS-R downstream tasks. With a one-time initial pretraining, GEM (FT) converges faster to lower RMSE values compared to other models which allows faster adaptation while reducing computational costs.

fine-tuning. Although pretraining introduces an initial computational cost, the resulting model is task-agnostic and can be reused across various retrieval tasks. This trade-off leads to a computationally efficient solution when deployed across multiple applications and emphasizes the value of representation learning on multimodal GNSS-R data for achieving cross-task transferability and rapid adaptation.

4.2. Geographical analysis

To enhance the understanding of the global distribution of estimations obtained from our GEM framework, Fig. 6 illustrates the average ground truth labels and predicted values for ocean wind speed, surface soil moisture, and VWC throughout the entire test period. With an average grid resolution of 1-degree, the qualitative comparison of global distributions indicates that the estimations align reasonably well with reference values. This demonstrates the strong generalization capability of the proposed pretraining framework for Earth surface monitoring at scale. For ocean wind speed retrieval, our method accurately captures both moderate and high wind regions (e.g., wind speed higher than 12 m/s). The middle and bottom figures also depict strong agreement between estimated and ground truth maps for terrestrial applications. In particular, in regions such as the Amazon Basin, Central Africa, and Southeast Asia, the model effectively captures distinct variations in soil moisture and VWC values.

The global RMSE and bias distributions for downstream tasks are shown in Fig. 7. The triptych of visualizations is calculated with a grid resolution of $1° \times 1°$. Most regions in Fig. 7(a) are covered with orange to green fields, indicating relatively low RMSE values (0.5–1.5 m/s) compared to the CYGNSS mission requirements of 2 m/s. While minor underestimations of strong winds appear in the middle Pacific and southern Indian Ocean, the overall agreement with ERA5 labels still holds. In terms of surface soil moisture estimation, the lowest RMSE and bias values can be found in arid and semi-arid regions, whereas the highest errors occur in mountainous regions and archipelagos areas with complex topographies. Likewise, higher estimation errors in Fig. 7(c) appear in regions with strong surface roughness variations, including the Himalayas, the east coast of Australia, and the Andes Mountains in South America. The increased estimation errors observed in these terrestrial regions are closely linked to spatial heterogeneity, which introduces variability in signal scattering. Mountainous areas exhibit large variations in elevation and slope within the GNSS-R footprint, resulting in non-uniform incidence angles and increased multipath effects. Similarly, transitions between dense vegetation and bare land (e.g., Central Africa or parts of Southeast Asia) result in spatially heterogeneous dielectric properties and surface roughness, which cause inconsistent signal attenuation and reflection patterns within the glistening zone. These factors disrupt the model's performance in regions with complex surface conditions and contribute to localized increases in retrieval uncertainty. Nonetheless, given the generalizability of the pretrained model, such issues could be mitigated by incorporating more training samples from these regions or integrating additional geophysical parameters during the pretraining stage. The geographical distributions further emphasize the feasibility of using pretrained GNSS-R models for large-scale Earth surface monitoring with improved spatiotemporal coverage.

