

# Assessment of Interpolation Errors of CYGNSS Soil Moisture Estimations

Volkan Senyurek , Ali C. Gurbuz , *Senior Member, IEEE*, Mehmet Kurum , *Senior Member, IEEE*

**Abstract**—High spatio-temporal soil moisture (SM) is essential for many meteorological, hydrological, and agricultural applications and studies. Spaceborne Global Navigation Satellite System Reflectometry (GNSS-R) provides a promising opportunity for high-resolution SM retrievals. NASA's Cyclone Global Navigation Satellite System (CYGNSS) is a recent GNSS-R application that offers relatively high spatial and temporal resolution observations from Earth's surface. However, the quasi-random sampling of land surface by the CYGNSS constellation circumvents obtaining fully observed daily SM predictions at high spatial resolutions. Spatial interpolation techniques may fill this gap and provide a fully covered high-resolution daily SM estimation. However, the spatial interpolation errors need to be assessed when applied to the quasi-random 9-km CYGNSS based SM estimations. In this paper, we conduct interpolation error analysis using the SMAP Enhanced L3 Radiometer Global Daily 9-km product, sampled at the CYGNSS observation locations. The results indicate that the overall interpolation error (RMSE) was  $0.013 \text{ m}^3\text{m}^{-3}$  over SMAP's recommended grids. In addition, sparse CYGNSS SM observations are directly interpolated. The achieved results show that interpolated and observed CYGNSS SM values have similar performance metrics when validated with the SMAP 9-km gridded SM product as well as sparse soil moisture networks.

**Index Terms**—CYGNSS, reflectometry, soil moisture, SMAP, interpolation, Kriging, BLUE.

## I. INTRODUCTION

SOIL moisture (SM) is one of the critical components of agriculture and terrestrial water management, and global climate studies [1]. Accurate, global, and high-resolution SM estimation is required for many studies, including water resource management, irrigation scheduling, prediction of agricultural yields, and flood forecasting [2], [3]. Microwave remote sensing technologies have been widely used for SM estimation on a global scale [4]. Currently, there are several ongoing satellite missions that provide SM estimation with different spatial and temporal resolutions. The European Space Agency's (ESA) Soil Moisture and Ocean Salinity (SMOS) [5] and the National Aeronautics and Space Administration's (NASA) Soil Moisture Active Passive (SMAP) [6] are both dedicated SM missions that are instrumented with L-band passive radiometers and provide 36-km spatial and 1-3 days temporal resolution. Sentinel-1 is another ESA mission equipped

This work was sponsored by USDA-ARS Award NACA 58-6064-9-007.

V. Senyurek is with Geosystems Research Institute and , A.C. Gurbuz and M. Kurum are with the Department of Electrical and Computer Engineering, and Information Processing and Sensing Lab, Mississippi State University, Mississippi State, MS 39672 USA (e-mail: volkan@gri.msstate.edu; gurbuz@ece.msstate.edu; kurum@ece.msstate.edu; ).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier XX/JSTARS.2020.XX

with C-band synthetic aperture radar, and it can facilitate SM mapping with 1-km spatial and 6-12 days temporal resolution [7].

Global Navigation Satellite System-Reflectometry (GNSS-R) is a promising, alternative microwave technique that has been recently gaining high interest within the scientific community as it has the potential to provide higher spatio-temporal resolution observations than conventional, passive microwave remote sensing methods. This bistatic forward scattering approach collects reflected signals from the Earth's surface, covering the space-time gap of the current traditional active/passive monostatic satellite missions [8]. NASA's Cyclone Global Navigation Satellite System (CYGNSS, launched in December 2016) is primarily an ocean surface wind mission to improve extreme weather prediction, but it operates continuously over both land and ocean from  $38^\circ$  North to  $38^\circ$  South latitudes, providing many land observations with clear sensitivity to changing land surface conditions. Many studies have recently taken advantage of the large amount of CYGNSS land observations to develop models and algorithms to retrieve SM from CYGNSS observations at various spatial and temporal scales [9]–[14].

CYGNSS acquires GNSS-R measurements from 32 channels with 8 small satellites and 4 channels per satellite. Its mean revisit time can be as short as seven hours with 25-km spatial resolution over the ocean under a dominantly diffuse scattering regime. The use of a bistatic receiver constellation for GNSS Signals-of-Opportunity (SoOp) measurements introduces many factors which should be considered when handling this dataset. Because of the ever-changing geometries of the GNSS transmitters and CYGNSS satellites, CYGNSS observations have a quasi-random sampling characteristic which increases the measurement uncertainty when compared to the repeatable, swath-like sampling of traditional remote sensing satellites. Additionally, the reflections of GNSS signals from land are much more diverse and are a composite of coherent and/or incoherent scattering depending on surface conditions. This makes the spatial resolution of the CYGNSS observations over land variable. There are several ongoing research activities that attempt to identify relative contributions of coherent and incoherent reflections [15]–[17], yet the actual spatial resolution of GNSS-R measurements is still a subject of debate within the research community. Following previous studies that produce SM products which assume the presence of coherent signals [9], [11], [13], we will also assume coherent reflections over land. Under such assumption, the CYGNSS observations have a relatively high spatial resolution (on the order of a few kilometers), but less frequent observations

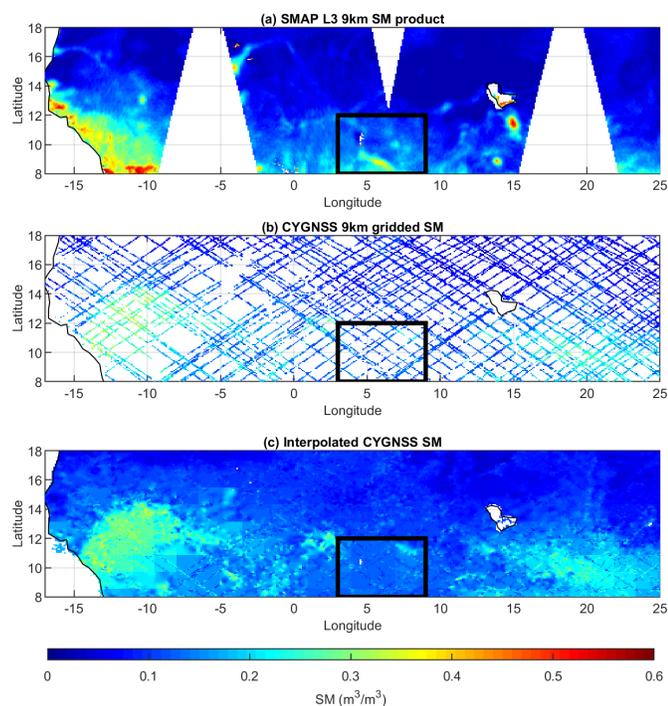


Fig. 1. SM results in Sahel region (a transitional zone) on 06 November 2018. (a) SMAP Enhanced L3 Radiometer 9-km EASE-Grid SM product, (b) Mississippi State University Geosystems Research Institute (MSU-GRI) CYGNSS SM product estimated from quasi-random surface samples of the CYGNSS [14], [18], (c) Interpolated CYGNSS SM estimations with Kriging algorithm. Black rectangular area shows selected region for Fig. 7

(increased revisit time).

To illustrate the spatial sampling capabilities, observations of SMAP, CYGNSS, and interpolated CYGNSS SM on 06 November 2018 are given in Fig. 1. Although the composite SMAP SM product provides 100% coverage on its track after excluding large, static water bodies, the daily SMAP SM product contains many unmeasured regions due to spatiotemporal resolution provided by the SMAP orbit and ground track. On the other hand, it can be seen that CYGNSS provides high-resolution observations over a larger area in a single day, but with sparse sampling. To fill the unsampled grids, spatial or temporal averaging of CYGNSS observations can be performed. However, this degrades the high-resolution potential of CYGNSS-based SM estimation both in space and time. Fig. 2 provides the spatial coverage performance of CYGNSS for different spatial gridding and temporal averaging scenarios. For instance, to achieve a daily product with full spatial coverage (at least one CYGNSS sample within each grid) between  $\pm 38^\circ$  latitudes, a spatial gridding that is greater than a 36-km aggregation is necessary. Also, more than ten days of temporal averaging is required to provide full coverage SM estimations at 9-km spatial resolution. Hence, without further processing, CYGNSS does not provide both a high spatial and temporal resolution product simultaneously.

