Best Linear Unbiased Estimators for Fusion of Multiple CYGNSS Soil Moisture Products

M M Nabi, Volkan Senyurek, Mehmet Kurum, *Senior Member, IEEE*, and Ali Cafer Gurbuz, *Senior Member, IEEE*

***Abstract*—NASA's Cyclone Global Navigation Satellite System (CYGNSS) mission has gained significant attention within the land remote sensing community for estimating soil moisture (SM) using the Global Navigation System Reflectometry technique. Multiple algorithms have been developed to generate global SM data products from CYGNSS observations in combination with other remotely sensed geophysical data products. However, different algorithms exhibit variations in performance concerning both time and space due to model capabilities, complexities, and loss calculations. To address these limitations, the fusion of various SM products can be an effective solution. In this study, we explore different fusion algorithms, including the minimum variance estimator, best linear unbiased estimator, and linear weight fusion, to fuse distinct global CYGNSS-based SM products. We consider three SM data products publicly available from the Geosystems Research Institute at Mississippi State University. To assess our model's performance, we compare our fused data product with the Soil Moisture Active Passive (SMAP) mission's enhanced SM products at a resolution of 9 km × 9 km. Our findings reveal notable performance enhancements in several regions when combining different SM data products. The results demonstrate that the minimum variance estimator achieves a mean unbiased root-mean-square difference of 0.0359 m^3/m^3 with a correlation coefficient of 0.91 for SMAP-recommended grids and also linear weight fusion achieves 0.0389 m^3/m^3 with a correlation coefficient of 0.90 when no label data are used in the training of fusion.**

***Index Terms*—Best linear unbiased estimator (BLUE), Cyclone Global Navigation Satellite System (CYGNSS), linear weight fusion (LWF), minimum variance estimator (MVE), soil moisture (SM) fusion, triple collocation (TC).**

I. Introduction

SOIL moisture (SM) plays an essential role in many areas, such as crop yields, agricultural water management, weather forecasts, hydrology, natural disasters monitoring, energy and water exchanges at the atmosphere/land surface interface, and different Earth science applications [1], [2], [3]. NASA launched a mission in December 2016 called Cyclone Global Navigation Satellite System (CYGNSS), which leverages Global Navigation Satellite System-Reflectometry (GNSS-R), a technology gaining much attention within the scientific community due to its great potential to provide higher spatio–temporal coverage measurements over traditional microwave remote sensing techniques. GNSS-R operates by capturing signals reflected from the Earth's surface through bistatic scattering, effectively bridging the spatio–temporal gaps inherent in monostatic active or passive satellite missions. This approach involves cross-correlating GNSS signals received from a scattering surface with either a direct signal reception or a replica of the GNSS signal. GNSS-R enables the determination of various geophysical parameters of the observed surface, rendering it highly effective for monitoring sea surface roughness and wind vectors using both spaceborne and airborne systems [4], [5], [6]. GNSS-R is also the subject of extensive research for applications, such as biomass retrieval [7], sea ice monitoring [8], [9], ocean altimetry [10], and SM estimation [11], [12], [13], [14], [15], [16], [17].

One of the main advantages of using the CYGNSS mission is to utilize its high spatial and temporal coverage. CYGNSS operates with eight small microsatellites, each equipped to receive GNSS-R measurements from 32 channels during its 95-min orbital period. This configuration allows for a mean revisit time as short as 7 h, offering a spatial resolution of 25 km across ocean surfaces, primarily under diffuse scattering conditions. The primary objective of the CYGNSS mission is to enhance hurricane forecasting by deepening our understanding of the interactions between the atmospheric conditions near the core of a storm and the underlying sea surface. Remarkably, CYGNSS covers a wide geographical range, spanning from 38° north to 38° south latitudes, encompassing both land and ocean areas. This wide coverage provides valuable observations not only over water but also over terrestrial surfaces. CYGNSS works at *L*-band frequency, which can penetrate up to 5 cm from the Earth's surface providing observations suitable for SM estimation.

Several efforts have been conducted to retrieve surface SM using CYGNSS observations [12], [14], [15], [16], [18], [19], [20], [21], [22], [23], [24]. Out of these, the authors have

Manuscript received 20 March 2024; revised 7 May 2024; accepted 6 August 2024. Date of publication 14 August 2024; date of current version 18 September 2024. This work was supported in part by the National Science Foundation under Grant 2047771 and in part by the USDA-ARS Award NACA under Grant 58-6064-3-007. *(Corresponding author: Ali Cafer Gurbuz.)*

M M Nabi and Ali Cafer Gurbuz are with the Department of Electrical and Computer Engineering, and Information Processing and Sensing Lab, Mississippi State University, Mississippi State, MS 39762 USA (e-mail: mn918@msstate.edu; gurbuz@ece.msstate.edu).

Volkan Senyurek is with the Geosystems Research Institute, Mississippi State University, Mississippi State, MS 39762 USA (e-mail: volkan@gri.msstate.edu).

Mehmet Kurum is with the School of Electrical and Computer Engineering, University of Georgia, Athens, GA 30602 USA (e-mail: kurum@uga.edu).

Digital Object Identifier 10.1109/JSTARS.2024.3443100

previously developed several CYGNSS-based SM products [16], [18], [25] and made the global SM products publicly available.[1] These algorithms exhibit performance variation concerning time and space based on model capability, model complexity, and evaluation methods. In order to overcome these challenges, a fused SM product is needed that can combine multiple products and keep the properties of individual products. The primary research question we are trying to answer is "Can we fuse SM products in such a way that the fused products can achieve better estimation performance and can have the desired properties of each individual product? If we can fuse them, what will be the best approach?" Therefore, the main goal of this article is to create a fused SM product that will provide high-resolution SM with comparatively less error than the individual products and publicly accessible SM products. In this study, we apply fusion algorithms using CYGNSS-based SM products obtained from the Geosystems Research Institute (GRI) at Mississippi State University (MSU). These products, namely, MSU-GRI-V1.0A [18], MSU-GRI-V1.0B [16], and MSU-GRI-V1.0C [26], are publicly available. Additional details about these data products can be found in Section II.

There are different approaches for merging different remote sensing-based products, but only a few of them are related to SM fusion [27], [28], [29]. Recently, Hodges et al. [30] developed algorithms [minimum variance estimator (MVE) and hash lookup] for fusing multiple CYGNSS-based SM products.

Most of the existing techniques depend on statistical approaches on a constrained linear estimation [31] with varying assumptions depending on the application. In this article, we compare three different techniques i.e., MVE, best linear unbiased estimator (BLUE), and linear weight fusion (LWF) to fused CYGNSS-based SM products.

We will apply these three different algorithms to our MSU-GRI SM data products and evaluate their performance against the Soil Moisture Active Passive (SMAP) SM. The contributions of this article are summarized as follows.

1) Three different fusion algorithms have been proposed to combine CYGNSS-based SM data products.
2) We demonstrate the models' capability with and without using the reference/label (SMAP SM) data product.
3) The fused SM products are evaluated against SMAP mission's enhanced SM products at a 9 km $\times$ 9 km resolution.
4) MVE provides best performance when combining multiple products and shows an unbiased root-mean-square difference (ubRMSD) of 0.0359 m^3/m^3 with a correlation coefficient of 0.91. On the other hand, LWF method shows better performance with ubRMSD of 0.0389 m^3/m^3 with a correlation coefficient of 0.90 when reference SM value is not considered.

The rest of this article is organized as follows. Section II summarizes input datasets and the reference dataset for fusion algorithms. Details of the different fusion approaches are described in Section III. Results are presented in Section IV. The findings, difficulties, and implications for further research are discussed in Section V. Finally, Section VI concludes this article.