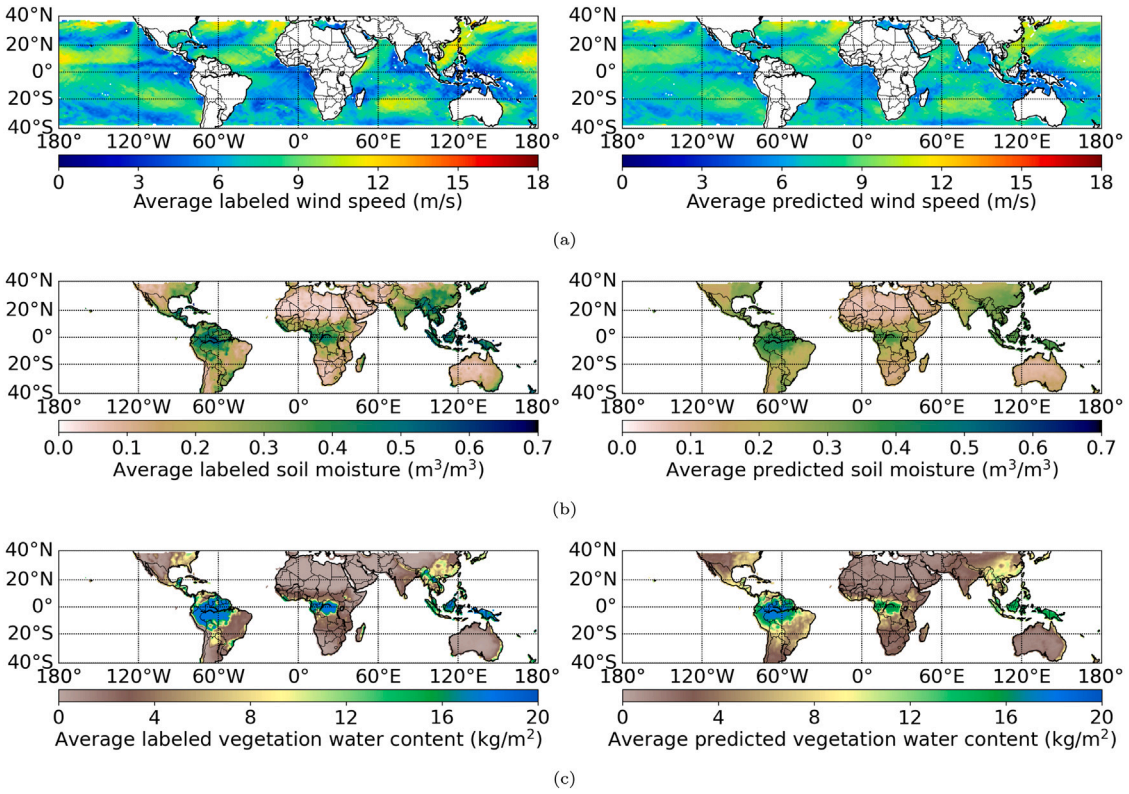


Fig. 6. Average ground truth and predicted spatial distributions of (a) ocean wind speed, (b) surface soil moisture, and (c) vegetation water content in $1° \times 1°$ grid resolution over the test dataset.

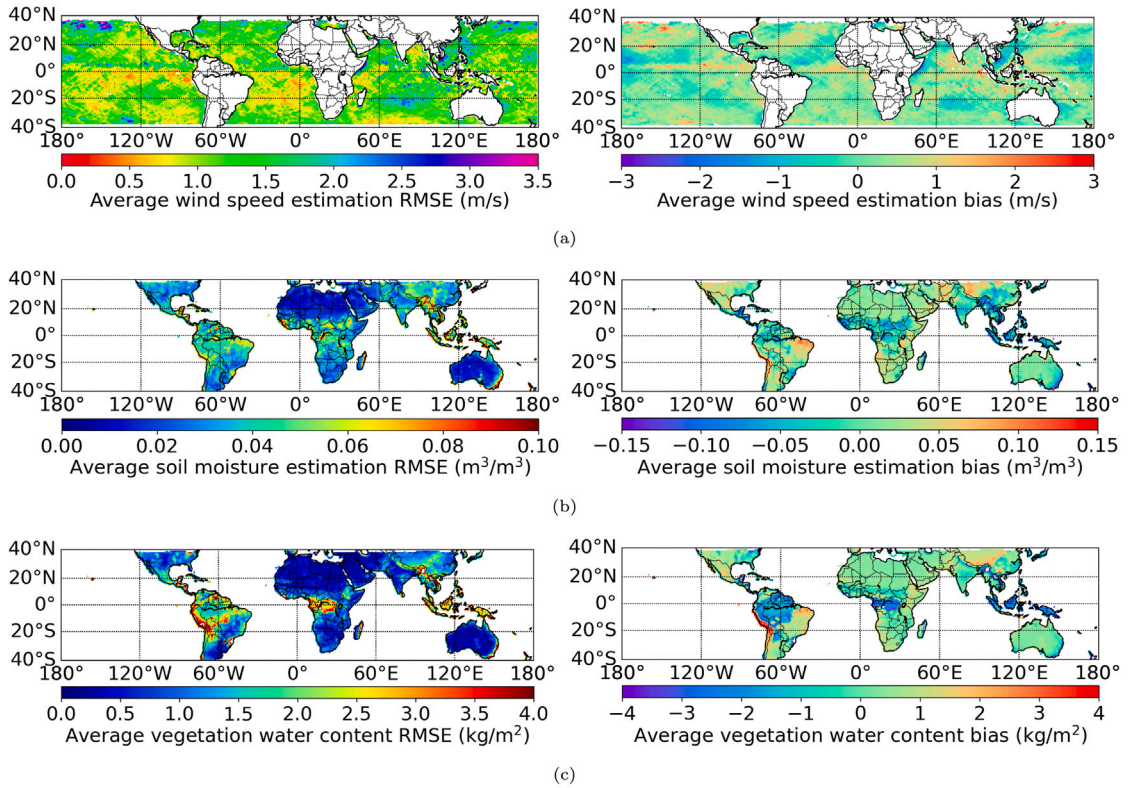


Fig. 7. Average RMSE and bias spatial distributions of (a) ocean wind speed, (b) surface soil moisture, and (c) vegetation water content in $1° \times 1°$ grid resolution over the test dataset.