Furthermore, CYGNSS's revisit or temporal sampling rate varies spatially within its coverage (between  $\pm 38^\circ$  latitudes) as illustrated in Fig. 3. For 9-km  $\times$  9-km aggregated grids,

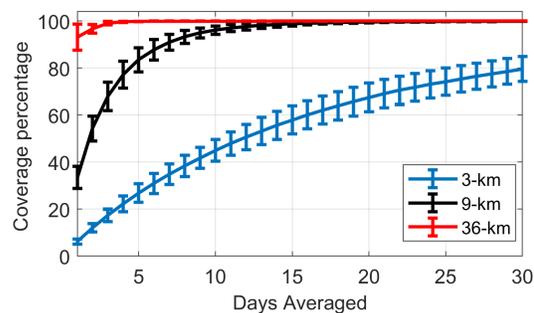


Fig. 2. Coverage performance of CYGNSS in different spatial and temporal resolutions. Each CYGNSS measurement is gridded at 3-km spatial resolution around the specular point assuming a coherent reflection over land. If a pixel contains at least one CYGNSS specular reflection point, it is assumed to be representative of the whole pixel. Bars show one standard deviation around the mean.

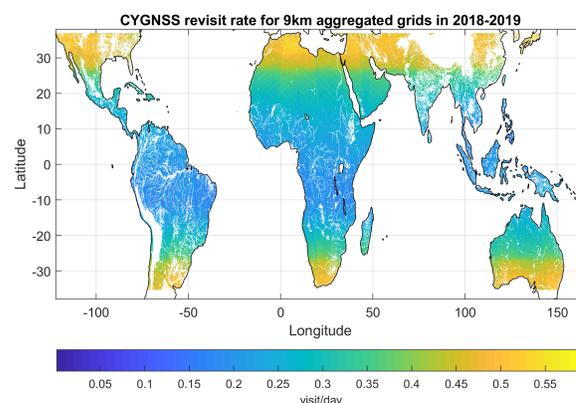


Fig. 3. Revisit map of the 9km aggregated CYGNSS samples for the period of 2018 and 2019.

the temporal sample rate of CYGNSS can vary between 0.5 and 0.1 visits/day. While it will take 2 days for CYGNSS to provide an observation for locations above  $\pm 30^\circ$  on average, it takes 3-10 days for locations between  $+30^\circ$  and  $-30^\circ$  latitudes for 9-km  $\times$  9-km grids. Due to the spatially changing observation rate of CYGNSS, the averaging requirements differ for different regions of the world.

As an alternative to spatial and/or temporal averaging, which are often employed in retrieval algorithms, we investigate the use of interpolation methods to provide CYGNSS-based SM estimates in unmeasured regions of CYGNSS's ground track. This approach has the potential to generate a high resolution SM product for both the spatial and temporal domains as illustrated in Fig. 1(c). Spatial and temporal interpolation of SM data is not new, and there have been some studies focusing on spatial interpolation/extrapolation of SM measurements for continuous data [19]–[21]. However, these studies were performed using *in-situ* measurements for limited areas.

The idea of spatial interpolation of CYGNSS reflectivity observations was also presented using a previously-observed behavior interpolation (POBI) by Chew [22]. Our previous study performed spatial and spatiotemporal interpolation of the CYGNSS SM estimates using linear, nearest, and natural

neighbor interpolation techniques [23]. The present study extends our previous work by incorporating ordinary Kriging [24], inverse distance weighting (IDW) [25], best linear unbiased estimation (BLUE) [26] and POBI [22] techniques for spatial interpolation and analyzes the interpolation performance for different land cover and geographic factors.

The purpose of this study is two-fold. First, we assess error sources in the spatial interpolation of CYGNSS-based SM data to obtain a daily SM product within the full CYGNSS coverage (between  $\pm 38^\circ$  latitudes). To accomplish this, we utilize the SMAP Enhanced L3 Radiometer Global Daily 9-km dataset as a testbed since its 1000-km swath width provides fully covered regions on its track. It is imperative to note that this SMAP product is a 9-km gridded version of the original SMAP measurements, and the native resolution is still 36-km [27]. In this scheme, CYGNSS spatial locations are used to sample the SMAP data and the interpolated results are compared with the actual SMAP observations to assess the errors due to the utilized interpolation approaches. Second, we interpolate sparse CYGNSS SM estimates directly and compare the resultant estimates against SMAP and sparse *in-situ* SM networks.

The rest of the paper is organized as follows: Section II summarizes datasets used. Details on our assessment approach and the utilized interpolation algorithms are provided in Section III. Results and discussions are provided in Section IV. Finally, conclusions are drawn in Section V.

## II. DATASET

### A. MSU-GRI CYGNSS-derived surface SM product

In this study, we use our CYGNSS-derived 9-km gridded SM product, MSU-GRI surface SM v1.0a, which is publicly available in [18]. This product was generated by a machine learning model that uses CYGNSS land observations (reflection power, incident angle, and trailing edge slope) and ancillary inputs (elevation, soil texture, difference vegetation index, and vegetation water content). We used the 1-km resolution Digital Elevation Model GTOPO30 product from the United States Geological Survey Earth Resources Observation and Science archive for surface elevation information (<https://lpdaac.usgs.gov/products/myd13a1v006/>). The presence of surface inland water body is identified by utilizing a 30-m Global Surface Water Dataset from the Joint Research Centre (GSW-JRC) [28]. Global soil texture data from the Global Gridded Soil Information [29] were used to characterize the spatial variations of soil type. Several Moderate Resolution Imaging Spectrometer (MODIS) land surface products were used to characterize the underlying vegetation conditions [30]. The model was trained using only *in-situ* SM data from 170 International Soil Moisture Network (ISMN) sites and is independent of SMAP observations in its training. The produced SM product was then successfully validated using both SMAP enhanced L3 SM product and ISMN's SM measurements. The details of the ancillary data and algorithm behind the MSU-GRI product can be found in [14].

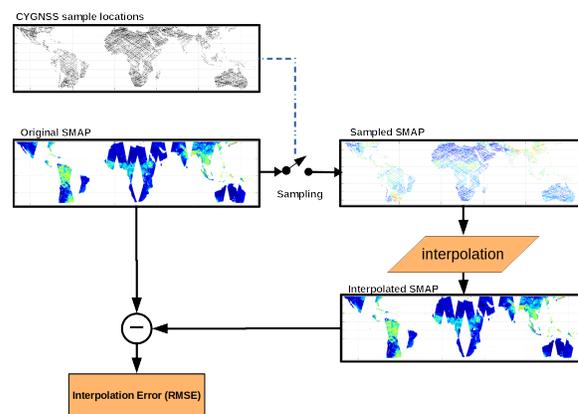


Fig. 4. Interpolation error calculation diagram. CYGNSS observation locations spatially sample the SMAP data. The sampled SMAP data were interpolated then reconstructed data was compared with original SMAP data to obtain the interpolation error.

### B. SMAP SM

The SMAP Enhanced L3 Radiometer Global Daily 9-km EASE-Grid SM product is used here for error analysis of the interpolation approaches. This product is an enhanced version of the 36-km radiometer SM product using Backus–Gilbert optimal interpolation [31]. SMAP datasets contain SM estimation and the associated coordinates for both the descending (A.M.) and ascending (P.M.) overpasses. Both passes are combined to obtain daily 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 SM retrieval quality flags that indicate whether the SM retrieval is recommended or not. SMAP SM estimations can have an uncertain quality for several reasons such as water body fraction, coastal proximity, urban area, precipitation, slope, vegetation water content. The data is freely available through the National Snow and Ice Data Center (NSIDC) at [https://nsidc.org/data/SPL3SMP\\_E/versions/3](https://nsidc.org/data/SPL3SMP_E/versions/3).

## III. METHODOLOGY

Interpolation can be applied to either CYGNSS reflectivities before SM predictions as in [22] or directly on the CYGNSS-derived SM products as done here. In this study, we first use the locations of the CYGNSS SM product for sampling the SMAP L3 9-km gridded data set in order to assess the interpolation performance of well-known interpolation algorithms (listed in subsection III-A) under the CYGNSS quasi-random observation scheme. We then interpolate the MSU-GRI CYGNSS SM product directly to produce a full-coverage SM product at 9 km, which is independently evaluated against both SMAP's SM product and *in-situ* observation at various ISMN sites. Due to the SMAP's native 36-km resolution [27], it is expected that interpolation error for CYGNSS SM estimations against 9-km gridded version of the original SMAP measurements could vary depending on the heterogeneity of the surface being interpolated.