[1][Online]. Available: https://ssm.hpc.msstate.edu/

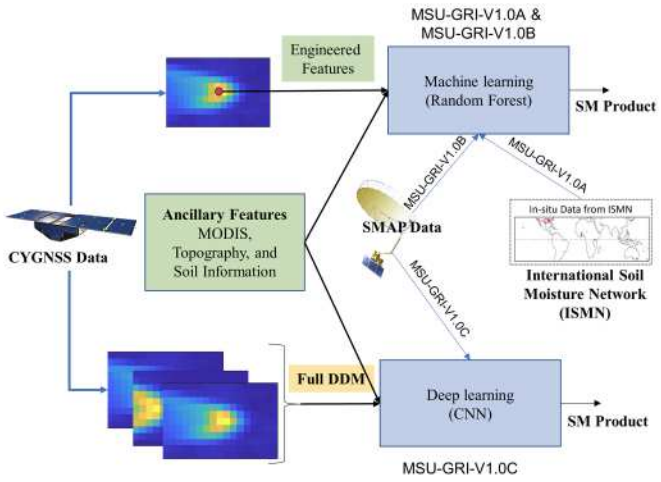

Fig. 1. MSU-GRI CYGNSS-based SM product generation using different approaches.

II. Dataset

For the analysis of fusion performance, three different CYGNSS-based SM products are considered [15], [16], [25]. These products are generated using different machine learning (ML) [16], [18] and deep learning [32] approaches using the available CYGNSS data and other auxiliary data from different sources. The overall SM product generation processes are illustrated in Fig. 1. Table I gives the general information, i.e., trained models, label source, resolution, and the reference for individual SM data products. All considered SM products have global coverage and are publicly available.[2] It is important to note that the input data products are chosen based on their availability and similar generation assumptions. Before applying the fusion model to each input dataset, we ensure that similar quality control flags are applied and that the datasets are in comparable resolutions. The products' details are described in the following sections.

A. MSU-GRI-V1.0A

This version of the product is trained using in situ SM data from the International Soil Moisture Network (ISMN) sites, along with various space-borne ancillary data [15], [18]. Initially, CYGNSS observables within the contiguous United States are employed to train the model. This ML-based SM product uses three features derived from CYGNSS, i.e., specular point incidence angle, trailing edge slope (TES), and the reflectivity [33]. In addition to CYGNSS measurements, five ancillary datasets are employed as input features for the ML model: 1) normalized difference vegetation index, 2) vegetation water content, 3) elevation, 4) soil clay ratio, and 5) soil silt ratio. In order to generate this product, a supervised learning problem ML model is used that maps a set of input features to the SM value, which is the final output label. Subsequently, the trained model is applied to predict global SM estimates. This approach

[2][Online]. Available: https://www.gri.msstate.edu/research/ssm/naca-uas-ssm-datasets.php

TABLE I
CYGNSS-BASED SM DATA PRODUCTS FROM GRI AT MSU

Products	Trained Model	Label source	Evaluated by	Output spatial resolution	CYGNSS Product version	Period	Main reference
MSU-GRI-V1.0A	RF	ISMN	ISMN and SMAP	9 km	L1 version 2.1	2017–2022	Senyurek et al. [15]
MSU-GRI-V1.0B	RF	SMAP	ISMN and SMAP	9 km	L1 version 2.1	2017–2022	Lei et al. [16]
MSU-GRI-V1.0C	Deep learning	SMAP	ISMN and SMAP	9 km	L1 version 3.1	2017–2022	Nabi et al. [25], [26]

yields daily SM retrievals, which are then gridded to resolutions of 3 and 9 km, corresponding to the CYGNSS spatial coverage.

The model's performance is independently assessed at various temporal scales, including daily, three-day, weekly, and monthly intervals, against both ISMN data and the enhanced SMAP mission's SM products, which have a resolution of 9 km $\times$ 9 km. The level-1 (L1) version 2.1 product CYGNSS data are used to generate the SM product. Random forest (RF) is used as an ML model. The training set is constructed using 170 ISMN sites over the globe. As the ground truth data availability is less for this method, the global error is slightly higher when compared with SMAP SM.

B. MSU-GRI-V1.0B

In this version of the product, a daily quasi-global SM estimate at a 9-km resolution is produced using the RF ML technique [16]. This approach differs significantly from MSU-GRI-V1.0A, which predominantly relied on a single ML algorithm applied to all available data for SM product generation. Instead of relying on localized information from in situ sites to represent a spatial grid, this work explores the use of multiple separately trained models, each employing distinct data stratification strategies to determine SM and assess its accuracy, with SMAP SM data serving as the base reference.

Multiple remote sensing land surface products characterizing the land surface conditions are incorporated to describe the nonlinear relationship between CYGNSS signals and SM. The ML-based model, optimally constructed for this purpose, has been designed to be adaptable for future SM retrieval efforts. Its effectiveness is assessed through year-based cross-validation and independent validation using data from sparsely distributed in situ measurement networks.

In addition, this quasi-global CYGNSS product is subjected to a rigorous and independent evaluation using the triple collocation (TC) technique. This comprehensive assessment ensures the reliability and robustness of the SM estimates derived from CYGNSS-based observations.

C. MSU-GRI-V1.0C

This product is generated [25] using one of the important observations from the CYGNSS mission along with other land surface data products. Delay-Doppler maps (DDMs) are used primarily as input features for this version of the CYGNSS-based SM product. A novel deep learning framework is introduced that considers entire DDMs (analog power, effective scattering area, and bistatic radar cross section) as input and other additional features as secondary inputs. In previous works, instead of using the whole DDMs, a few features were extracted, such as peak reflectivity, TES, leading edge slope, etc. Also, similar to the previous research [16], whole regions are stratified into several boxes, and a deep learning model is applied individually to each box. Instead of one single model, several models are generated that show improved performance and reduce the computational complexity [26].

Although all MSU-GRI products are developed using an ML or deep learning model, they are distinguished by differences in input data or features, ML models, and labeled data. Both MSU-GRI-V1.0A and MSU-GRI-V1.0B are developed using the RF model, but they are trained with ISMN (in situ SM) and SMAP (remotely sensed SM) data, as their true label data, respectively. However, due to the limited availability of ISMN sites worldwide, global SM retrieval in MSU-GRI-V1.0A is not as comprehensive as in MSU-GRI-V1.0B. Both of these versions consider only a few generated features from CYGNSS DDMs. The primary distinction between MSU-GRI-V1.0C and the previously mentioned versions lies in its consideration of full DDMs as the input to ML model. In MSU-GRI-V1.0C, the entire DDM images are utilized for SM retrieval. To directly employ the entire DDM dataset, a deep learning model is developed, which learns features from DDM images itself instead of designed features for other models. This shift to utilizing the complete DDM dataset represents a significant advancement in methodology, addressing the limitations of previous versions, and improving the reliability of SM retrieval on a global scale. Specifications on the ML model architecture for each SM product can be found in their respective publications.

D. SMAP Radiometer SM Data

The SMAP Enhanced L3 Radiometer Global Daily 9-km EASE-Grid SM product is used as a reference value for the fusion algorithms to evaluate the performance. SMAP uses the *L*-band microwave radiometer to collect brightness temperature data and produces SM estimates. Although the SMAP SM product is generated at 36-km resolution, it has also a 9-km enhanced grid product by using Backus–Gilbert optimal interpolation techniques [34]. SMAP datasets containing the associated coordinates for the descending (A.M.) and ascending (P.M.) overpasses are combined to obtain daily SM results. With the help of a 1000-km swath width, a daily SMAP product can cover about 70% of all land areas within the CYGNSS coverage ($\pm 38^\circ$ latitudes). The SMAP product also contains quality flags that indicate whether the SM retrieval is recommended or not. The data are freely available through the National Snow and Ice Data Center.[3] In this study, SMAP SM with a 9-km EASE-Grid product is used and CYGNSS specular points location is used to obtain the SM value from the SMAP data.