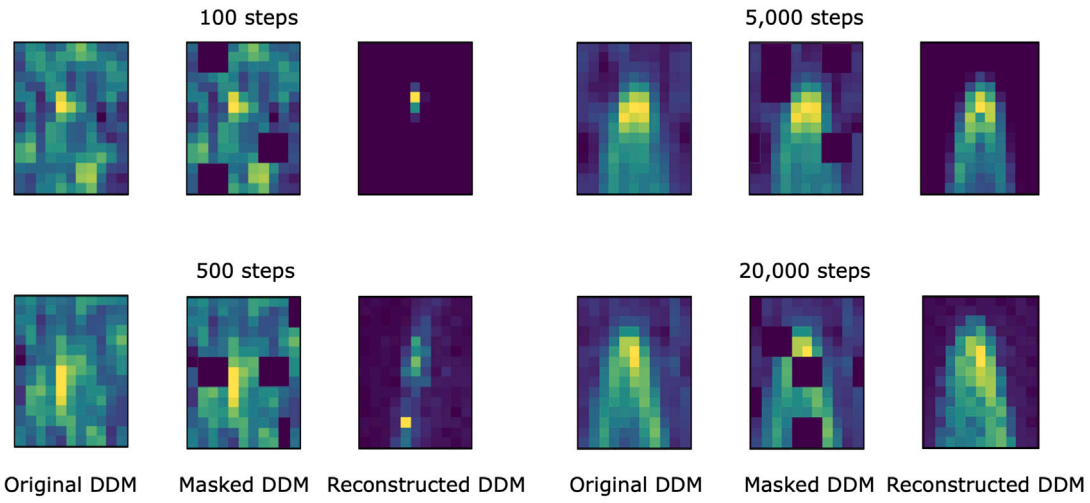


Fig. 8. Reconstruction of DDMs at different pretraining steps for the proposed GEM framework. From left to right at each step: original DDMs, masked DDMs with randomly occluded patches, and reconstructions based only on the visible portions.

Table 4
Comparison of RMSE and bias for different fusion strategies using multimodal GNSS-R data across downstream tasks.

Fusion method	Ocean wind speed		Surface soil moisture		Vegetation water content	
	RMSE (m/s)	Bias (m/s)	RMSE (m^3/m^3)	Bias (m^3/m^3)	RMSE (kg/m^2)	Bias (kg/m^2)
1 × 1 convolution	1.57	−0.08	**0.087**	0.003	2.92	**0.05**
Cross-attention	1.57	**−0.07**	0.090	**0.002**	2.93	0.06
DA-Fuse	**1.55**	−0.10	0.088	0.003	**2.91**	0.06

4.3. DDM reconstruction

A non-trivial aspect of self-supervised learning approaches is the difficulty of interpreting a model's representation learning process. Given the complexity of multimodal GNSS-R data, effectively learning useful representations for downstream applications is crucial. However, since the underlying mapping and decision-making processes are not easily explainable, it is challenging to determine whether the learned representations capture meaningful correlation patterns.

To gain insights into the representation learning process of our GEM framework, we visualize the reconstruction of DDMs during pretraining, as shown in Fig. 8. The BRCS of original input DDMs, randomly masked DDMs, and their reconstructions based only on the visible portions are presented at different training steps. It is obvious that at the early training stage, the reconstructed BRCS DDMs are highly incomplete. While the model initially learns to identify peak values in a DDM, which are essential for calculating normalized BRCS values, it struggles to capture meaningful patterns for reconstructing fine-grained cross-correlations. After 5,000 steps, some missing signals are still noticeable, but the overall alignment between the reconstructions and original inputs shows clear improvement. At the final stage of pretraining, the model effectively reconstructs fine-grained delay-Doppler correlations and signal intensity, which reflects an improved representation of learning outcomes.

These findings indicate that the proposed GEM framework allows the model to learn rich DDM representations in a self-supervised manner. During fine-tuning on downstream applications, the learned representations help to reduce the dependency on labeled data and improve performance. The progressively enhancement in reconstruction quality also suggests that the model captures the essential delay-Doppler correlation of GNSS-R observables, which contains valuable information for surface parameter retrieval over both ocean and land.

4.4. Ablation study on fusion strategies

To assess the impact of different fusion strategies on downstream performance, we conduct an ablation study comparing the DA-Fuse module with two alternative approaches: 1 × 1 convolution and cross-attention. The cross-attention method allows the model to selectively focus on relevant information from both modalities, while 1 × 1 convolution integrates both modalities through a learnable channel-wise projection in a simple yet effective manner. Table 4 summarizes the RMSE and bias obtained using these fusion methods on multimodal GNSS-R data across three downstream tasks. Although both alternative methods perform competitively, the proposed method shows modest improvements in RMSE for ocean wind speed retrieval and VWC prediction. Similar to the design of our pretraining loss, the DA-Fuse module assigns weights to each modality but does so more dynamically through adaptive fusion during pretraining. This alignment between model architecture and learning objective contributes to effective representation learning. The consistent performance of the DA-Fuse module across tasks suggests that adaptively balancing modalities during fusion provides an applicable benefit.