Both SMAP SM retrievals and CYGNSS SM retrievals within the time period of 18 March 2017 to 31 December 2019 are arranged as  $1019 \times 920 \times 3200$  (Days  $\times$  Latitude  $\times$

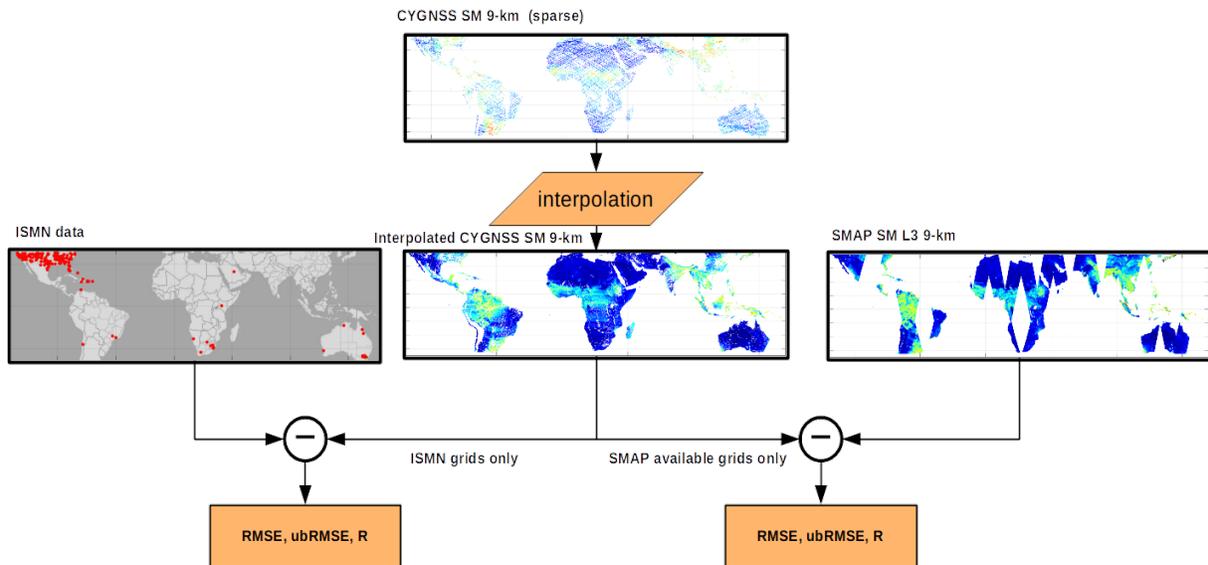


Fig. 5. Comparison of interpolated CYGNSS SM with SMAP SM and in-situ measurements.

Longitude) matrices. Then, we followed the same quality controls described in [14]. For instance, the samples that correspond to water bodies and high elevation ( $> 2000$  m) were removed from both CYGNSS and SMAP datasets. In order to analyze the interpolation error, daily SMAP SM estimates whose EASE grid center are closest to CYGNSS specular point locations were sampled. Then the resulting sparse SMAP daily SM data was interpolated independently using the different algorithms discussed in subsection III-A. A total of 1019 days of spatial interpolation was performed, and an interpolated SMAP SM matrix was generated. The interpolated and original SMAP SM estimations were compared to calculate the interpolation error over the EASE grids only if SMAP has measurement at that grid location. Root mean square error (RMSE) metric is used to quantify the interpolation error. The interpolation and error calculation process is illustrated in Fig. 4.

After the error analysis, we study direct interpolation of the CYGNSS SM product. This includes evaluations of interpolated CYGNSS SM values independently against overlapping SMAP SM data with grid-based performance metrics for both (1) all grids and (2) recommended SMAP grids. In addition to error analysis comparing SMAP and CYGNSS, an error analysis is performed using individual sparse *in-situ* SM networks as an independent metric for the CYGNSS SM interpolation methods. Fig. 5 illustrates the comparison approach.

#### A. Interpolation algorithms

In the study, we used well-known spatial interpolation methods such as linear [32], natural neighbor [33], IDW [24], Kriging [34], and best linear unbiased estimation (BLUE) [26] and POBI [22] for daily SM estimations. These interpolation approaches use a weighted combination of the sparse obser-

vations as

$$\hat{u}(\mathbf{x}_0) = \sum_{i=1}^N \lambda_i u(\mathbf{x}_i), \quad (1)$$

where  $\hat{u}(\mathbf{x}_0)$  is the predicted/interpolated value for the unsampled point  $\mathbf{x}_0$ ,  $u(\mathbf{x}_i)$  is the SM observations at locations  $\mathbf{x}_i$ , and  $\lambda_i$  are the weights for each  $u(\mathbf{x}_i)$  value where  $N$  is the number of observations within the window of interpolation.

1) *Linear*: Classical linear interpolation fits a plane between measured samples to estimate the unknown values. For spatial interpolation, we performed the Delaunay triangulation-based [35] linear interpolation. In this approach the grid is subdivided into a series of triangles such that the measured points are connected with lines in such a way that no triangle edges are intersected by other triangles. Once the triangulation is done, a bivariate linear interpolation is applied within each triangle to find the unknown pixel value using the ones measured at the three edges of the triangle. This approach is an exact interpolator and implemented with MATLAB's *griddata* function.

2) *Inverse Distance Weighting (IDW)*: The inverse distance weighting method estimates the unsampled point value using a linear combination of sampled points values and their corresponding weights. The weights are calculated based on the exponentiated Euclidean distance between sampled and unsampled points. IDW weights with low Euclidean distance values will have more influence than weights which feature high Euclidean distance. The IDW estimation of an unknown point is given by

$$\hat{u}(\mathbf{x}_0) = \frac{\sum_{i=1}^N u(\mathbf{x}_i) d_i^{-p}}{\sum_{i=1}^N d_i^{-p}} \quad (2)$$

where  $\hat{u}(\mathbf{x}_0)$  is the interpolated value at any location  $\mathbf{x}_0$ ,  $u(\mathbf{x}_i)$  are the sampled values ( $i = 1, 2, \dots, N$ ),  $d_i$  is the distance between  $\mathbf{x}_0$  and  $\mathbf{x}_i$ ,  $p$  is the a power parameter and  $N$  is the number of sample points within the interpolation

window used for the estimation. IDW is implemented with window sizes of 5, 7 or 9 and  $p = 3$  is used as the power parameter.

3) *Natural Neighbor*: Like IDW, the natural neighbor interpolation method [33] is a weighted-average interpolation technique. The method is based on Voronoi tessellation of a discrete set of spatial points and provides smoother approximation. The method is based on Voronoi tessellation of a discrete set of spatial observation points, where basically a partition of a plane into multiple regions that are close to each observation point is done. The partitioning of the plane is updated after inserting the interpolation point to the list of all spatial points. The weights for each measured sample are calculated by using the intersection area of the new partitioning cell with the existing areas of the cells of the closest samples. For the presented results in this paper, natural interpolation is implemented utilizing MATLAB's *griddata* function using the natural neighbor option.

4) *Kriging*: Kriging [24] or Gaussian process regression is another weighted spatial interpolation method. In Kriging, the weights are based on the overall spatial arrangement among the measured points which is quantified through the spatial correlation between the sample points. Thus, in Kriging, the weight depends on a fitted model of the field semi-variogram derived from the distance to the unsampled point, and the spatial relationships among the measured points around the unsampled point. Assuming an average mean adds the constraint of  $\sum_{i=1}^N \lambda_i = 1$  to the system. The Kriging equation is expressed as:

$$\sum_{i=1}^N \lambda_i \gamma(\mathbf{x}_i, \mathbf{x}_j) + \mu = \gamma(\mathbf{x}_i, \mathbf{x}_0) \quad (3)$$

where,  $\gamma(\mathbf{x}_i, \mathbf{x}_j)$  is the semivariogram between observation point  $\mathbf{x}_i$  and  $\mathbf{x}_j$ ,  $\gamma(\mathbf{x}_i, \mathbf{x}_0)$  is the semivariogram between sampling point  $\mathbf{x}_i$  and interpolation point  $\mathbf{x}_0$ , and  $\mu$  is the Lagrange multiplier parameter. Kriging equations can be written in the form of a matrix:

$$\begin{bmatrix} \gamma(\mathbf{x}_1, \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_1, \mathbf{x}_N) & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(\mathbf{x}_N, \mathbf{x}_1) & \cdot & \gamma(\mathbf{x}_N, \mathbf{x}_N) & 1 \\ 1 & \cdot & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(\mathbf{x}_1, \mathbf{x}_0) \\ \vdots \\ \gamma(\mathbf{x}_N, \mathbf{x}_0) \\ 1 \end{bmatrix} \quad (4)$$

In the study, a Gaussian model fitted semivariogram function was used for each segmented windows (for detail see section III-B). After solving Eq. (4) the weights  $\lambda$  will be obtained, and they can be used in Eq. (1) to estimate the value at the unsampled point.

