Remote Sensing of Environment 247 (2020) 111944

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Pan-tropical soil moisture mapping based on a three-layer model from CYGNSS GNSS-R data

Qingyun Yan[a], Weimin Huang[b,*], Shuanggen Jin[a,c], Yan Jia[d]

[a] School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China
[b] Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1B 3X5, Canada
[c] Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China
[d] Department of Surveying and Geoinformatics, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

ARTICLE INFO

Keywords:
GNSS-Reflectometry
CYGNSS
Soil moisture
SMAP

ABSTRACT

In this paper, an effective schematic is developed for estimating soil moisture (SM) from CYclone Global Navigation Satellite System (CYGNSS) data. Here, a three-layer model of air, vegetation cover, and soil is considered. In practice, the surface reflectivity (Γ) along with its statistics derived from the CYGNSS data and the ancillary vegetation opacity (τ) data from Soil Moisture Active Passive (SMAP) are employed. The expression for empirically retrieving SM from τ, Γ and its corresponding statistics are determined through the linear regression technique. The CYGNSS data collected over the land surfaces within $\pm$ 37° (latitude) during the whole year of 2018 are investigated. The reference SM data are obtained from SMAP, and are regarded as ground-truth in this work. Validation and assessment are performed on the pan-tropical daily data collected throughout the annual circle at a resolution of 36 $\times$ 36 km^2. Experimental evaluation demonstrates good consistency between the SM derived from CYGNSS data and the ground-truth, with a correlation coefficient of 0.80 and a root-mean-square error of 0.07 cm^3/cm^3. This method succeeds in providing SM estimations on a pan-tropical scale that does not rely on ongoing knowledge of SM and merely employs the least ancillary data. Furthermore, the intense temporal and spatial coverages of CYGNSS SM results are also illustrated. The use of CYGNSS SM significantly enhances the pan-tropical coverage of SMAP SM by about 22% on average. The satisfactory outcomes achieved here prove CYGNSS as an efficient complementary tool for pan-tropical SM sensing on a daily basis.

1. Introduction

Soil moisture (SM) is a key parameter for improving our understanding of the hydrological, geophysical, and agricultural processes as well as weather and climate studies. It has been reported that updating SM on a daily basis in a numerical weather model is beneficial in increasing precipitation forecast skill (Zhan et al., 2016; Zheng et al., 2018). Availability of daily SM has potential for improving hydrological models and predicting water availability, infiltration, percolation, runoff, plant transpiration, bare soil evaporation, and flood discharge (e.g., Jin and Henderson, 2011; Seneviratne et al., 2010). Therefore, knowledge of SM with both temporally and spatially intensive coverage is important. Remote sensing techniques offer an exciting solution to acquiring global SM data in an efficient and timely manner.

Global-scale SM data can be obtained via microwave remote sensing techniques since the signals at microwave frequencies are sensitive to the dielectric constant of soil that is a function of SM (Dobson et al., 1985). Existing satellites dedicated for SM estimation include the Soil Moisture Active Passive (SMAP) (Entekhabi et al., 2010) and Soil Moisture and Ocean Salinity (SMOS) (Kerr et al., 2001), both carrying L-band radiometers on-board and measuring brightness temperature. Both missions provide SM data with a spatial resolution about 40 km and global coverage every 2–3 days. In particular, the L-band signals are the optimal for SM sensing because they are more sensitive to SM in the top 5 cm of the soil column and less affected by the attenuation due to surface roughness and vegetation (De Roo and Ulaby, 1994). Moreover, L-band signals can effectively penetrate the atmosphere and are immune from weather conditions, and thus, they can guarantee the all-day and all-weather surveillance over the regions of interest (Jin and Komjathy, 2010; Jin et al., 2011). In addition to these passive sensors, active platforms e.g. Sentinel-1 (Paloscia et al., 2013) and ERRASAR-X (Aubert et al., 2011) can also provide SM estimation through radar backscattering measurements.

Global Navigation Satellite System (GNSS)-Reflectometry (GNSS-R)

* Corresponding author.
E-mail address: weimin@mun.ca (W. Huang).

https://doi.org/10.1016/j.rse.2020.111944
Received 23 December 2019; Received in revised form 6 June 2020; Accepted 7 June 2020