[3][Online]. Available: https://nsidc.org/data/SPL3SMP_E/versions/3

III. Methods

In this study, our goal is to combine MSU-GRI CYGNSS-based SM products to achieve an unbiased and minimum variance estimate of the underlying true SM parameter. To achieve this goal, we will explore the following statistical fusion models.

1) BLUE.
2) MVE.
3) LWF using TC.

These three statistical models are chosen for their widespread adoption and inherent linearity, enabling them to efficiently minimize variance while fusing multiple data products. Despite their robustness, these models are relatively simple and easy to implement.

Let us assume that $x_i[n]$ represents the nth SM product for the ith spatial grid. Here, $n \in \{1, 2, 3\}$ corresponds to the three different SM products described in Section II. Each statistical fusion model mainly has different data models relating the underlying true SM parameter to the observations and achieves this with different goals and constraints. Although we have three SM products to combine in this study, BLUE and MVE approaches are flexible enough to combine any more given SM product. Next, the working principle of each statistical estimation approach is described.

A. Best Linear Unbiased Estimator

A common approach in estimation theory is to restrict the estimator to be linear in the data and find the linear estimator that is unbiased and has the minimum variance, which is named BLUE [31]. Our data model relating the underlying unknown true parameter to the observed data product is

$$x_i[n] = \theta_i s[n] + w_i[n] \quad n = 1, 2, \ldots, N \tag{1}$$

where θ_i is the true SM value for the ith grid location that we would like to estimate. $x_i[n]$ are different independent data products that are achieved through different approaches, algorithms, or sensing systems. Here, n corresponds to a different product and we could have a total of N different ones. In our case $N = 3$, since we are fusing the three different SM products but in general we could have N data samples. The data model in (1) relates the data samples $x_i[n]$ to the underlying unknown true SM value θ_i through an unknown scalar $s[n]$, and $w_i[n]$ is white Gaussian noise with zero mean and unknown variance.

In BLUE, we restrict the estimator to be linear in data as

$$\widehat{\theta_i} = \sum_{n=1}^{N} a_n x_i[n] \tag{2}$$

where $\widehat{\theta_i}$ is the estimate of the underlying true parameter θ_i, and a_n's are constants that are yet to be determined. Depending on various selections of a_n's, different estimators can be achieved, but BLUE is defined as the one that is unbiased and has minimum variance. Letting the vector notations $\mathrm{a} = [a_1, a_2, \ldots, a_N]^T$, $\mathrm{s} = [s[1], s[2], \ldots, s[N]]^T$, and $\mathrm{x_i} = [x_i[1], x_i[2], \ldots, x_i[N]]^T$, the unbiased constraint is

$$\begin{aligned}\mathbb{E}(\widehat{\theta_i}) &= \sum_{n=1}^{N} a_n \mathbb{E}(x_i[n]) = \theta_i \\ &= \sum_{n=1}^{N} a_n s[n] \theta_i = \theta_i \\ &= \sum_{n=1}^{N} a_n s[n] = 1\end{aligned} \tag{3}$$

which can be simply stated as $\mathrm{a}^T\mathrm{s} = 1$. The variance of the estimator $\widehat{\theta_i}$ will be

$$\begin{aligned}\mathrm{var}(\widehat{\theta_i}) &= \mathbb{E}\left[\left(\widehat{\theta_i} - \mathbb{E}(\widehat{\theta_i})\right)^2\right] \\ &= \mathbb{E}\left[(\mathrm{a}^T\mathrm{x_i} - \mathrm{a}^T\mathbb{E}(\mathrm{x_i}))^2\right] \\ &= \mathbb{E}\left[\mathrm{a}^T(\mathrm{x_i} - \mathbb{E}(\mathrm{x_i}))(\mathrm{x_i} - \mathbb{E}(\mathrm{x_i}))^T\mathrm{a}\right] \\ &= \mathrm{a}^T\mathrm{C_i}\mathrm{a}\end{aligned} \tag{4}$$

where $\mathrm{C_i} = (\mathrm{x_i} - \mathbb{E}(\mathrm{x_i}))(\mathrm{x_i} - \mathbb{E}(\mathrm{x_i}))^T$ is the covariance matrix of the data. To find the BLUE, we need to minimize the variance $\mathrm{var}(\widehat{\theta_i}) = \mathrm{a}^T\mathrm{C_i}\mathrm{a}$ subject to the unbiased constraint $\mathrm{a}^T\mathrm{s} = 1$. This can be done using the method of Lagrangian multipliers where the Lagrangian function can be defined as $J = \mathrm{a}^T\mathrm{C_i}\mathrm{a} + \lambda(\mathrm{a}^T\mathrm{s} - 1)$. To solve for a, we take the gradient of J with respect to a as $\partial J/\partial \mathrm{a} = 2\mathrm{C_i}\mathrm{a} + \lambda\mathrm{s}$. Setting the gradient to zero and solving for a leads to $\mathrm{a} = -\lambda\mathrm{C_i}^{-1}\mathrm{s}/2$. The vector a is the weights used in the linear combination of data products, and to fully determine it, we need to find the Lagrangian multiplier λ using the solved a in the unbiased constraint as $\mathrm{a}^T\mathrm{s} = -\lambda\mathrm{s}^T\mathrm{C_i}^{-1}\mathrm{s}/2 = 1$. So, the Lagrangian multiplier is $\lambda = -2/(\mathrm{s}^T\mathrm{C_i}^{-1}\mathrm{s})$. Finally, the solution to the constraint minimization problem leading the optimal weights as

$$\mathrm{a_{opt}} = \frac{\mathrm{C_i}^{-1}\mathrm{s}}{\mathrm{s}^T\mathrm{C_i}^{-1}\mathrm{s}}. \tag{5}$$

Using the optimal weight vector $\mathrm{a_{opt}}$ in our estimator in (1) will lead to the BLUE, which can be directly computed as

$$\widehat{\theta_i} = \frac{\mathrm{s}^T\mathrm{C_i}^{-1}\mathrm{x_i}}{\mathrm{s}^T\mathrm{C_i}^{-1}\mathrm{s}}. \tag{6}$$

Computation of BLUE requires the covariance $\mathrm{C_i}$ and the scaled mean s to generate the fused SM product at the ith spatial grid, as shown in (6). Both $\mathrm{C_i}$ and s need to be estimated from available historical data.

B. Minimum Variance Estimator

Combining different individual estimates of a physical quantity to achieve an enhanced estimate has been highly used in the literature. One example is the MVE in [35] to combine five individual estimates of the wind speed exploiting the degree of decorrelation between the errors in the individual estimates to minimize the rms error in its wind speed estimate. In fact, MVE is a subcategory of BLUE where the scaled mean parameter is

assumed as $s[n] = 1$ leading to a data model

$$x_i[n] = \theta_i + w_i[n] \quad n = 1, 2, \ldots, N. \tag{7}$$

In this case, the unbiased constraint becomes $\mathbf{a}^T 1 = 1$, which fundamentally states that fusion weights should add up to one. Following a similar solution outlined in Section III-A, the optimal weights vector can be achieved as

$$\mathbf{a_{opt}} = \frac{\mathbf{C_i}^{-1} 1}{1^T \mathbf{C_i}^{-1} 1}. \tag{8}$$

While MVE does not need to estimate a scaled mean parameter s as BLUE does, it still needs to estimate the covariance $\mathbf{C_i}$ from historical data. Estimation performance of scaled mean and how well the data products fit into the assumed data models in (1) and (7) determines the performance of the fused data product, which is discussed in Section IV.