5) *Best Linear Unbiased Estimation (BLUE)*: BLUE [26] is an unbiased linear estimator and it has minimum variance among all other linear estimators. Since it is a linear estimator, the unknown point can be represented as a linear combination of measured points as in Eq. (1). The unbiased constraint states that  $E(\hat{u}(\mathbf{x}_0)) = \sum_{i=1}^N \lambda_i E(u(\mathbf{x}_i)) = u(\mathbf{x}_0)$ . The variance on the estimate  $\hat{u}(\mathbf{x}_0)$  is given by  $var(\hat{u}(\mathbf{x}_0)) = \lambda^T C \lambda$  where  $C$  is the covariance of the measured points. The BLUE algorithm minimizes the estimate variance constraint to the unbiased property. In order to satisfy the unbiased constraint,

$E(u(\mathbf{x}_i))$  should be linear in  $\hat{u}(\mathbf{x}_0)$  as  $E(u(\mathbf{x}_i)) = s_i \hat{u}(\mathbf{x}_0)$  where  $s$  is the scaled mean. The solution to the variance minimization will result in the optimal weights as

$$\lambda = \frac{C^{-1} s}{s^T C^{-1} s} \quad (5)$$

BLUE requires the covariance and the scaled mean to estimate the SM at the unknown point. In our implementation both  $C$  and  $s$  are estimated from all historical data over the study period of 2017-2019.

6) *Previously-observed behavior interpolation (POBI)*: This interpolation technique interpolates an unsampled point by quantifying how past observations of the quantity at the unsampled location varies with the past observations with the nearby points [22]. This relationship is defined by a linear regression model between the unsampled location and each nearby point separately using the past observations of the unsampled point and the nearby points. The estimated SM value at the unsampled point can be expressed as

$$\hat{u}(\mathbf{x}_0) = \frac{\sum_i^N \lambda_i (a_i u(\mathbf{x}_i) + b_i)}{\sum_i^N \lambda_i}, \quad (6)$$

where  $N$  is the number of neighbor pixels,  $a_i$  and  $b_i$  are the slope and intercept parameters of the regression line applied on previous measurements at locations  $\mathbf{x}_i$  and  $\mathbf{x}_0$ .  $\lambda_i$  are the weights for each  $u(\mathbf{x}_i)$ . Where  $\lambda_i$  is the square of the correlation coefficient between  $u(\mathbf{x}_0)$  and  $u(\mathbf{x}_i)$  computed over the past observations. In [22], the POBI is applied on the CYGNSS reflectivity values. Instead, in this paper, interpolation over the SM observations is implemented using the same approach via (6).

## B. Implementation of the algorithms

Each interpolation method can work on a different size region or apply varying weighting mechanisms for an optimal performance. Next we provide the implementation details of the presented interpolation approaches:

- For IDW, BLUE, and POBI, a window size of  $n \times n$  (e.g.,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ ) pixel is used to interpolate an unknown point. Here, the unknown point is in the center of the window. The sampled measurements within the  $n \times n$  window are used to calculate unknown points. This procedure is repeated for each unknown point in the dataset.
- For the Kriging interpolation algorithm, a moving window size of  $24 \times 24$  is utilized. The window is moved horizontally and vertically with 4 pixels overlapping. First, a Gaussian fitted semivariogram model was estimated for each  $24 \times 24$  window using sampled SMAP SM data. These semivariograms were calculated for each day independently using the sparse observations within the specified window size. The number of data for semivariogram calculation can vary between 170 - 310 depending on the number of CYGNSS observation in the region. In this study, a constant window size of  $24 \times 24$  is utilized since it provides a good tradeoff between the number of available measurements and the local information. It

should be noted that using a small size window makes it difficult to obtain a proper fit for the semivariogram model because of decreased number of observations. However, increasing the size of the window will degrade the semivariogram model because of spatial heterogeneity of SM. All unsampled points within the  $24 \times 24$  window are interpolated using the detailed Kriging approach.

#### IV. RESULTS AND DISCUSSION

In this section, the performance of the aforementioned interpolation approaches have been analyzed. First, spatial heterogeneity within the SMAP 9-km gridded data, the MSU-GRI CYGNSS SM data, and the CYGNSS reflectivity is examined. The interpolated images from a selected region is then illustrated. Then, the overall interpolation error of the tested algorithms over the whole CYGNSS observation region and time-span is presented. We then provide more detailed analysis of the observed interpolation error with respect to CYGNSS revisit rate, geographical distribution, land cover, SMAP retrieval quality flags, and number of annual CYGNSS observations. Finally, the overall performance of interpolated CYGNSS SM against SMAP and *in-situ* SM products were reported.

##### A. Heterogeneity Check

With the error analysis using SMAP data, we implicitly assume that the interpolation error metrics calculated using SMAP as a proxy for CYGNSS will be similar to what the true interpolation errors would be when interpolating CYGNSS SM. To assess the validity of this assumption, spatial heterogeneity maps for SMAP 9-km gridded data, the MSU-GRI CYGNSS SM data and CYGNSS reflectivity for the year of 2019 were generated for each EASE grid. This was achieved by calculating the standard deviation of the SM estimates and reflectivity for various window sizes (i.e.,  $45 \times 45$  km,  $63 \times 63$  km,  $81 \times 81$  km, and  $99 \times 99$  km) and averaging throughout 2019. The comparisons are provided in Table I that provides the overall SM standard deviation for different calculation sizes and different land cover types. In addition, an example map is provided in Fig. 6. The results indicate that SMAP L3 9-km SM is more heterogeneous than CYGNSS SM except for shrublands and barren areas. This results, at first, might sound counter intuitive as one expects more variations in CYGNSS than SMAP due to CYGNSS's sensitivity to higher spatial scales and SMAP's 36 km native resolution. It is clear that the CYGNSS reflectivity has very large variations (up to 8 dB standard deviation) as seen in Fig. 6 (c). However, the sensitivity of CYGNSS reflectivity stems mostly from the presence of inland water bodies. On the other hand, the MSU-GRI CYGNSS SM algorithms explicitly removes transient open water fractions larger than 2% within the first Fresnel zone while the baseline SMAP retrieval algorithm corrects for large static water bodies only. This usually makes the dynamic range of SMAP larger than that of CYGNSS SM estimates (e.g., see Figs. 10-13 of [14]). To summarize, while there are some inherent differences in the heterogeneity of both SM products, both data sets follow similar spatial patterns

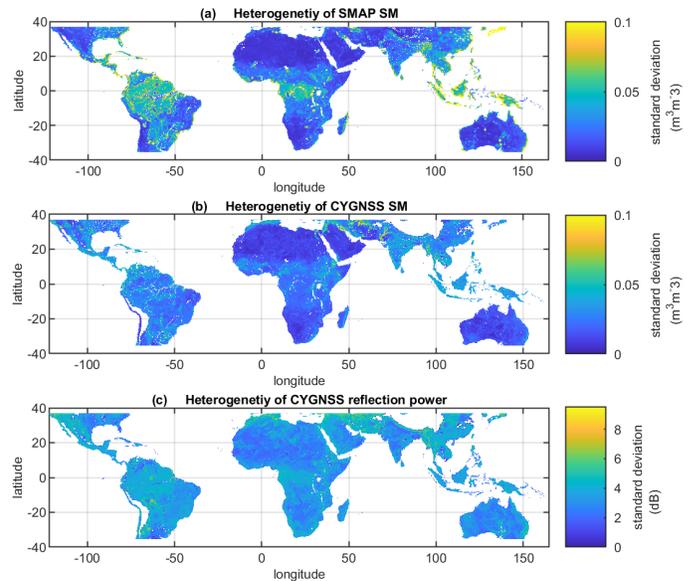


Fig. 6. Heterogeneity map of (a) SMAP L3 9-km SM (b) MSU-GRI CYGNSS 9-km SM (c) CYGNSS reflectivity 9-km.

and SMAP is found to be a good candidate as a proxy for CYGNSS interpolation testing.