4.5. Limitations

While our work demonstrates the feasibility of using a pretrained GNSS-R model as a versatile Earth surface monitor, there are still challenges and limitations that need to be addressed. One notable issue is the degraded performance in terrestrial applications compared to oceanic tasks such as wind speed estimation. This is primarily due to the complex and spatially heterogeneous scattering behavior over land, where a mixture of coherent and incoherent scattering arises from vegetation, topography, and land cover variability. Unlike ocean surfaces, which provide relatively homogeneous reflections, land surfaces introduce signal decorrelation due to surface roughness and diverse dielectric properties, making it relatively challenging to extract consistent features from GNSS-R-only measurements. Additionally, the impact of vegetation canopy and soil characteristics often results in nonlinear and location-dependent correlations between GNSS-R observables and retrieval targets. While our GEM framework incorporates auxiliary parameters from GNSS-R missions to help mitigate these effects, the current auxiliary feature set may be insufficient to fully

compensate for complex land interactions. Corroborated by previous task-specific studies, incorporating additional parameters such as soil state, topography, and land use from other resources could further improve estimation precision (Nabi et al., 2023; Santi et al., 2024). Understanding how to effectively combine GNSS-R observables with other geophysical indicator sources in a unified self-supervised learning framework is critical for further improving terrestrial performance.

Another issue is related to the effective spatial resolution of GNSS-R observations, which is coarser than that of optical remote sensing sensors due to the large and diffuse nature of the scattering area. As a signals-of-opportunity bistatic radar, the spatial resolution of GNSS-R measurements is governed by the size of the first Fresnel zone, which varies with incidence angle and orbital geometry. Over land, scattered signals are more complex due to mixed coherent and incoherent scattering, further altering the footprint and increasing measurement uncertainty. Future studies could explore interpolation strategies using multi-source GNSS-R data or integrate observations from other satellite missions to enable data fusion, fill measurement gaps, and enhance the spatiotemporal coverage of global monitoring. Despite these limitations, our results demonstrate that the GEM framework serves as an initial cornerstone for developing cross-task transferable models in GNSS-R applications. These findings highlight the importance of developing flexible, multimodal pretraining approaches that fit the unique signal characteristics of GNSS-R data across land and ocean surfaces.

5. Conclusion

Monitoring Earth's surface with increased spatiotemporal coverage is critical for climate action. GNSS-R offers new potential to provide rapid global measurements with unprecedented spatiotemporal coverage. An increasing amount of GNSS-R satellite constellations may help to efficiently build deep learning models for surface property monitoring. However, existing data-driven methods are typically designed for task-specific applications using supervised learning, which requires large annotated datasets and limits the exploration of GNSS-R applications.

In this work, we introduced the first multimodal GNSS-R self-supervised learning framework fusing DDMs with auxiliary parametric data. Across three downstream tasks, namely, ocean wind speed retrieval, surface soil moisture estimation, and VWC prediction, the proposed model demonstrates decent performance in retrieving surface parameters. While acknowledging the computational demands of pretraining, our findings highlight the potential of pretrained models to advance GNSS-R applications by reducing the need for extensive feature engineering, fine-tuning, and downstream task annotations.

We hope this work will encourage further research in GNSS-R self-supervised learning and foundation models for Earth surface parameter monitoring, as these approaches could lead to cost-efficient and scalable GNSS-R applications. Future studies could incorporate more diverse data sources to enhance estimation accuracy, address spatial resolution gaps, and explore more advanced designs for GNSS-R foundation models to further improve efficiency and expand applications to novel GNSS-R use cases.

CRediT authorship contribution statement

Daixin Zhao: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Konrad Heidler:** Writing – review & editing, Visualization, Supervision, Methodology. **Milad Asgarimehr:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition. **Conrad M. Albrecht:** Writing – review & editing, Visualization, Methodology. **Jens Wickert:** Writing – review & editing, Supervision, Resources, Project administration. **Xiao Xiang Zhu:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Lichao Mou:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is jointly supported by the Hermann von Helmholtz-Gemeinschaft Deutscher Forschungszentren e.V. project "Artificial Intelligence for GNSS Reflectometry: Novel Remote Sensing of Ocean and Atmosphere (AI4GNSSR)" [grant number: ZT-I-PF-5-091], by the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" [grant number: 01DD20001], and by the German Federal Ministry of Economics and Technology in the framework of the "national center of excellence ML4Earth" [grant number: 50EE2201C]. We would like to thank the scientific teams associated with the CYGNSS and SMAP mission at NASA and the University of Michigan. Also, the authors are grateful to Yi Wang, Tianqi Xiao, Caroline Arnold, Anja Rösel, and Andrés Camero for their valuable discussions.

Data availability

Data will be made available on request.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems [software]. Zenodo. http://dx.doi.org/10.5281/zenodo.4724125.

Asgarimehr, M., Arnold, C., Weigel, T., Ruf, C., Wickert, J., 2022. GNSS reflectometry global ocean wind speed using deep learning: Development and assessment of CyGNSSnet. Remote Sens. Environ. 269, 112801. http://dx.doi.org/10.1016/J.RSE.2021.112801.