C. Linear Weight Fusion

TC analysis is a widely used technique for evaluating large-scale remote sensing products by incorporating a minimum of three mutually independent measurement systems. The data that are considered as "true" or "label" values could also contain their own measurement errors, biased calibration, and invalid validation errors. Assuming that each independent product is linearly related to the underlying true parameter and measurement errors for each product are uncorrelated with each other and the true values, TC is used to estimate the errors as well as weights to combine these products linearly [36], [37].

An affine error model [36] is commonly used in the TC literature for relating the observed data $x_i[n]$ to an underlying true geophysical variable θ_i as

$$x_i[n] = \alpha[n]\theta_i + \beta[n] + w_i[n] \quad n = 1, 2, \ldots, N \tag{9}$$

where $\alpha[n]$ and $\beta[n]$ are the deterministic coefficients of the affine model and $w_i[n]$ is the random error with unknown distribution. We assume that the errors from different data products have zero mean ($\mathrm{Cov}(w_i[n]) = 0, \ \forall n$) and uncorrelated with each other ($\mathrm{Cov}(w_i[n], w_i[m]) = 0, \ n \neq m$) and with true parameter θ_i ($\mathrm{Cov}(w_i[n], \theta_i) = 0, \ \forall n$). In TC analysis, the error variance values on the three independent SM data products are estimated as

$$\begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \sigma_3^2 \end{bmatrix} = \begin{bmatrix} C_{11} - \frac{C_{12}C_{13}}{C_{23}} \\ C_{22} - \frac{C_{12}C_{23}}{C_{13}} \\ C_{33} - \frac{C_{13}C_{23}}{C_{12}} \end{bmatrix} \tag{10}$$

where σ_n^2 is the error variance estimate of the nth data product and C_{kl} is the klth entry of the covariance matrix C. The LWF approach combines the three SM products depending on the error variance values of each data product. The weight vector $\mathbf{a} = [a_1, a_2, a_3]^T$ for TC in linearly combining data products as in (2) are given in [37] as

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \frac{1/\sigma_1^2}{1/\sigma_1^2+1/\sigma_2^2+1/\sigma_3^2} \\ \frac{1/\sigma_2^2}{1/\sigma_1^2+1/\sigma_2^2+1/\sigma_3^2} \\ \frac{1/\sigma_3^2}{1/\sigma_1^2+1/\sigma_2^2+1/\sigma_3^2} \end{bmatrix}. \tag{11}$$

Note that the weights in (11) add up to one, $a_1 + a_2 + a_3 = 1$, similar to the unbiased constraint in MVE. Although TC is practical to implement, the outlined form can only combine three different products. However, several studies have utilized TC to combine more than three products as well. For example, in [29], a multistep TC approach is presented to merge five active and passive microwave surface SM products. The process was done in two steps where the combined data product from the first step is used as another data product to combine with the remaining two products in the second step. In our case, three different SM data products are used to combine, hence it will be a single-step process.

In this work, we consider three different fusion algorithms selected for their straightforward implementation and potential for improved fusion estimation accuracy. However, each algorithm varies based on its underlying principles and assumptions, which in turn influence its performance characteristics. The BLUE algorithm aims to estimate a linear combination of the individual data sources, which minimizes the mean squared error while maintaining unbiasedness. Similarly, MVE seeks to minimize the variance of the estimated parameters, assuming normally distributed errors but with scaled mean assumed as one. In contrast, the LWF algorithm, based on TC, assumes linearity between each product and the underlying true parameter. It is noteworthy that while both BLUE and MVE can accommodate multiple input datasets simultaneously, LWF using TC is limited to three inputs at a time due to the nature of TC computation and a higher number of inputs can be fused three at a time in multiple stages. These distinctions highlight unique characteristics and requirements of the presented approaches, influencing their applicability and performance in various contexts.

IV. Results

In this section, we will present the performance of detailed CYGNSS-based SM fusion algorithms. To evaluate the fusion algorithm's performance, we use SMAP as the reference or label SM data. As performance metrics, we consider the root-mean-square difference (RMSD) and ubRMSD. In addition, we calculate the correlation coefficient (R-value) for individual data products and fused data products with and without the reference SMAP data. We also provide the fusion weights for each SM product presenting how much each single data product contributes to the final fused result.

Before applying the fusion algorithms, we aggregate three distinct SM products into separate vectors. For our analysis, we encompass the entire world, utilizing SM products spanning from 2017 to 2022 as input data products for various fusion methods detailed in Section III. As previously mentioned, SMAP serves as the reference dataset for calculating performance metrics. We assess the performance of each individual product to validate

TABLE II
PERFORMANCE ANALYSIS AGAINST SMAP SM FOR ALL AVAILABLE GRIDS ACROSS THE WORLD

Data categories	Products	RMSD	Bias	ubRMSD	R-value
Input Data Products	MSU-GRI-V1.0A	0.0918	0.0008	0.0821	0.87
	MSU-GRI-V1.0B	0.0538	-0.0125	0.0527	0.93
	MSU-GRI-V1.0C	0.0518	-0.0215	0.0489	0.93
Fused data (not using SMAP data)	BLUE	0.1018	-0.0551	0.0857	0.84
	MVE	0.0895	-0.0218	0.0868	0.79
	LWF	0.0552	-0.0095	0.0544	0.92
Fused data (using SMAP data)	BLUE	0.0513	0.0004	0.0513	0.93
	MVE	0.0514	-0.0022	0.0441	0.93
	LWF	0.0582	-0.0120	0.0570	0.92

the efficacy of the fusion methods. Section IV is divided into two parts. The first part applies a single estimator for the entire world. Hence, we obtain a single-weight vector to combine SM data products for all grids. The second part is applying fusion regionwise or a small number of grids. For this part, different estimators and corresponding weight vectors are obtained for each region.

A. Single Fusion Estimator for Entire World

In this part, we learn a single fusion estimator and hence a single fusion weight vector a that is used to generate the fused SM product for all grids. Note that all three fusion techniques depend on the covariance matrix of the data, while BLUE also needs to estimate an additional scaled mean parameter s. In the estimation of these parameters, we follow two different approaches. The first approach is unsupervised and only the three data products are used with no reference data in this case. To estimate the true covariance between different data products, we use the sample covariance which can be stated as

$$\widehat{C} = \frac{1}{N-1}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T \tag{12}$$

where $\mathbf{x}_n$ is the 3×1 size vector holding the SM values for the three different data products for a specific day and grid location, $\bar{\mathbf{x}}$ is the mean vector of all the data used in fusion, and N represents the total number of data samples. For the supervised case, we use the SMAP reference data while generating the sample covariance matrix. For the supervised case, we randomly split the dataset into 20% and 80% ratios. The 20% part of the data are used in the fusion to estimate the sample covariance for this case. After the fusion weights are determined, the remaining 80% of the data are used to evaluate the performance of the fused product. For the supervised case, the sample covariance estimation is done as

$$\widehat{C} = \frac{1}{N-1}\sum_{n=1}^{N}(\mathbf{x}_n - \mathbf{x}_n^{\text{ref}})(\mathbf{x}_n - \mathbf{x}_n^{\text{ref}})^T \tag{13}$$

where $\mathbf{x}_n^{\text{ref}}$ is the reference vector containing the SMAP SM values for the corresponding data sample.