##### B. Illustration over a Selected Region

Fig. 7 shows the original and sparsely sampled SMAP SM observations as well as the interpolated results for the tested approaches over a selected region ( $8N:12N, 3E:9E$ ) within the Sahel on 06 November 2018. Again, CYGNSS points are only used to spatially sample the original SMAP SM estimations and it covers only 21 % of the whole image's pixels for this particular example. The sampled SMAP image shown in Fig. 7(b) is then interpolated by tested algorithms and the obtained interpolated SM results and residuals are shown in Fig. 7(d-h) and (i-n), respectively. The figure indicates that all interpolation algorithms demonstrated high correlation with the original SMAP SM ( $R^2 = 0.92$  to  $0.95$ ). In addition, we calculate the root mean squared error (RMSE), average absolute error (AAE), and maximum absolute error (MAE) over only the interpolated pixels. For the illustrated example area in Fig. 7, RMSE varies from  $0.015 \text{ m}^3\text{m}^{-3}$  to  $0.020 \text{ m}^3\text{m}^{-3}$  and the performance of each technique is specific to this particular region. The MAE is found to be between  $0.097 - 0.18 \text{ m}^3\text{m}^{-3}$ , while AAE changes between  $0.008 - 0.012 \text{ m}^3\text{m}^{-3}$  depending on interpolation algorithms. In order to provide the overall performance of the tested approaches, we now provide average results for the entire CYGNSS coverage.

##### C. Overall Interpolation Performance

This section presents the overall interpolation error (RMSE) by comparing original and interpolated SMAP data observations using different algorithms over the whole CYGNSS coverage area. The comparison was performed for all SMAP observation grids and SMAP recommended grids separately.

TABLE I  
HETEROGENEITY COMPARISON OF SMAP L3 9-KM AND CYGNSS 9-KM SM PRODUCTS FOR 2019.

	Overall Std Dev. for different sizes [km <sup>2</sup> ]				Overall Std Dev. for land cover types (63×63 [km <sup>2</sup> ])						
	45 × 45	63 × 63	81 × 81	99 × 99	forest	shrublands	woody	savanna	grass	croplands	barren
SMAP L3 SM [m <sup>3</sup> m <sup>-3</sup> ]	0.024	0.028	0.032	0.035	0.054	0.013	0.045	0.032	0.023	0.032	0.011
CYGNSS SM [m <sup>3</sup> m <sup>-3</sup> ]	0.021	0.023	0.025	0.026	0.029	0.017	0.030	0.027	0.023	0.031	0.014
CYGNSS Reflectivity [dB]	2.43	2.92	3.26	3.56	3.32	2.99	3.13	2.88	2.67	2.66	2.90

TABLE II  
INTERPOLATION ERROR OF SMAP 9-KM EASE-GRID SM PRODUCT. SMAP RETRIEVAL QUALITY FLAG (RQF) INDICATES WHETHER SM RETRIEVAL IS RECOMMENDED OR NOT RECOMMENDED (UNCERTAIN).

Algorithm	Interpolation error RMSE [m <sup>3</sup> m <sup>-3</sup> ]					
	All Grids	SMAP Recom.	All Grids	SMAP Recom.	All Grids	SMAP Recom.
Linear	0.032	0.016				
Natural	0.030	0.015				
Kriging	0.030	0.014				
Window size/coverage rate	9×9 / 97%		7×7 / 94%		5×5 / 87%	
IDW	0.032	0.016	0.031	0.015	0.028	0.013
BLUE	0.025	0.014	0.023	0.012	0.022	0.010
POBI	0.026	0.013	0.024	0.011	0.023	0.010

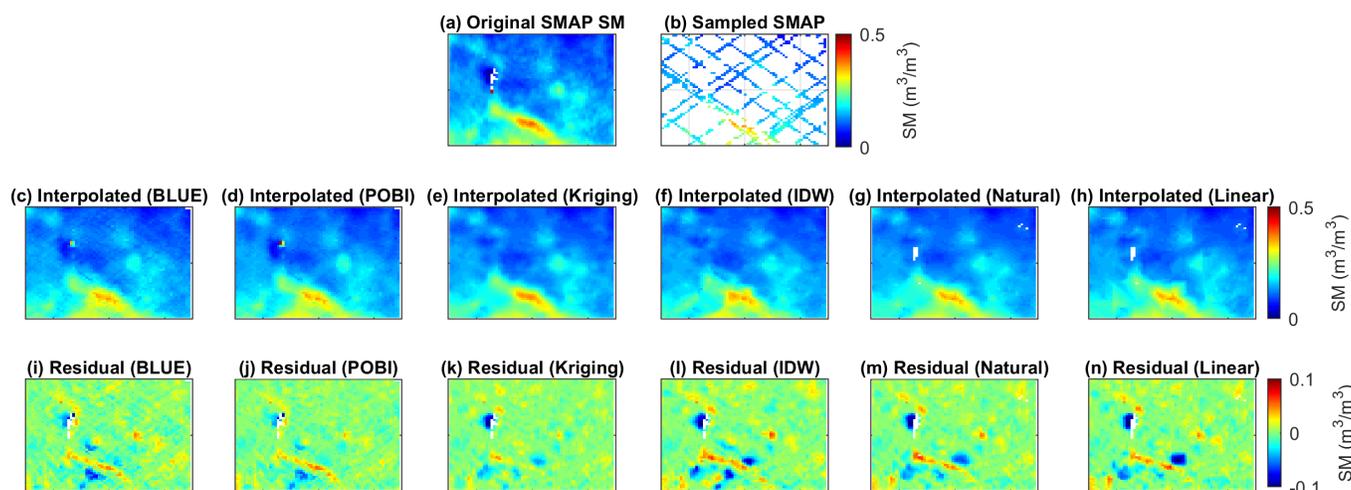


Fig. 7. Examples of the methodology for selected region (8N:12N,3E:9E) on 06 November 2018. (a) original SMAP SM, (b) sampled SMAP SM, (c-h) interpolated SMAP SM by, BLUE (RMSE=0.017m<sup>3</sup>m<sup>-3</sup>, R=0.94, MAE=0.118, AAE=0.011), POBI (RMSE=0.016m<sup>3</sup>m<sup>-3</sup>, R=0.95, MAE=0.181, AAE=0.009), Kriging (RMSE=0.015m<sup>3</sup>m<sup>-3</sup>, R=0.95, MAE=0.097, AAE=0.008), IDW (RMSE=0.020m<sup>3</sup>m<sup>-3</sup>, R=0.92, MAE=0.112, AAE=0.012), natural neighbors (RMSE=0.018m<sup>3</sup>m<sup>-3</sup>, R=0.93, MAE=0.010, AAE=0.01), and linear (RMSE=0.019m<sup>3</sup>m<sup>-3</sup>, R=0.92, MAE=0.115, AAE=0.011), respectively. (i-n) shows residuals of interpolation algorithms.

For IDW, BLUE, and POBI algorithms, the different sizes of interpolation windows were tested, while linear and natural interpolations are applied over the whole region at once.

Table II shows the spatial interpolation errors of different methods over the SMAP SM estimations. The results indicate the RMSE of SM samples may vary between 0.025 m<sup>3</sup>m<sup>-3</sup> and 0.032 m<sup>3</sup>m<sup>-3</sup> when computed over all grids. The BLUE and POBI interpolation algorithms reached a score of 0.025 m<sup>3</sup>m<sup>-3</sup> and 0.026 m<sup>3</sup>m<sup>-3</sup>, respectively. However, these two algorithms are the only algorithms that uses prior statistics of SM values. For the other methods, the average interpolation error is around 0.030 m<sup>3</sup>m<sup>-3</sup>. If we only consider SMAP's recommended grids the interpolation errors are much lower and vary between 0.013 and 0.016 m<sup>3</sup>m<sup>-3</sup>, with smaller performance difference between interpolation algorithms. While

using smaller window sizes for IDW, BLUE and POBI, a decrease in the achieved RMSE values is observed, and the coverage rate (ratio between the number of sampled or interpolated grids and the total number of grids on land within ±38° latitudes) for smaller windows is also less. For example, while using a window size of 5 × 5 provides an average of 87% spatial coverage, a 9 × 9 window provides near full coverage of 97% as shown in Table II. It should be emphasized that the RMSE values presented in Table II are only errors due to the interpolation, they are not absolute SM errors. If the SMAP SM values were correct, the interpolation approaches would contribute to the presented error levels in Table II.