Asgarimehr, M., Zhelavskaya, I., Foti, G., Reich, S., Wickert, J., 2020. A GNSS-R geophysical model function: Machine learning for wind speed retrievals. IEEE Geosci. Remote. Sens. Lett. 17 (8), 1333–1337. http://dx.doi.org/10.1109/LGRS.2019.2948566.

Balsamo, G., Agusti-Panareda, A., Albergel, C., Arduini, G., Beljaars, A., Bidlot, J., Blyth, E., Bousserez, N., Boussetta, S., Brown, A., Buizza, R., Buontempo, C., Chevallier, F., Choulga, M., Cloke, H., Cronin, M.F., Dahoui, M., De Rosnay, P., Dirmeyer, P.A., Drusch, M., Dutra, E., Ek, M.B., Gentine, P., Hewitt, H., Keeley, S.P., Kerr, Y., Kumar, S., Lupu, C., Mahfouf, J.F., McNorton, J., Mecklenburg, S., Mogensen, K., Muñoz-Sabater, J., Orth, R., Rabier, F., Reichle, R., Ruston, B., Pappenberger, F., Sandu, I., Seneviratne, S.I., Tietsche, S., Trigo, I.F., Uijlenhoet, R., Wedi, N., Woolway, R.I., Zeng, X., 2018. Satellite and in situ observations for advancing global earth surface modelling: A review. Remote. Sens. 10 (12), 2038. http://dx.doi.org/10.3390/rs10122038.

Chan, S., Bindlish, R., Hunt, R., Jackson, T., Kimball, J., 2013. Ancillary Data Report: Vegetation Water Content. Technical Report, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, SMAP Science Document No. 047.

Chen, F., Guo, F., Liu, L., Nan, Y., 2021. An improved method for pan-tropical aboveground biomass and canopy height retrieval using CYGNSS. Remote. Sens. 13 (13), 2491. http://dx.doi.org/10.3390/rs13132491.

Chen, F., Liu, L., Guo, F., Huang, L., 2024. A new vegetation observable derived from spaceborne GNSS-R and its application to vegetation water content retrieval. Remote. Sens. 16 (5), 931. http://dx.doi.org/10.3390/rs16050931.

Chew, C., Small, E., 2020. Estimating inundation extent using CYGNSS data: A conceptual modeling study. Remote Sens. Environ. 246, 111869. http://dx.doi.org/10.1016/j.rse.2020.111869.

Cong, Y., Khanna, S., Meng, C., Liu, P., Rozi, E., He, Y., Burke, M., Lobell, D., Ermon, S., 2022. SatMAE: Pre-training transformers for temporal and multi-spectral satellite imagery. In: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (Eds.), In: Adv. Neural Inf. Process. Syst., vol. 35, Curran Associates, Inc., pp. 197–211.

CYGNSS, 2021. CYGNSS Level 1 Science Data Record Version 3.1. NASA Physical Oceanography Distributed Active Archive Center (PO.DAAC), CA, USA, http://dx.doi.org/10.5067/CYGNS-L1X31, Ver. 3.1. Dataset (Accessed 01 June 2023).

Dielacher, A., Fragner, H., Koudelka, O., 2022. PRETTY - passive GNSS-reflectometry for CubeSats. E+I Elektrotech. Infotech. 139 (1), 25–32. http://dx.doi.org/10.1007/s00502-022-00993-7.

Du, H., Li, W., Cardellach, E., Ribó, S., Rius, A., Nan, Y., 2024. Deep residual fully connected network for GNSS-R wind speed retrieval and its interpretation. Remote Sens. Environ. 313, 114375. http://dx.doi.org/10.1016/j.rse.2024.114375.

Eroglu, O., Kurum, M., Boyd, D., Gurbuz, A.C., 2019. High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks. Remote. Sens. 11 (19), 2272. http://dx.doi.org/10.3390/rs11192272.

Guo, W., Du, H., Guo, C., Southwell, B.J., Cheong, J.W., Dempster, A.G., 2022. Information fusion for GNSS-R wind speed retrieval using statistically modified convolutional neural network. Remote Sens. Environ. 272, 112934. http://dx.doi.org/10.1016/J.RSE.2022.112934.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R., 2022. Masked autoencoders are scalable vision learners. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.. pp. 16000–16009. http://dx.doi.org/10.1109/CVPR52688.2022.01553.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146 (730), 1999–2049. http://dx.doi.org/10.1002/QJ.3803.

Hong, D., Zhang, B., Li, X., Li, Y., Li, C., Yao, J., Yokoya, N., Li, H., Ghamisi, P., Jia, X., Plaza, A., Gamba, P., Benediktsson, J.A., Chanussot, J., 2024. SpectralGPT: Spectral remote sensing foundation model. IEEE Trans. Pattern Anal. Mach. Intell. 46 (8), 5227–5244. http://dx.doi.org/10.1109/tpami.2024.3362475.