The RMSD, bias, ubRMSD, and R-value metrics are calculated for the individual data products and each fused data product that is generated with the proposed fusion approaches under both supervised and unsupervised settings. The resultant performance metrics are provided in Table II for all available grids across the world. When the individual input data products are considered, it can be seen that the products that are trained using the SMAP dataset perform better than the product generated using a model trained using ISMN station data. Even though two of the data products have similar average performances, the deep learning-based approach (MSU-GRI-V1.0C) achieves lower RMSD and ubRMSD values. The next part of the table presents fusion performance in an unsupervised scenario. It is evident that when no reference data are used, best performing fusion approach is LWF. However, in this case, none of the fused products can provide lower ubRMSD or higher R-values than the best individual product. The MVE approach results in a high ubRMSD and a low R-value. The BLUE algorithm exhibits a similar performance, providing a ubRMSD of 0.0857 m^3m^{-3} and a slightly higher correlation than the MVE algorithm. However, we observe comparatively better performance for LWF, which yields a global ubRMSD of 0.0544 m^3m^{-3} and a correlation coefficient of 0.92 for the unsupervised case. For the supervised fusion case, the achieved performance metrics are better compared with the unsupervised fusion case, specifically for BLUE and MVE approaches. We observe that MVE produces an ubRMSD of 0.0441 m^3m^{-3} and a correlation of 0.93 when reference data are used for the fusion algorithm. The fused product with the MVE approach shows better performance than any of the individual data products. BLUE shows slightly higher errors than MVE. LWF does not seem to improve in this case compared to not using the reference data.

Table III presents the performance for the SMAP recommended grids. Similar to the previous table, we first present the performance of each of our input data products and later present the performance of the fusion algorithms, both for supervised and unsupervised cases. We observe similar performance for LWF when we do not use the reference value. It provides an ubRMSD of 0.0389 m^3m^{-3} and a correlation of 0.90, which is comparatively higher than the individual data products. On the other hand, when we consider the reference, MVE provides the best performance for fusion, with an ubRMSD of 0.0359 m^3m^{-3}. We observe that BLUE does not perform significantly well for both cases.

Tables II and III present results for fusion algorithms generated using individual weights for each single model and multiplying them with individual products. This means that for global SM fusion, there are three different weights for three

TABLE III
PERFORMANCE ANALYSIS AGAINST SMAP SM FOR SMAP RECOMMENDED GRIDS

Data categories	Products	RMSD	Bias	ubRMSD	R-value
Input Data Products	MSU-GRI-V1.0A	0.0693	9.78e-04	0.0693	0.71
	MSU-GRI-V1.0B	0.0429	-0.0080	0.0422	0.88
	MSU-GRI-V1.0C	0.0371	0.0025	0.0370	0.91
Fused data (not using SMAP data)	BLUE	0.0602	-0.0383	0.0465	0.90
	MVE	0.0442	-0.0078	0.0435	0.87
	LWF	0.0393	-0.0051	0.0389	0.90
Fused data (using SMAP data)	BLUE	0.0359	-6.15e-04	0.0359	0.91
	MVE	0.0359	-4.44e-04	0.0359	0.91
	LWF	0.0380	-0.0038	0.0378	0.90

TABLE IV
WEIGHT CONTRIBUTIONS FOR EACH PRODUCT AFTER APPLYING THE FUSION ALGORITHMS

Products	Weights (BLUE)	Weights (MVE)	Weights (LWF)
MSU-GRI-V1.0A	-0.0933	-0.0355	0.0497
MSU-GRI-V1.0B	0.2012	0.2890	0.2601
MSU-GRI-V1.0C	0.8922	0.7465	0.6902

This table gives overall global weights values for all three products when we consider supervised case.

TABLE V
WEIGHT CONTRIBUTIONS FOR EACH PRODUCT AFTER APPLYING THE FUSION ALGORITHMS

Products	Mean Weights (BLUE)	Mean Weights (MVE)	Mean Weights (LWF)
MSU-GRI-V1.0A	-0.0477	-0.0566	0.1042
MSU-GRI-V1.0B	0.4883	0.3716	0.5802
MSU-GRI-V1.0C	0.6003	0.6850	0.3156

This table gives mean of weights for 144-km grid regions for all three products considering supervised case.

different products. In each SM fusion algorithm, we have individual weights for each SM product. In Table IV, we illustrate how individual SM products contribute to the fusion algorithms considering the supervised case. For the MVE algorithm, we observe weight values of -0.0355, 0.2890, and 0.7465 for MSU-GRI-V1.0A, MSU-GRI-V1.0B, and MSU-GRI-V1.0C, respectively. In the case of the BLUE algorithm, approximately 90% of the weights are attributed to the MSU-GRI-V1.0C product, whereas MSU-GRI-V1.0A has negative weights on the fusion algorithm. This indicates that the MSU-GRI-V1.0C product contributes more weight to MVE compared with the other two SM products. For LWF, we observe a slightly lower contribution from the MSU-GRI-V1.0C product, but it still maintains a high contribution compared with the other two methods.

B. Applying Regionwise (144-Km) Fusion

All the previous results were generated using single-weight values for individual products for the fusion algorithms. It is difficult to observe the regionwise product contribution for the fusion algorithm with this approach. To understand the regionwise contribution, we divide the global data into 144-km small regions and apply each fusion algorithm within those regions. This allows us to determine the contribution of each product in different regions. We hypothesize that all input data products should not perform similar way for different regions. Table V summarizes the mean contribution for each 144-km small region. For the BLUE algorithm, we notice that the MSU-GRI-V1.0B product contributes more weight compared with MSU-GRI-V1.0A, but this method still predominantly uses the MSU-GRI-V1.0C product.

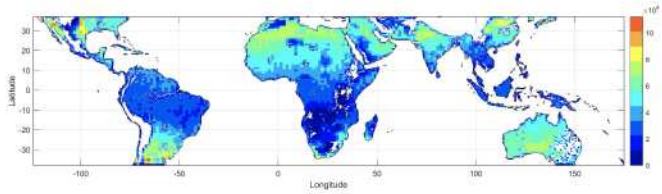

Fig. 2. Number of data samples for each 144-km grids globally.

For the MVE algorithm, we observe weight values of -0.0566, 0.3716, and 0.6850 for MSU-GRI-V1.0A, MSU-GRI-V1.0B, and MSU-GRI-V1.0C, respectively. This mean result is slightly different from the single-valued weight results presented in Table IV. Still, the MSU-GRI-V1.0C product receives more weight when considering the 144-km grid case.

On the other hand, we observe significantly different results for the LWF method. The MSU-GRI-V1.0B product provides more weight than the other two products during the fusion algorithm in this context. LWF relies on the TC algorithm, and it is highly data dependent with several assumptions. Since we are considering a small region, there may not be enough data for the model to execute the fusion algorithm, resulting in the failure to generate the fusion product. Although the other two algorithms work very similarly for a single model and the 144-km scenarios, LWF surprisingly provides different results. There may be a question about choosing the 144-km grid case. To address this, we need a sufficient amount of data for the fusion algorithm. If we consider less than 144 km, the data sample size becomes smaller, leading to significant discrepancies in results. Therefore, we have chosen to continue with the 144-km case. In Fig. 2, the data sample distributions for each 144-km region are presented. This visualization helps us understand how the models are affected by the data samples.