TABLE III  
EVALUATION OF THE INTERPOLATION ERROR FOR DIFFERENT LAND COVER TYPES.

SMAP RQFs	All Grids		Recommended Grids	
	RMSE	#of samples	RMSE	#of samples
Land cover type				
Forest	0.032	17M	0.019	9M
Closed shrublands	0.013	1M	0.011	0.8M
Open shrublands	0.011	19M	0.010	16M
Woody	0.031	5.4M	0.021	2.9M
Savanna	0.026	14M	0.020	11M
Grass	0.019	29M	0.016	25M
Croplands	0.023	13M	0.018	9.6M
Cropland natural veg.	0.028	1.6M	0.018	1M
Barren	0.009	54M	0.008	49M

#### D. Analysis of Interpolation Errors

In this section, the results obtained only using the BLUE algorithm (without loss of generality) are considered and analyzed. The spatial distribution of the interpolation error over the entire CYGNSS coverage is provided in Fig. 8. It indicates that interpolation error is very low (less than  $0.01 \text{ m}^3\text{m}^{-3}$ ) in dry regions such as North Africa, Australia, and the Western U.S. However, the error dramatically increases for regions with substantial seasonal variations, such as south India, the north part of South America, and the equatorial band. These regions are frequently flagged by SMAP due to uncertainties in SM retrievals.

1) *The Effect of SMAP Retrieval Quality Flag (RQF):* While we use the SMAP L3 SM estimation as a reference SM value, its retrieval performance varies with respect to several factors such as water body fraction, coastal proximity, urban area, precipitation, terrain slope, and high vegetation water content. Uncertain RQFs generally belong to Amazon, central Africa, and seasonal flooding regions (e.g., Southeast Asia). Fig. 8 shows that the interpolation error is high for the region where SMAP retrieval is not recommended. We further divided the overall interpolation errors in Table II for all SMAP grids as well as only SMAP recommended grids. For all compared techniques, the interpolation errors for SMAP recommended grids are consistently much lower. The possible reasons are: (1) The regions where the SMAP RQFs are "recommended" generally have dry and homogeneous SM characteristics. (2) Because of low VWC, SMAP SM measurements are more consistent in time and spatial domain. However, SM variation in the regions where the SMAP RQFs are "not recommended" is high, and the retrieval system has its own error due to high VWC.

2) *The Effect of Land Cover:* In Table III, we show the interpolation errors with respect to different land cover types. Open regions such as open/closed shrublands and barren land cover are generally associated with relatively limited SM variation. Thus, the interpolation is more successful, and errors are relatively small ( $< 0.008 \text{ m}^3\text{m}^{-3}$ ) than other land cover regions. For the relatively high vegetation land cover areas (woody, savanna, grass, and cropland), the interpolation error is calculated as  $0.016 \text{ m}^3\text{m}^{-3}$ . Spatial and temporal SM variation is relatively high for these areas, and this increases interpolation error.

3) *Effect of CYGNSS Revisit Distribution:* As mentioned in the introduction section (see Fig. 3), CYGNSS has unequal sampling characteristics. The sampling rate is high near  $37^\circ$  latitudes while it decreases near the equator regions. Figure 9 shows the interpolation error changes by revisit rate of CYGNSS for different land cover types. The general trend for all land cover types is that interpolation error decreases with the increase of the CYGNSS revisit rate. In Fig. 9 barren and shrublands' relatively flat characteristics indicate that SM's spatial variation is low for these land covers. However, other land cover types' revisit rates for CYGNSS become more effective for interpolation performance due to the relatively high spatial SM variation since the performance gets better as revisit rate increases for these land types.

4) *Effect of Number of Annual CYGNSS Observations:* The availability of daily CYGNSS satellite data varies each year. For instance, there are a total of 2089 files for 2017, 2588 files for 2018, and 2910 files for 2019. Each file represents daily observations acquired by one of 8 satellites. When we plot the daily interpolation error distribution as a function of the daily CYGNSS coverage rate for all grids in CYGNSS coverage region in Fig. 10, we observe that interpolation error is decreasing with the increasing daily coverage rate, which is defined as the ratio of the number of 9-km EASE grids that have at least one CYGNSS observation during the considered time period and the total number of 9-km EASE grids on land within  $\pm 38^\circ$  latitudes. This is well aligned with the fact that as more samples are available for each day, lower interpolation errors can be achieved.

#### E. The Performance of Interpolated CYGNSS SM against SMAP and ISMN sites

In this section, we interpolate the CYGNSS SM product [18] directly to produce a fully-covered SM product at 9-km over the entire CYGNSS land coverage. We then evaluate independently against overlapping SMAP SM data for the "all grids" and "recommended SMAP grids" evaluation datasets by calculating the grid-based performance metrics. The RMSE and ubRMSE are computed for each EASE grid using all available samples during the study period and the mean of all grids are presented in Table IV. This includes the overall performance scores for sparse grids, interpolated grids only, and all grids. The all grids case includes both the original sparse CYGNSS observation and the interpolated SM observations to create a spatially complete, CYGNSS-based SM product. All cases are compared against SMAP observations that are available at the same grids on the same days. The obtained scores for all cases are similar. This indicates that the interpolated CYGNSS SM data has similar errors compared to SMAP as the CYGNSS SM observations are computed from CYGNSS measurements. Additionally, there is a relatively lower ubRMSE for interpolated CYGNSS estimations. This is because the average number of samples per grid of the interpolated CYGNSS product is higher than the un-interpolated product as given in Table IV.

We also compare interpolated CYGNSS SM estimations against the *in-situ* SM measurements at 170 ISMN stations.

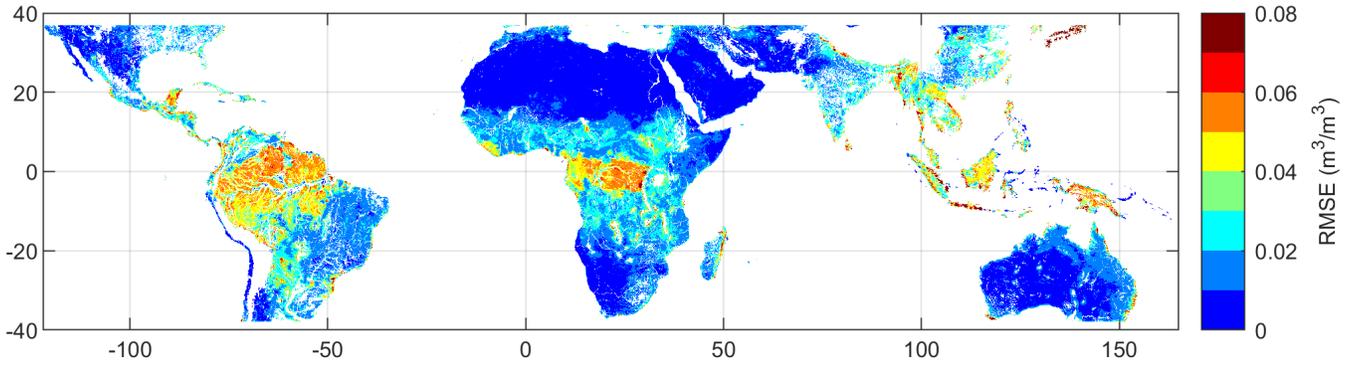


Fig. 8. Interpolation error (RMSE) map for 9-km gridded SMAP SM over the entire CYGNSS coverage region. The interpolation is performed using BLUE algorithm.