Jakubik, J., Roy, S., Phillips, C.E., Fraccaro, P., Godwin, D., Zadrozny, B., Szwarcman, D., Gomes, C., Nyirjesy, G., Edwards, B., Kimura, D., Simumba, N., Chu, L., Mukkavilli, S.K., Lambhate, D., Das, K., Bangalore, R., Oliveira, D., Muszynski, M., Ankur, K., Ramasubramanian, M., Gurung, I., Khallaghi, S., Li, H.S., Cecil, M., Ahmadi, M., Kordi, F., Alemohammad, H., Maskey, M., Ganti, R., Weldemariam, K., Ramachandran, R., 2023. Foundation models for generalist geospatial artificial intelligence. http://dx.doi.org/10.48550/arXiv.2310.18660, arXiv Preprint arXiv:2310.18660.

Jales, P., Esterhuizen, S., Masters, D., Nguyen, V., Nogués-Correig, O., Yuasa, T., Cartwright, J., 2020. The new spire GNSS-R satellite missions and products. In: Proc. SPIE, Image Signal Process. Remote Sens., vol. 11533, p. 115330H. http://dx.doi.org/10.1117/12.2574127.

Jia, Y., Jin, S., Savi, P., Yan, Q., Li, W., 2020. Modeling and theoretical analysis of GNSS-R soil moisture retrieval based on the random forest and support vector machine learning approach. Remote. Sens. 12 (22), 3679. http://dx.doi.org/10.3390/rs12223679.

Kossieris, S., Asgarimehr, M., Wickert, J., 2023. Unsupervised machine learning for GNSS reflectometry inland water body detection. Remote. Sens. 15 (12), 3206. http://dx.doi.org/10.3390/rs15123206.

Lei, F., Senyurek, V., Kurum, M., Gurbuz, A.C., Boyd, D., Moorhead, R., Crow, W.T., Eroglu, O., 2022. Quasi-global machine learning-based soil moisture estimates at high spatio-temporal scales using CYGNSS and SMAP observations. Remote Sens. Environ. 276, 113041. http://dx.doi.org/10.1016/j.rse.2022.113041.

Li, Z., Guo, F., Zhang, X., Zhang, Z., Zhu, Y., Yang, W., Wu, Z., Yue, L., 2024. Integrating spaceborne GNSS-R and SMOS for sea surface salinity retrieval using artificial neural network. GPS Solut. 28 (4), 162. http://dx.doi.org/10.1007/s10291-024-01709-4.

Li, X., Yang, D., Yang, J., Zheng, G., Han, G., Nan, Y., Li, W., 2021. Analysis of coastal wind speed retrieval from CYGNSS mission using artificial neural network. Remote Sens. Environ. 260, 112454. http://dx.doi.org/10.1016/J.RSE.2021.112454.

Liu, X., Bai, W., Tan, G., Huang, F., Xia, J., Yin, C., Sun, Y., Du, Q., Meng, X., Liu, C., Hu, P., 2023. GNSS-R global sea surface wind speed retrieval based on deep learning. IEEE Trans. Geosci. Remote Sens. 61, 1–15. http://dx.doi.org/10.1109/TGRS.2023.3309690.

Liu, Y., Collett, I., Morton, Y.J., 2019. Application of neural network to GNSS-R wind speed retrieval. IEEE Trans. Geosci. Remote Sens. 57 (12), 9756–9766. http://dx.doi.org/10.1109/TGRS.2019.2929002.

Masters, D.S., Roberts, M., Chew, C., Lowe, S., Tan, L., Colwell, I., McCleese, D., Ruf, C.S., 2023. The muon space signals of opportunity small satellite constellation and its role in sustainable earth observations. In: AGU Fall Meeting 2023.

Nabi, M.M., Senyurek, V., Lei, F., Kurum, M., Gurbuz, A.C., 2023. Quasi-global assessment of deep learning-based CYGNSS soil moisture retrieval. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 16, 5629–5644. http://dx.doi.org/10.1109/jstars.2023.3287591.

Nguyen, H.H., Kim, H., Crow, W., Yueh, S., Wagner, W., Lei, F., Wigneron, J.-P., Colliander, A., Frappart, F., 2025. From theory to hydrological practice: Leveraging CYGNSS data over seven years for advanced soil moisture monitoring. Remote Sens. Environ. 316, 114509. http://dx.doi.org/10.1016/j.rse.2024.114509.

O'Neill, P., Chan, S., Njoku, E., Jackson, T., Bindlish, R., Chaubell, J., Colliander, A., 2021. SMAP Enhanced L3 Radiometer Global and Polar Grid Daily 9 km EASE-Grid Soil Moisture, Version 5.. NASA National Snow and Ice Data Center Distributed Active Archive Center, http://dx.doi.org/10.5067/4DQ54OUIJ9DL, (Accessed 06 January 2023).