In Fig. 3, 144-km gridwise weight values are presented using the BLUE algorithm. These maps are generated using MSU-GRI-V1.0A, MSU-GRI-V1.0B, and MSU-GRI-V1.0C (from top to bottom). As we can observe, the top map in Fig. 3 represents MSU-GRI-V1.0A, which indicates that the BLUE algorithm does not use this product at all in the fusion process. BLUE generated low weights for MSU-GRI-V1.0A. The MSU-GRI-V1.0B gives relatively higher weight toward the Sahara

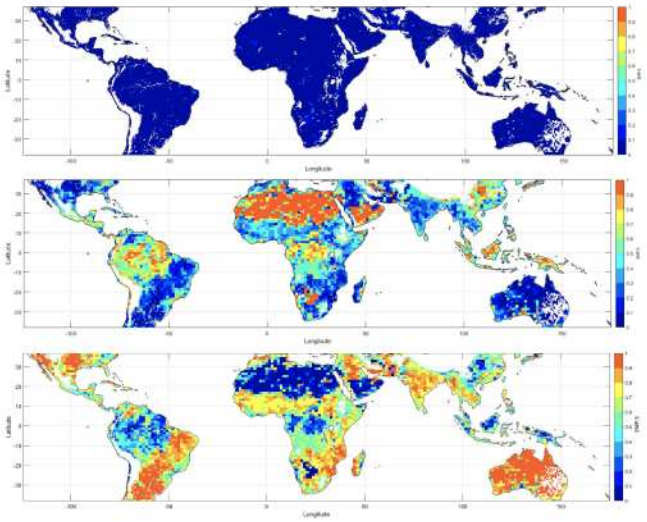

Fig. 3. 144-km gridwise global weight values generated using the BLUE for each data inputs. The top map is for MSU-GRI-V1.0A, the middle is MSU-GRI-V1.0B, and the bottom one is MSU-GRI-V1.0C. It is evident that MSU-GRI-V1.0C maintains high weight values compared to other two products.

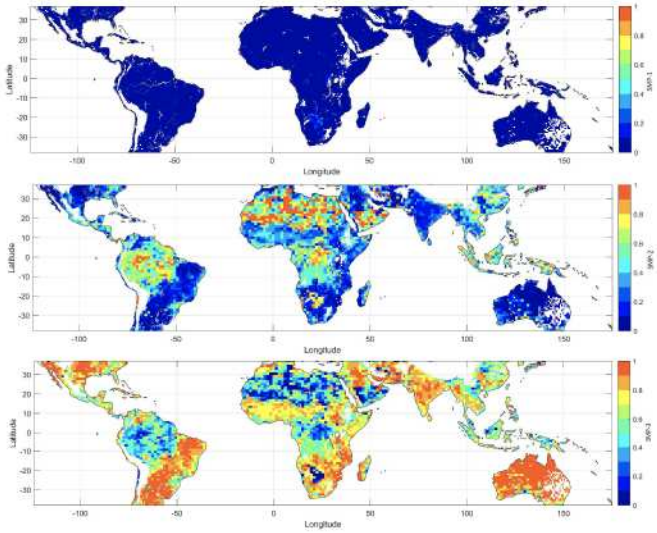

Fig. 4. 144-km gridwise global weight values generated using the MVE for each data inputs. The top map is for MSU-GRI-V1.0A, the middle is MSU-GRI-V1.0B, and the bottom one is MSU-GRI-V1.0C. It is evident that MSU-GRI-V1.0C maintains high weight values compared to other two products.

and Amazon regions, but the other parts of the world have low weights compared to MSU-GRI-V1.0C. MSU-GRI-V1.0B product shows comparatively high weights across the world, which proves that it has a good contribution to the SM fusion method.

In Fig. 4, 144-km gridwise weight values are presented for each individual data product case for MVE. Similar to the BLUE, MVE also generated low weights for MSU-GRI-V1.0A. Instead, it primarily relies on MSU-GRI-V1.0B and MSU-GRI-V1.0C, with MSU-GRI-V1.0C making a significant contribution to the MVE algorithm.

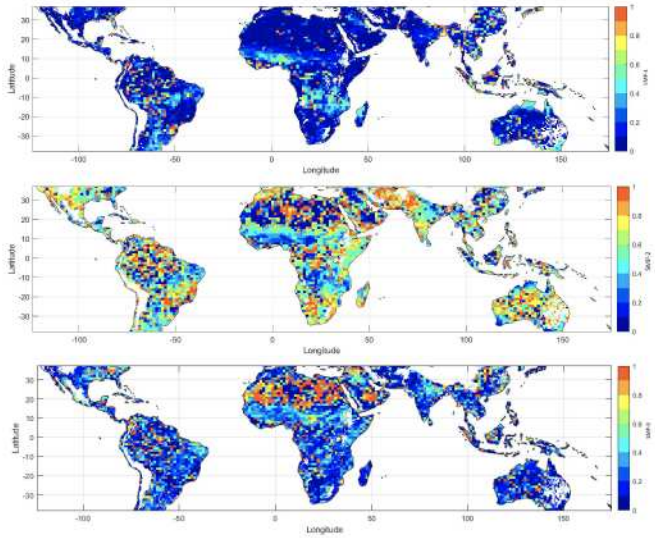

Fig. 5. 144-km gridwise global weight values generated using the LWF for each data inputs. The top map is for MSU-GRI-V1.0A, the middle is MSU-GRI-V1.0B, and the bottom one is MSU-GRI-V1.0C. Here, MSU-GRI-V1.0B contributes more than the other two products.

In Fig. 5, 144-km gridwise weight values are presented using the LWF algorithm. As we have already observed the quantitative results in Table V for LWF, the maps demonstrate a similar trend. MSU-GRI-V1.0C products receive lower weight compared with the MSU-GRI-V1.0B. It is also evident that MSU-GRI-V1.0A has some little weights for the LWF, whereas we did not see any impact for BLUE and MVE. As we already mentioned, LWF is data dependent and works based on several assumption that might impact the fusion process.

V. Discussion

SM is an essential property for Earth observation, hydrology, and several other land surface applications. Space-borne GNSS-R observations have gained popularity due to their wide coverage and cost-effectiveness. NASA's CYGNSS mission, developed for its primary purpose of wind estimation, has had a significant impact by providing crucial information for SM estimation. Researchers have developed a wide range of techniques to estimate SM and create the best global SM product. The estimation techniques depend upon various factors, including data processing, quality control, and model complexity, resulting in different types of estimation outputs. Different global SM products derived from the CYGNSS mission exhibit strong performance in different regions. This underscores the importance of merging or fusing multiple SM products. However, it is important to note that not all CYGNSS-based SM data products are publicly available for use. Therefore, this study focuses solely on the publicly accessible data provided by the GRI at MSU. Over the years, the GRI has developed various approaches. It has become necessary to consolidate these efforts into a single official data product for the institute, rather than relying on multiple versions. Such an initiative can serve as a model for the CYGNSS-based

SM community, facilitating the adoption of an official product for the CYGNSS mission.

In this study, we demonstrate three different approaches for merging various SM products. We selected these approaches due to their straightforward implementation and common usage in remote sensing-based data products. The MVE approach, known for its effectiveness in merging wind speed estimation products [35], provided excellent fusion results. Hence, we introduce it for global SM fusion. This approach is renowned for delivering lower or equal RMSD error compared with individual RMSD errors. In addition, we demonstrate the BLUE algorithm [31], which is advantageous when full knowledge of the probability density function is not required. Previously, the BLUE algorithm was employed as an interpolation technique for CYGNSS-based SM products [38], and it is now utilized for SM fusion in this study. Although, as we observe in the results, BLUE does not perform as well as MVE, its overall performance is quite satisfactory. MVE does not require estimation of the scaled mean parameter, unlike BLUE, which necessitates the estimation of both the scaled mean and covariance. Instead, MVE solely estimates the covariance from historical data. Since the estimation performance is influenced by how well the data fit into the model, the accuracy of the scaled mean estimation could be a reason for enhanced MVE fusion performance compared with BLUE.

Both MVE and BLUE approaches can be employed when reference SM data (in situ or SMAP) are available. However, if access to reference data is challenging, these approaches may not perform optimally. To address the issue of reference data unavailability, we can consider the LWF algorithm [29], which uses a TC approach to estimate error variances, making it a viable alternative for SM fusion.