TABLE IV  
PERFORMANCE METRICS OF CYGNSS SM AND INTERPOLATED CYGNSS SM ESTIMATIONS AT 9 KM EASE GRIDS

	CYGNSS / SMAP					
	All Grids			SMAP Recommended Grids		
	RMSE [ $m^3m^{-3}$ ]	ubRMSE [ $m^3m^{-3}$ ]	# of days	RMSE [ $m^3m^{-3}$ ]	ubRMSE [ $m^3m^{-3}$ ]	# of days
All grids	0.091	0.049	880	0.060	0.040	428
Sparse grids	0.093	0.051	281	0.060	0.042	282
Interpolated grids	0.091	0.048	598	0.058	0.040	144

TABLE V  
PERFORMANCE COMPARISON OF CYGNSS SM AND INTERPOLATED CYGNSS SM WITH ISMNS.

<i>in-situ</i> vs. CYGNSS	RMSE [ $m^3m^{-3}$ ]	R value [ $m^3m^{-3}$ ]	mean ubRMSE [ $m^3m^{-3}$ ]	median ubRMSE [ $m^3m^{-3}$ ]
All grids	0.082	0.81	0.056	0.050
Sparse grids	0.078	0.80	0.057	0.054
Interpolated grids	0.084	0.80	0.055	0.049

TABLE VI  
PERFORMANCE COMPARISON OF SPARSE SMAP SM (SAMPLED WITH CYGNSS OBSERVATIONS) AND INTERPOLATED SMAP SM WITH ISMNS.

<i>in-situ</i> vs. SMAP	RMSE ( $m^3m^{-3}$ )	R value ( $m^3m^{-3}$ )	mean ubRMSE ( $m^3m^{-3}$ )	median ubRMSE ( $m^3m^{-3}$ )
All grids	0.109	0.65	0.049	0.047
Sparse grids	0.118	0.63	0.049	0.049
Interpolated grids	0.109	0.66	0.049	0.047

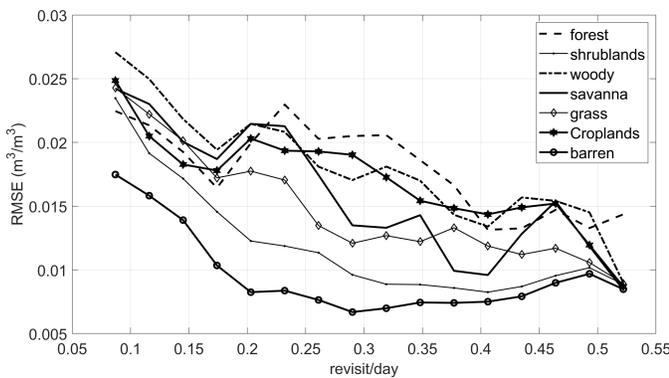


Fig. 9. The relationship between CYGNSS revisits rate and interpolation error for different land cover types.

Table V presents the performance of both the original sparse CYGNSS-based SM data and the interpolated CYGNSS-based

SM data against *in-situ* measurements over the study period. The similar RMSE, ubRMSE and R-value for the original and the interpolated CYGNSS samples are other indicators that the interpolated SM estimations agree with *in-situ* measurements. We also compared sparsely and interpolated SMAP SM data with *in-situ* measurements in Table VI to demonstrate the efficiency of the interpolation algorithm. Again similar performance metrics are obtained for interpolated and sparse SMAP. Thus, the results suggest that interpolated CYGNSS-based SM products can provide a fully covered, daily product with a performance similar to sparse CYGNSS SM estimations against SMAP SM and *in-situ* values.

## V. SUMMARY AND CONCLUSION

This paper assessed interpolated CYGNSS-based SM estimations to generate a fully covered, daily high spatial resolution SM product. The overall results demonstrated that the gaps of quasi-random sampling of the CYGNSS could be filled successfully with the tested interpolation algorithms. The

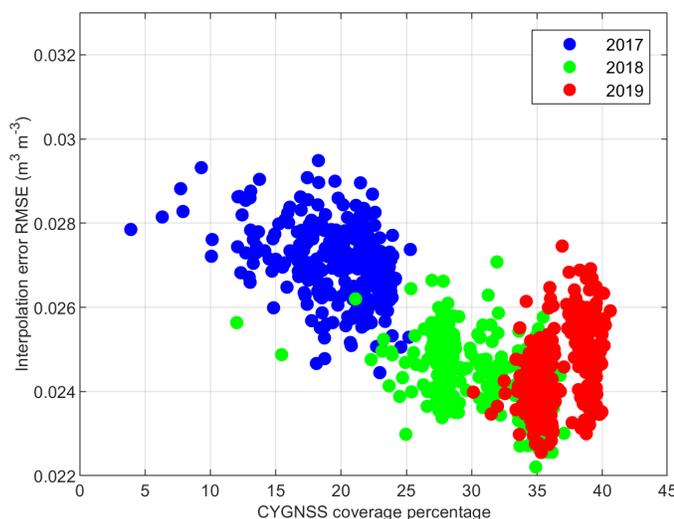


Fig. 10. The relationship between CYGNSS daily coverage percentage and linear interpolation error.

overall interpolation RMSE error computed over the SMAP SM data vary between  $0.013 \text{ m}^3\text{m}^{-3}$  and  $0.016 \text{ m}^3\text{m}^{-3}$  for SMAP recommended grids and  $0.025\text{-}0.032 \text{ m}^3\text{m}^{-3}$  for all grids depending on the selected interpolation algorithm. In addition, when sparse CYGNSS data is interpolated with the BLUE approach and compared with SMAP or ISMN SM measurements, similar SM error levels are obtained for both the original sparse CYGNSS data and the interpolated data which indicates the benefit of the interpolation.

When interpolation errors are analyzed, substantially higher interpolation errors for high vegetation content regions (e.g.,  $> 5 \text{ kg} \cdot \text{m}^{-2}$ ) are observed since the SMAP RQF is uncertain for such regions including the Amazon, central Africa, and seasonal flooded areas. Interpolation error also depends on the revisit rate of CYGNSS. There are slight increments in interpolation error due to the decrement of revisit rate for the same land cover areas. Near the equator (e.g., between  $\pm 10^\circ$  latitudes), the sample rate of CYGNSS is lower, and the interpolation error of these regions remain relatively high. We observe a minor relationship between land cover type and interpolation error. Due to the high spatio-temporal SM variation, croplands, woody, forest, and savanna areas have a relatively higher interpolation errors.

Although this study did not aim to find the best interpolation technique, algorithms that use prior statistics (e.g., covariance and mean) of unsampled points and their neighbors such as BLUE or POBI demonstrated higher performance overall. Simpler techniques such as Linear and IDW showed slightly worse performance than other algorithms with  $0.003 - 0.007$  higher RMSE values. Natural neighbor and Kriging techniques performed similarly. However, different types of variogram models and area size for the variogram calculations may increase the performance of the Kriging method. As a future study, semivariograms of SMAP and CYGNSS sparse retrievals or Fourier-transform-based approaches could be compared to investigate the different spatial scales of the two

data sets. Regional or seasonal comparisons of interpolation algorithms may also be the subject of future studies.

This study used the SMAP L3 9-km gridded SM product to calculate interpolation error. However, this product is already an enhanced version of the 36-km radiometer SM product using Backus–Gilbert optimal interpolation techniques and may not represent real SM values. Because of this, the expected interpolation error could be slightly higher than the results presented in the study. In the study, we used a 9-km gridded CYGNSS product. Readers should consider that interpolation error will vary for the CYGNSS products with different aggregated resolutions. This study used popular spatial interpolation algorithms. The effect of more sophisticated techniques on both temporal and spatial domain still needs to be evaluated and compared. However, this study shows that CYGNSS SM products and interpolation have high potentials to provide daily high-resolution SM products over its quasi-global coverage.