Pierdicca, N., Comite, D., Camps, A., Carreno-Luengo, H., Cenci, L., Clarizia, M.P., Costantini, F., Dente, L., Guerriero, L., Mollfulleda, A., Paloscia, S., Park, H., Santi, E., Zribi, M., Floury, N., 2022. The potential of spaceborne GNSS reflectometry for soil moisture, biomass, and freeze–thaw monitoring: Summary of a European space agency-funded study. IEEE Geosci. Remote. Sens. Mag. 10 (2), 8–38. http://dx.doi.org/10.1109/MGRS.2021.3115448.

Pilikos, G., Clarizia, M.P., Floury, N., 2024. Biomass estimation with GNSS reflectometry using a deep learning retrieval model. Remote. Sens. 16 (7), 1125. http://dx.doi.org/10.3390/rs16071125.

Prexl, J., Schmitt, M., 2024. SenPa-MAE: Sensor parameter aware masked autoencoder for multi-satellite self-supervised pretraining. arXiv:2408.11000, http://dx.doi.org/10.48550/arXiv.2408.11000, arXiv Preprint arXiv:2408.11000.

Reed, C.J., Gupta, R., Li, S., Brockman, S., Funk, C., Clipp, B., Keutzer, K., Candido, S., Uyttendaele, M., Darrell, T., 2023. Scale-MAE: A scale-aware masked autoencoder for multiscale geospatial representation learning. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. pp. 4088–4099. http://dx.doi.org/10.1109/ICCV51070.2023.00378.

Reynolds, J., Clarizia, M.P., Santi, E., 2020. Wind speed estimation from CYGNSS using artificial neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 13, 708–716. http://dx.doi.org/10.1109/JSTARS.2020.2968156.

Roberts, T.M., Colwell, I., Chew, C., Lowe, S., Shah, R., 2022. A deep-learning approach to soil moisture estimation with GNSS-R. Remote. Sens. 14 (14), 3299. http://dx.doi.org/10.3390/rs14143299.

Rodriguez-Alvarez, N., Podest, E., Jensen, K., McDonald, K.C., 2019. Classifying inundation in a tropical wetlands complex with GNSS-R. Remote. Sens. 11 (9), 1053. http://dx.doi.org/10.3390/rs11091053.

Ruf, C.S., Chew, C., Lang, T., Morris, M.G., Nave, K., Ridley, A., Balasubramaniam, R., 2018. A new paradigm in earth environmental monitoring with the CYGNSS small satellite constellation. Sci. Rep. 8 (1), 1–13. http://dx.doi.org/10.1038/S41598-018-27127-4.

Ruf, C., Gleason, S., Ridley, A., Rose, R., Scherrer, J., 2017. The nasa cygnss mission: Overview and status update. In: 2017 IEEE International Geoscience and Remote Sensing Symposium. IGARSS, pp. 2641–2643. http://dx.doi.org/10.1109/IGARSS.2017.8127537.

Ruf, C., Mckague, D., Posselt, D., Gleason, S., Clarizia, M.P., Zavorotny, V., Butler, T., Redfern, J., Wells, W., Morris, M., Crespo, J., Chew, C., Small, E., Pasqual, D., Wang, T., Warnock, A., Mayers, D., Al-Khaldi, M., O'Brien, A.J., 2022. CYGNSS Handbook. Michigan Publishing Services, Ann Arbor, MI, http://dx.doi.org/10.3998/mpub.12741920.

Santi, E., Clarizia, M., Comite, D., Dente, L., Guerriero, L., Pierdicca, N., 2022. Detecting fire disturbances in forests by using GNSS reflectometry and machine learning: A case study in Angola. Remote Sens. Environ. 270, 112878. http://dx.doi.org/10.1016/j.rse.2021.112878.

Santi, E., Comite, D., Dente, L., Guerriero, L., Pierdicca, N., Clarizia, M.P., Floury, N., 2024. Global soil moisture mapping at 5 km by combining GNSS reflectometry and machine learning in view of HydroGNSS. Sci. Remote. Sens. 10, 100177. http://dx.doi.org/10.1016/j.srs.2024.100177.

Santi, E., Paloscia, S., Pettinato, S., Fontanelli, G., Clarizia, M.P., Comite, D., Dente, L., Guerriero, L., Pierdicca, N., Floury, N., 2020. Remote sensing of forest biomass using GNSS reflectometry. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 13, 2351–2368. http://dx.doi.org/10.1109/JSTARS.2020.2982993.

Senyurek, V., Lei, F., Boyd, D., Gurbuz, A.C., Kurum, M., Moorhead, R., 2020a. Evaluations of machine learning-based CYGNSS soil moisture estimates against SMAP observations. Remote. Sens. 12 (21), 3503. http://dx.doi.org/10.3390/rs12213503.

Senyurek, V., Lei, F., Boyd, D., Kurum, M., Gurbuz, A.C., Moorhead, R., 2020b. Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS. Remote. Sens. 12 (7), 1168. http://dx.doi.org/10.3390/rs12071168.