We demonstrate two different analyses for fusion cases, one is to generate product single weight globally for each product and another is to generate multiple weights for each 144-km small region. Generating global fusion is easier compared to the 144-km grid scenarios. But to get the regionwise product contribution, it is better to utilize small regions. If the data availability is limited for the small, the fusion algorithm can provide high error. So, it is important to note that choosing a fusion algorithm needs to be carefully considered when merging several products. In addition, these three methods can fuse data within a fraction of a second if we perform a single global fusion. However, performing fusion for small regions individually and subsequently integrating them on a global scale will require additional processing time for each region.

Although this study focuses on the fusion of various CYGNSS-based SM products, the developed methodology holds applicability to the fusion of any GNSS-R-based remotely sensed data products. Several GNSS-R missions, including the U.K. Disaster Monitoring Constellation satellite [39], U.K. TechDemoSat-1 [40], Chinese BuFeng-1 A/B satellites [41], and FSSCat [42], provide Earth surface information. By leveraging our proposed fusion algorithm, researchers can enhance estimation performance for different Earth observations by integrating data from these diverse sources. Furthermore, our methodology can be extended to fuse additional environmental parameters from the CYGNSS mission, such as wind speed data [43] and freeze/thaw surface state products [44], offering a comprehensive approach for multisensor data fusion in environmental monitoring and modeling.

VI. Conclusion

This study presented different SM fusion algorithms to combine estimated CYGNSS-based SM data products. Three different SM data products are utilized to demonstrate the capability of three different fusion algorithms. Based on our analysis, we observed that MVE showed a good potential for merging different SM products. In order to evaluate the fusion methods, we utilized SMAP as a reference. This method achieved an ubRMSD of 0.0359 m^3m^{-3} and a correlation of 0.91 for SMAP-recommended grids. This approach requires reference SM data during the fusion process. It is observed that LWF can achieve an ubRMSD of 0.0389 m^3m^{-3} with a correlation of 0.90 when reference SM data are not used. LWF shows good potential when true data are not present.

References

[1] D. A. Robinson et al., "Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review," *Vadose Zone J.*, vol. 7, no. 1, pp. 358–389, 2008.

[2] M. E. Holzman, R. Rivas, and M. C. Piccolo, "Estimating soil moisture and the relationship with crop yield using surface temperature and vegetation index," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 28, pp. 181–192, 2014.

[3] K. Das and P. K. Paul, "Present status of soil moisture estimation by microwave remote sensing," *Cogent Geosci.*, vol. 1, no. 1, 2015, Art. no. 1084669.

[4] E. Valencia, V. U. Zavorotny, D. M. Akos, and A. Camps, "Using DDM asymmetry metrics for wind direction retrieval from GPS ocean-scattered signals in airborne experiments," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 3924–3936, Jul. 2014.

[5] A. Komjathy, M. Armatys, D. Masters, P. Axelrad, V. Zavorotny, and S. Katzberg, "Retrieval of ocean surface wind speed and wind direction using reflected GPS signals," *J. Atmospheric Ocean. Technol.*, vol. 21, no. 3, pp. 515–526, 2004.

[6] D. Guan et al., "Wind direction signatures in GNSS-R observables from space," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 198.

[7] E. Santi et al., "Forest biomass estimate on local and global scales through GNSS reflectometry techniques," in *Proc. IGARSS 2019-2019 IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 8680–8683.

[8] Q. Yan and W. Huang, "Spaceborne GNSS-R sea ice detection using delay-Doppler maps: First results from the U.K. TechDemoSat-1 mission," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 10, pp. 4795–4801, Oct. 2016.

[9] N. Rodriguez-Alvarez, B. Holt, S. Jaruwatanadilok, E. Podest, and K. C. Cavanaugh, "An Arctic sea ice multi-step classification based on GNSS-R data from the TDS-1 mission," *Remote Sens. Environ.*, vol. 230, 2019, Art. no. 111202.

[10] W. Li, E. Cardellach, F. Fabra, S. Ribó, and A. Rius, "Assessment of spaceborne GNSS-R ocean altimetry performance using CYGNSS mission raw data," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 238–250, Jan. 2020.

[11] C. C. Chew and E. E. Small, "Soil moisture sensing using spaceborne GNSS reflections: Comparison of CYGNSS reflectivity to SMAP soil moisture," *Geophys. Res. Lett*, vol. 45, no. 9, pp. 4049–4057, 2018.

[12] M. P. Clarizia, N. Pierdicca, F. Costantini, and N. Floury, "Analysis of CYGNSS data for soil moisture retrieval," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2227–2235, Jul. 2019.

[13] H. Kim and V. Lakshmi, "Use of Cyclone Global Navigation Satellite System (CYGNSS) observations for estimation of soil moisture," *Geophys. Res. Lett*, vol. 45, no. 16, pp. 8272–8282, 2018.