## REFERENCES

- [1] H. Vereecken, J.A. Huisman, H. Bogena, J. Vanderborght, et al., "On the value of soil moisture measurements in vadose zone hydrology: A review," *Water resources research*, vol. 44, no. 4, 2008.
- [2] D. Entekhabi, I. Rodriguez-Iturbe, and F. Castelli, "Mutual interaction of soil moisture state and atmospheric processes," *J. Hydrol.*, vol. 184, no. 1-2, pp. 3–17, 1996.
- [3] M. Jung, M. Reichstein, P. Ciais, S. Seneviratne, J. Sheffield, M. Goulden, et al., "Recent decline in the global land evapotranspiration trend due to limited moisture supply," *Nature*, vol. 467, no. 7318, pp. 951–954, 2010.
- [4] J. Peng and A. Loew, "Recent advances in soil moisture estimation from remote sensing," 2017.
- [5] Y. Kerr, A. Al-Yaari, N. Rodriguez-Fernandez, M. Parrens, B. Molero, et al., "Overview of SMOS performance in terms of global soil moisture monitoring after six years in operation," *Remote Sensing of Environment*, vol. 180, pp. 40–63, 2016.
- [6] D. Entekhabi, E. G. Njoku, P. E. O'Neill, et al., "The soil moisture active passive SMAP mission," *Proc. IEEE*, vol. 98, no. 5, pp. 704–716, 2010.
- [7] Ramon Torres, Paul Snoeij, Dirk Geudtner, David Bibby, Malcolm Davidson, Evert Attema, Pierre Potin, Björn Rommen, Nicolas Floury, Mike Brown, et al., "Gmes sentinel-1 mission," *Remote Sensing of Environment*, vol. 120, pp. 9–24, 2012.
- [8] V. Zavorotny, S. Gleason, E. Cardellach, and A. Camps, "Tutorial on remote sensing using GNSS bistatic radar of opportunity," *IEEE Geoscience and Remote Sensing Magazine*, vol. 2, no. 4, pp. 8–45, 2014.
- [9] 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.
- [10] M. Al-Khaldi, J. Johnson, A. O'Brien, A. Balenzano, et al., "Time-series retrieval of soil moisture using CYGNSS," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4322–4331, 2019.
- [11] 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, pp. 2272, 2019.
- [12] 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 Sensing*, vol. 12, no. 7, pp. 1168, 2020.
- [13] 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 Sensing of Environment*, vol. 247, pp. 111944, 2020.
- [14] 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 Sensing*, vol. 12, no. 21, pp. 3503, 2020.
- [15] Adriano Camps, "Spatial resolution in gnss-r under coherent scattering," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 1, pp. 32–36, 2019.
- [16] Davide Comite and Nazzareno Pierdicca, "Decorrelation of the near-specular land scattering in bistatic radar systems," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

- [17] Ahmed M Balakhder, Mohammad M Al-Khaldi, and Joel T Johnson, "On the coherency of ocean and land surface specular scattering for gnss-r and signals of opportunity systems," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 10426–10436, 2019.
- [18] "MSU-GRI surface soil moisture product," <http://https://www.gri.msstate.edu/research/ssm/>, note = Accessed: 2021-04-04.
- [19] Prashant K Srivastava, Prem C Pandey, George P Petropoulos, Nektarios N Kourgialas, Varsha Pandey, and Ujjwal Singh, "Gis and remote sensing aided information for soil moisture estimation: A comparative study of interpolation techniques," *Resources*, vol. 8, no. 2, pp. 70, 2019.
- [20] Hui Chen, Li Fan, Wei Wu, and Hong-Bin Liu, "Comparison of spatial interpolation methods for soil moisture and its application for monitoring drought," *Environmental monitoring and assessment*, vol. 189, no. 10, pp. 1–13, 2017.
- [21] Kaikai Fang, Huike Li, Zhikang Wang, Yifei Du, and Jing Wang, "Comparative analysis on spatial variability of soil moisture under different land use types in orchard," *Scientia Horticulturae*, vol. 207, pp. 65–72, 2016.
- [22] Clara Chew, "Spatial interpolation based on previously-observed behavior: a framework for interpolating spaceborne gnss-r data from cygnss," *Journal of Spatial Science*, pp. 1–14, 2021.
- [23] Volkan Senyurek, Mehmet Kurum, Ali Gurbuz, Fangni Lei, Robert Moorhead, and Dylan Boyd, "Spatial and temporal interpolation of CYGNSS soil moisture estimations," in *IGARSS 2021-2021 IEEE International Geoscience and Remote Sensing Symposium*. IEEE.
- [24] Donald Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the 1968 23rd ACM national conference*, 1968, pp. 517–524.
- [25] Patrick M Bartier and C Peter Keller, "Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (idw)," *Computers & Geosciences*, vol. 22, no. 7, pp. 795–799, 1996.
- [26] Steven M Kay, *Fundamentals of statistical signal processing*, Prentice Hall PTR, 1993.
- [27] S. K. Chan, R. Bindlish, P. O'Neill, T. Jackson, et al., "Development and assessment of the SMAP enhanced passive soil moisture product," *Remote Sens. Environ.*, vol. 204, pp. 931–941, 2018.
- [28] Jean-François Pekel, Andrew Cottam, Noel Gorelick, and Alan S Belward, "High-resolution mapping of global surface water and its long-term changes," *Nature*, vol. 540, no. 7633, pp. 418, 2016.
- [29] T. Hengl, J.M. de Jesus, G. B. Heuvelink, M. R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, and M. A. Guevara, "Soilgrids250m: Global gridded soil information based on machine learning," *PLoS one*, vol. 12, no. 2, 2017.
- [30] P. E. O'Neill, E. G. Njoku, T. J. Jackson, S Chan, and R Bindlish, "Smmap algorithm theoretical basis document: Level 2 & 3 soil moisture (passive) data products," *Jet Propulsion Laboratory. California Institute of Technology*, pp. JPL D–66480, 2015.
- [31] Gene A Poe, "Optimum interpolation of imaging microwave radiometer data," *IEEE Transactions on geoscience and remote sensing*, vol. 28, no. 5, pp. 800–810, 1990.
- [32] Philip J Davis, *Interpolation and approximation*, Courier Corporation, 1975.
- [33] Robin Sibson, "A brief description of natural neighbour interpolation," *Interpreting multivariate data*, 1981.
- [34] Margaret A Oliver and Richard Webster, "Kriging: a method of interpolation for geographical information systems," *International Journal of Geographical Information System*, vol. 4, no. 3, pp. 313–332, 1990.
- [35] Der-Tsai Lee and Bruce J Schachter, "Two algorithms for constructing a delaunay triangulation," *International Journal of Computer & Information Sciences*, vol. 9, no. 3, pp. 219–242, 1980.



Volkan Senyurek received his B.S., M.S., and Ph.D. degrees in electronics and communication engineering from Marmara University in 2003, 2007, 2013, respectively. After he got Ph.D, he worked as an assistant professor at Marmara University until 2015. Between 2015 and 2017, he worked as a Post-doctoral researcher in the Department of Mechanical and Materials Engineering at Florida International University. Between 2017 and 2019, he joined the Department of Electric and Computer Engineering at the University of Alabama as a Post-doctoral researcher. Currently, he is an assistant research professor at Geosystems Research Institute at Mississippi State University. His research interests are remote sensing, biomedical signal processing, wearable sensors, pattern recognition, fiber optic sensors, and structural health monitoring.



Ali Gurbuz received B.S. degree from Bilkent University, Ankara, Turkey, in 2003, in Electrical Engineering, and the M.S. and Ph.D. degrees from Georgia Institute of Technology, Atlanta, GA, USA, in 2005 and 2008, both in Electrical and Computer Engineering. From 2003 to 2009, he researched compressive sensing based computational imaging problems at Georgia Tech. He held faculty positions at TOBB University and University of Alabama between 2009 and 2017 where he pursued an active research program on the development of sparse signal representations, compressive sensing theory and applications, radar and sensor array signal processing, and machine learning. Currently, he is an Assistant Professor at Mississippi State University, Department of Electrical and Computer Engineering, where he is co-director of Information Processing and Sensing (IMPRESS) Lab. He is the recipient of The Best Paper Award for Signal Processing Journal in 2013 and the Turkish Academy of Sciences Best Young Scholar Award in Electrical Engineering in 2014, NSF CAREER award in 2021. He has served as an associate editor for several journals such as Digital Signal Processing, EURASIP Journal on Advances in Signal Processing and Physical Communications. He is a senior member of IEEE.



Mehmet Kurum (M'08–SM'14) received the B.S. degree in electrical and electronics engineering from Bogazici University, Istanbul, Turkey, 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 a Postdoctoral position with the Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA. In 2016, he was an Assistant Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA. His research interests include microwave and millimeter-wave remote sensing, RF sensors and systems, radiation and scattering theory, and signals of opportunity. Dr. Kurum was a recipient of the Leopold B. Felsen Award for excellence in electromagnetic in 2013 and the International Union of Radio Science (URSI) Young Scientist Award in 2014. He is an Early Career Representative for the International URSI Commission F (Wave Propagation and Remote Sensing).