that utilizes L-band signals has emerged as another promising tool for estimating SM. GNSS-R operates in a bistatic configuration, in which transmitted signals are first scattered by the Earth's surface in the forward direction and then captured by a receiver that is not collocated with the transmitter. For GNSS-R, numerous transmitters have already been in orbit as part of the GNSS constellation and thus only the deployment of receivers is needed. Furthermore, a GNSS-R receiver can be low-cost, low-mass, and low-power. This advantage allows more GNSS-R campaigns to be carried out based on various platforms, which ultimately improves the spatial and temporal coverages of GNSS-R measurements. After a few early attempts of GNSS-R-based SM sensing by Masters et al. (2000); Zavorotny and Voronovich (2000a), several field tests have been performed using ground-based and airborne receivers (Alonso-Arroyo et al., 2016; Jia et al., 2019; Rodriguez-Alvarez et al., 2009). A summary of measurements, approaches, and applications using the conventional and interferometric GNSS-R is provided by Edokossi et al. (2020). Some early spaceborne data-based research (Camps et al., 2016; Chew et al., 2016)\focused on developing the usage of TechDemoSat-1 (TDS-1) data for SM application. Unfortunately, the data acquisition of TDS-1 poses severe limitations in terms of spatial and temporal coverage since the TDS-1 GNSS-R instrument is active for only 2 days within each working period (8 days). Therefore, generating daily SM estimation cannot be realized using the TDS-1 data. However, this task can be achieved by the CYclone GNSS (CYGNSS) mission that has been operating daily since March 2017 and providing a data volume being several orders of magnitude more than that acquired by TDS-1 (Chew and Small, 2018). Recently, research of SM sensing from CYGNSS data has been a topic of interest (Al-Khaldi et al., 2019; Chew and Small, 2018; Clarizia et al., 2019; Eroglu et al., 2019; Kim and Lakshmi, 2018). Chew and Small (2018) found that the variations in the CYGNSS signal-to-noise ratio (SNR) were correlated with the SMAP SM results, and interpreted the correlation by a linear regression approach. They incorporated the mean SMAP SM values and the derived SM variations (based on CYGNSS SNR data) to calculate the SM estimations. Kim and Lakshmi (2018) developed a relative SNR (rSNR) for SM derivation. Daily SM estimations can be derived by combining rSNR with SMAP SM values. Clarizia et al. (2019) proposed a reflectivity-vegetation-roughness (R-*V*-R) algorithm. The daily SM estimations can be derived based on the CYGNSS reflectivity along with SMAP vegetation opacity (τ) and roughness coefficient through a trilinear regression function. Al-Khaldi et al. (2019) performed the SM retrieval through time-series analysis. This method requires both the maximum and minimum of SMAP SM values to limit their SM retrieval results. Eroglu et al. (2019) presented an artificial neural network (ANN)-based method for retrieving daily SM. The inputs for their designed ANN included CYGNSS data and other ancillary data that consist of normalized difference vegetation index (NDVI), vegetation water content (VWC), terrain elevation, terrain slope, and h-parameter (surface roughness). Through comparison (see Table 1), the results presented by Al-Khaldi et al. (2019); Chew and Small (2018); Eroglu et al. (2019) appear to be advantageous in terms of accuracy by evaluating root-mean-squared error (RMSE) or correlation coefficient (r). However, it should be noted that these methods lie on either the direct knowledge of SM or the heavy-loaded ancillary data. Some other relevant information about above-mentioned SM applications is summarized in Table 1.

This paper proposes a CYGNSS-based SM retrieval scheme that does not require ongoing knowledge of SM and employs the least ancillary data. This work aims at demonstrating CYGNSS's capability to complement the SMAP mission for daily monitoring SM on a pan-tropical scale. In this work, the received CYGNSS signals are considered to be affected by SM, surface roughness, and vegetation. The contributions of SM and surface roughness are resolved by the proposed CYGNSS observables and the attenuation by vegetation cover is compensated by SMAP vegetation opacity τ. This proposed method is applicable to all the land surfaces regardless of different surface elevations, seasonal changes or local vegetation conditions. Compared with the previous research, this work presents a simple, effective, and general method for retrieving soil moisture. The remainder of this paper is organized as follows: Section 2 introduces the employed CYGNSS and reference SMAP data. Section 3 describes the proposed SM estimation scheme. Section 4 presents results and discussions. Section 5 concludes with a summary and possible future work of this research.

2. Observation data

In this section, the acquisition of CYGNSS GNSS-R data is first described. Next, the processing of the reference SMAP data is introduced.

2.1. CYGNSS GNSS-R data

CYGNSS was launched in December 2016 (Ruf et al., 2016), and has been offering data of free access since March 2017 (available online: https://podaac.jpl.nasa.gov). CYGNSS has eight micro-satellites, each of them can provide four GNSS-R measurements (from four separate ground tracks) at the same time. Since they provide data every second, a total number of 32 measurements can be acquired simultaneously per second. In addition, the achieved revisit time is within several hours (Ruf et al., 2016). Therefore, CYGNSS is able to monitor SM with extensive spatial coverage and high temporal resolution. The data examined in this work span the year of 2018 and over the full region covered by CYGNSS (within ± 37° latitudes).

The CYGNSS metadata include the bistatic radar cross section (BRCS, or σ) and SNR at each specular point (SP) as well as their associated information about the measuring geometry and navigation message, such as, the incidence angle, coordinates of SP, distances from SP to the transmitter and receiver etc. In this work, data collected over land and with an SNR over 0 dB at SP are retained. It is worth mentioning that the CYGNSS BRCS is stored in a signal box with 17 delay × 11 Doppler bins. The position of CYGNSS BRCS in a signal box can vary from each other due to the variations of SP locations. In order to ensure the main BRCS component to be retained within its signal box, only BRCS data with a peak position between the 4th and 15th bins in the delay axis are persevered. In this work, the surface reflectivity is derived from CYGNSS data based on the assumption of coherent reflections (see more details in Section 3.1). Thus, the highest achievable resolution is about 7×0.5 km^2 (see e.g. Katzberg et al., 2006), by considering the resolution of a single CYGNSS coherent measurement.