Song, S., Zhu, Y., Qu, X., Tao, T., 2025. Spaceborne GNSS-R for sensing soil moisture using CYGNSS considering land cover type. Water Resour. Manag. 1–21. http://dx.doi.org/10.1007/s11269-025-04119-4.

Unwin, M.J., Pierdicca, N., Cardellach, E., Rautiainen, K., Foti, G., Blunt, P., Guerriero, L., Santi, E., Tossaint, M., 2021. An introduction to the HydroGNSS GNSS reflectometry remote sensing mission. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 14, 6987–6999. http://dx.doi.org/10.1109/JSTARS.2021.3089550.

Wang, Y., Albrecht, C.M., Braham, N.A.A., Mou, L., Zhu, X.X., 2022. Self-supervised learning in remote sensing: A review. IEEE Geosci. Remote. Sens. Mag. 10 (4), 213–247. http://dx.doi.org/10.1109/MGRS.2022.3198244.

Wang, Q., Bu, J., Ni, J., Li, L., Liu, X., Huang, W., 2024. Ocean swell height estimation from spaceborne GNSS-R data using hybrid deep learning model. GPS Solut. 28 (4), http://dx.doi.org/10.1007/s10291-024-01698-4.

Wang, Y., Hernández, H.H., Albrecht, C.M., Zhu, X.X., 2025. Feature guided masked autoencoder for self-supervised learning in remote sensing. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 18, 321–336. http://dx.doi.org/10.1109/JSTARS.2024.3493237.

Winkelried, J., Ruf, C., Gleason, S., 2023. Spatial and temporal sampling properties of a large GNSS-R satellite constellation. Remote. Sens. 15 (2), 333. http://dx.doi.org/10.3390/rs15020333.

Xiao, T., Arnold, C., Zhao, D., Mou, L., Wickert, J., Asgarimehr, M., 2024a. Deep learning in spaceborne GNSS reflectometry: Correcting precipitation effects on wind speed products. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 17, 17860–17875. http://dx.doi.org/10.1109/JSTARS.2024.3453999.

Xiao, T., Asgarimehr, M., Wickert, J., Zhao, D., Mou, L., Arnold, C., 2024b. Evaluating feature impact on ocean wind speed predictions: An application of explainable AI to GNSS reflectometry data. In: IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium. pp. 1854–1858. http://dx.doi.org/10.1109/IGARSS53475.2024.10642477.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716. http://dx.doi.org/10.1016/j.rse.2020.111716.

Yueh, S.H., Shah, R., Chaubell, M.J., Hayashi, A., Xu, X., Colliander, A., 2022. A semiempirical modeling of soil moisture, vegetation, and surface roughness impact on CYGNSS reflectometry data. IEEE Trans. Geosci. Remote Sens. 60, 1–17. http://dx.doi.org/10.1109/TGRS.2020.3035989.

Zhang, Y., Bu, J., Zuo, X., Yu, K., Wang, Q., Huang, W., 2024a. Vegetation water content retrieval from spaceborne GNSS-R and multi-source remote sensing data using ensemble machine learning methods. Remote. Sens. 16 (15), 2793. http://dx.doi.org/10.3390/rs16152793.

Zhang, L., Zhao, Y., Dong, R., Zhang, J., Yuan, S., Cao, S., Chen, M., Zheng, J., Li, W., Liu, W., Zhang, W., Feng, L., Fu, H., 2024b. A^2-MAE: A spatial-temporal-spectral unified remote sensing pre-training method based on anchor-aware masked autoencoder. http://dx.doi.org/10.48550/arXiv.2406.08079, arXiv Preprint arXiv:2406.08079.

Zhao, D., Asgarimehr, M., Heidler, K., Xiao, T., Wickert, J., Arnold, C., Zhu, X.X., Mou, L., 2024. AI for GNSS reflectometry: Empowering earth surface monitoring from space. In: AGU24 Annual Meeting. Authorea, Inc., http://dx.doi.org/10.22541/essoar.173499043.34319619/v1.

Zhao, D., Heidler, K., Asgarimehr, M., Arnold, C., Xiao, T., Wickert, J., Zhu, X.X., Mou, L., 2023. DDM-Former: Transformer networks for GNSS reflectometry global ocean wind speed estimation. Remote Sens. Environ. 294, 113629. http://dx.doi.org/10.1016/j.rse.2023.113629.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geosci. Remote. Sens. Mag. 5 (4), 8–36. http://dx.doi.org/10.1109/MGRS.2017.2762307.

Zhu, X.X., Xiong, Z., Wang, Y., Stewart, A.J., Heidler, K., Wang, Y., Yuan, Z., Dujardin, T., Xu, Q., Shi, Y., 2024. On the foundations of earth and climate foundation models. http://dx.doi.org/10.48550/arXiv.2405.04285, arXiv Preprint arXiv:2405.04285.