[14] O. Eroglu, M. Kurum, D. Boyd, and A. C. Gurbuz, "High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks," *Remote Sens.*, vol. 11, no. 19, 2019, Art. no. 2272.
[15] V. Senyurek, F. Lei, D. Boyd, M. Kurum, A. Gurbuz, and R. Moorhead, "Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS," *Remote Sens.*, vol. 12, no. 7, 2020, Art. no. 1168.
[16] F. Lei et al., "Quasi-global machine learning-based soil moisture estimates at high spatio-temporal scales using CYGNSS and SMAP observations," *Remote Sens. Environ.*, vol. 276, 2022, Art. no. 113041.
[17] S. H. Yueh, R. Shah, M. J. Chaubell, A. Hayashi, X. Xu, and A. Colliander, "A semiempirical modeling of soil moisture, vegetation, and surface roughness impact on CYGNSS reflectometry data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Nov., 2020, Art. no. 5800117.
[18] V. Senyurek, F. Lei, D. Boyd, A. C. Gurbuz, M. Kurum, and R. Moorhead, "Evaluations of a machine learning-based CYGNSS soil moisture estimates against SMAP observations," *Remote Sens.*, vol. 12, no. 21, 2020, Art. no. 3503.
[19] C. C. Chew and E. E. Small, "Soil moisture sensing using spaceborne GNSS reflections: Comparison of CYGNSS reflectivity to SMAP soil moisture," *Geophys. Res. Lett.*, vol. 45, no. 9, pp. 4049–4057, 2018.
[20] H. Kim and V. Lakshmi, "Use of Cyclone Global Navigation Satellite System (CyGNSS) observations for estimation of soil moisture," *Geophys. Res. Lett.*, vol. 45, no. 16, pp. 8272–8282, 2018.
[21] A. Calabia, I. Molina, and S. Jin, "Soil moisture content from GNSS reflectometry using dielectric permittivity from Fresnel reflection coefficients," *Remote Sens.*, vol. 12, no. 1, 2020, Art. no. 122.
[22] Q. Yan, W. Huang, S. Jin, and Y. Jia, "Pan-tropical soil moisture mapping based on a three-layer model from CYGNSS GNSS-R data," *Remote Sens. Environ.*, vol. 247, 2020, Art. no. 111944.
[23] T. Yang, W. Wan, Z. Sun, B. Liu, S. Li, and X. Chen, "Comprehensive evaluation of using TechDemoSat-1 and CYGNSS data to estimate soil moisture over mainland China," *Remote Sens.*, vol. 12, no. 11, 2020, Art. no. 1699.
[24] Y. Jia et al., "Temporal-spatial soil moisture estimation from CYGNSS using machine learning regression with a preclassification approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4879–4893, Apr. 29, 2021.
[25] M. Nabi, V. Senyurek, F. Lei, M. Kurum, and A. C. Gurbuz, "Quasi-global assessment of deep learning-based CYGNSS soil moisture retrieval," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 5629–5644, Jun. 20, 2023.
[26] M. Nabi, V. Senyurek, A. C. Gurbuz, and M. Kurum, "Deep learning-based soil moisture retrieval in CONUS using CYGNSS delay–Doppler maps," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6867–6881, Aug. 5 2022.
[27] R. Prakash, D. Singh, and N. P. Pathak, "A fusion approach to retrieve soil moisture with SAR and optical data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 196–206, Feb. 2012.
[28] A. S. Abowarda et al., "Generating surface soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale," *Remote Sens. Environ.*, vol. 255, 2021, Art. no. 112301.
[29] Q. Xie, L. Jia, M. Menenti, and G. Hu, "Global soil moisture data fusion by triple collocation analysis from 2011 to 2018," *Sci. Data*, vol. 9, no. 1, 2022, Art. no. 687.
[30] E. Hodges et al., "Foundations of a blended CYGNSS soil moisture product," in *Proc. IEEE GNSS R 2023*, Boulder, CO, USA, May 2023.
[31] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
[32] M. Nabi, V. Senyurek, A. C. Gurbuz, and M. Kurum, "A deep learning-based soil moisture estimation in CONUS region using CYGNSS delay Doppler maps," in *Proc. IGARSS 2022-2022 IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 6177–6180.
[33] N. Rodriguez-Alvarez, E. Podest, K. Jensen, and K. C. McDonald, "Classifying inundation in a tropical wetlands complex with GNSS-R," *Remote Sens.*, vol. 11, no. 9, 2019, Art. no. 1053.
[34] S. K. Chan et al., "Development and assessment of the SMAP enhanced passive soil moisture product," *Remote Sens. Environ.*, vol. 204, pp. 931–941, 2018.
[35] M. P. Clarizia, C. S. Ruf, P. Jales, and C. P. Gommenginger, "Spaceborne GNSS-R minimum variance wind speed estimator," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6829–6843, Nov. 2014.
[36] S. Zwieback, K. Scipal, W. Dorigo, and W. Wagner, "Structural and statistical properties of the collocation technique for error characterization," *Nonlinear Processes Geophys.*, vol. 19, no. 1, pp. 69–80, 2012.
[37] K. A. McColl, J. Vogelzang, A. G. Konings, D. Entekhabi, M. Piles, and A. Stoffelen, "Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target," *Geophys. Res. Lett.*, vol. 41, no. 17, pp. 6229–6236, 2014.
[38] V. Senyurek, A. C. Gurbuz, and M. Kurum, "Assessment of interpolation errors of CYGNSS soil moisture estimations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9815–9825, Sep. 20, 2021.
[39] S. Gleason et al., "Detection and processing of bistatically reflected GPS signals from low Earth orbit for the purpose of ocean remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1229–1241, Jun. 2005.
[40] M. Unwin, P. Jales, J. Tye, C. Gommenginger, G. Foti, and J. Rosello, "Spaceborne GNSS-reflectometry on TechDemoSat-1: Early mission operations and exploitation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 10, pp. 4525–4539, Oct. 2016.
[41] C. Jing, X. Niu, C. Duan, F. Lu, G. Di, and X. Yang, "Sea surface wind speed retrieval from the first Chinese GNSS-R mission: Technique and preliminary results," *Remote Sens.*, vol. 11, no. 24, 2019, Art. no. 3013.
[42] A. Camps et al., "FSSCaT, the 2017 Copernicus masters' "ESA Sentinel small satellite challenge" winner: A federated polar and soil moisture tandem mission based on 6 U CubeSats," in *Proc. IGARSS 2018-2018 IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 8285–8287.
[43] J. Reynolds, M. P. Clarizia, and E. Santi, "Wind speed estimation from CYGNSS using artificial neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 708–716, Feb. 5, 2020.
[44] H. Carreno-Luengo and C. S. Ruf, "Retrieving freeze/thaw surface state from CYGNSS measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 15, 2021, Art. no. 4302313.

M M Nabi received the bachelor's degree in electrical and electronics engineering from the Ahsanullah University of Science and Technology, Dhaka, Bangladesh, in 2014, the master's degree in information technology from the University of Dhaka, Dhaka, in 2016, and the Ph.D. degree in electrical and computer engineering from Mississippi State University (MSU), Mississippi State, MS, USA, in 2024.

He has been working as a Graduate Research Assistant with the High-Performance Computing Collaboratory and the Information Processing and Sensing Laboratory, MSU. He will be joining Western Kentucky University as an Assistant Professor in Fall 2024. His research interests include signal processing, remote sensing, machine learning, and deep learning.

Volkan Senyurek received the B.S., M.S., and Ph.D. degrees in electronics and communication engineering from Marmara University, Istanbul, Türkiye, in 2003, 2007, and 2013, respectively.

Until 2015, after he got the Ph.D. degree, he was an Assistant Professor with Marmara University. Between 2015 and 2017, he was a Postdoctoral Researcher with the Department of Mechanical and Materials Engineering, Florida International University. Between 2017 and 2019, he joined the Department of Electric and Computer Engineering, the University of Alabama, as a Postdoctoral Researcher. He is currently an Assistant Research Professor with Geosystems Research Institute, Mississippi State University, Mississippi State MS, USA. His research interests include remote sensing, biomedical signal processing, wearable sensors, pattern recognition, fiber optic sensors, and structural health monitoring.

Mehmet Kurum (Senior Member, IEEE) received the B.S. degree in electrical and electronics engineering from Bogazici University, Istanbul, Türkiye, in 2003, and the M.S. and Ph.D. degrees in electrical engineering from George Washington University, Washington, DC, USA, in 2005 and 2009, respectively.

He held Postdoctoral and Research Associate positions with the Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA. From 2016 to 2022, he was an Assistant Professor with Mississippi State University (MSU), and subsequently, he held the position of Associate Professor and the Paul B. Jacob endowed chair until 2023. He is currently an Associate Professor in electrical and computer engineering with the University of Georgia, Athens, Georgia, while also an Adjunct Professor with MSU. His current research focuses on recycling the radio spectrum to address the challenges of decreasing radio spectrum space for science while exploring entirely new microwave regions for land remote sensing.

Dr. Kurum is a Senior Member of IEEE Geoscience and Remote Sensing Society (GRSS) and a Member of U.S. National Committee for the International Union of Radio Science (USNC-URSI). He is currently the Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING since 2021. He was the recipient of the Leopold B. Felsen Award for excellence in electromagnetic in 2013 and URSI Young Scientist Award in 2014, and NSF CAREER award in 2022. He was an Early Career Representative for the International URSI Commission F (Wave Propagation and Remote Sensing) from 2014 to 2021.

Ali Cafer Gurbuz (Senior Member, IEEE) received the B.S. degree in electrical engineering from Bilkent University, Ankara, Türkiye, in 2003, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2005 and 2008, respectively.

From 2003 to 2009, he researched compressive sensing-based computational imaging problems with Georgia Tech. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA, where he is the Director of Information Processing and Sensing (IMPRESS) Laboratory. His research interests include signal/image processing and machine learning for radar, remote sensing and wireless communications with focus on compressive sensing, physics aware and explainable machine learning for inverse problems, computational imaging, multisensor activity sensing, radar/communication coexistence, passive sensing, autonomous aerial vehicle-based remote sensing and precision agriculture.

Dr. Gurbuz was the recipient of The Best Paper Award for *Signal Processing Journal* in 2013, the Turkish Academy of Sciences Best Young Scholar Award in electrical engineering in 2014, and NSF CAREER award in 2021. He is an Associate Editor for IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS and was an Associate Editor for journals, such as *Digital Signal Processing*, *EURASIP Journal on Advances in Signal Processing*, and *Physical Communications*.