Table 1
Applications of SM sensing using CYGNSS data.

Source	Time span	Spatial coverage	Reference SM	Require ongoing SM	No. of ancillary data	RMSE (cm^3/cm^3)	r
Chew and Small (2018)	1 year	Pan-tropical*	SMAP	No	2	0.045	/
Kim and Lakshmi (2018)	1 year	Regional	SMAP	Yes	1	/	0.68/0.77
Al-Khaldi et al. (2019)	6 months	Pan-tropical	SMAP	Yes	1	0.04	0.82
Eroglu et al. (2019)	2 years	Regional	In situ	No	5	0.054	0.90
Clarizia et al. (2019)	5 months	Pan-tropical*	SMAP	No	2	0.07	/
Proposed method	1 year	Pan-tropical	SMAP	No	1	0.07	0.80

* Regions with surface elevation exceeding 600 m were not examined.

2.2. SMAP data

SMAP provides SM estimation on a daily basis and its global coverage for land areas within ± 45° latitudes can be achieved at a three-day average interval (Entekhabi et al., 2010). In this work, the SMAP L3 radiometer global daily equal-area scalable earth grid (EASE-Grid) version 6 SM data (O'Neill et al., 2019) (available on https://nsidc.org/data/SPL3SMP/versions/6) are employed. This dataset has a spatial resolution of 36 × 36 km^2. To facilitate the later comparison and validation, daily CYGNSS data are also gridded over the same EASE-Grid as adopted by the SMAP data (see also Clarizia et al. (2019); Chew and Small (2018)).

SMAP daily data contain SM estimation, quality flag, vegetation opacity τ, and the associated coordinates for both the descending (A.M.) and ascending (P.M.) overpasses. In this work, the ascending and descending passes are averaged together to represent the 1-day results. Similarly, in order to produce the globally covered data, an averaging over three consecutive days is implemented, and hereafter, the corresponding data are denoted as 3-day results.

The SMAP vegetation opacity τ data are treated as ancillary data for CYGNSS. Therefore, it is only possible to provide CYGNSS-based SM estimate on the pixels that are also covered with SMAP in this current work. It is worth mentioning that τ can also be linked to *VWC*, and the latter can be calculated from vegetation indices, e.g., NDVI. Furthermore, NDVI can be obtained from, for example, the Moderate-resolution Imaging Spectroradiometer (MODIS) data. Thus, τ can be determined from MODIS NDVI data when the SMAP data are not available. The obtained SMAP SM estimation is regarded as the ground-truth for validating CYGNSS results presented later. In addition, the quality flag of "retrieval successful/unsuccessful" is employed for filtering SMAP data as done by Clarizia et al. (2019).

3. Theory and methods

In this section, a detailed procedure of estimating SM from CYGNSS data is described. Here, a model similar to the R-V-R algorithm introduced by Clarizia et al. (2019) is employed.

For a flat and smooth region covered by vegetation (see Fig. 1 for illustration), the surface reflectivity Γ can be modeled as (Choudhury et al., 1979)

$$\Gamma(\theta) = \Re(\theta)^2 \gamma^2 \exp(-4k^2 s^2 \cos^2(\theta)), \quad (1)$$

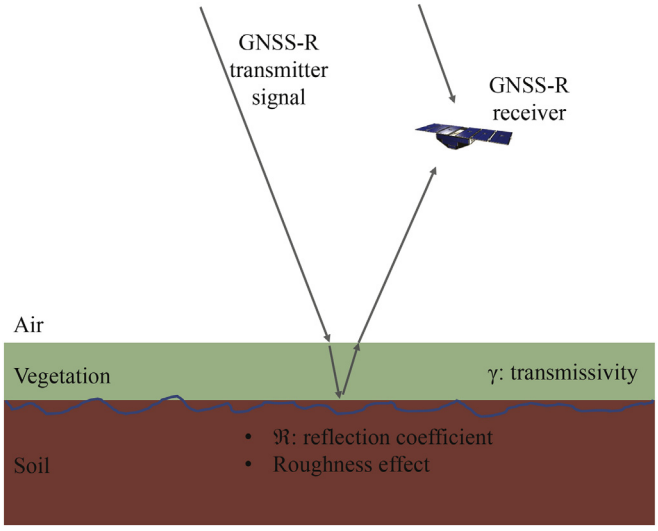

Fig. 1. Schematic of GNSS-R signal reflected from a three-layer model of air, vegetation, and soil. The blue curve represents surface roughness. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where θ is the incidence angle, $\Re$ is the Fresnel reflection coefficient, the transmissivity γ accounts for the attenuation of signal propagation by vegetation, and the exponential term represents the surface roughness effects with k being the signal wavenumber and s being the surface root-mean-squared height.

Since $\Re$ is dominated by SM (Dobson et al., 1985), SM is regarded as a proxy for $\Re$. In addition, γ is a function of vegetation opacity τ in the form of $\gamma = \exp(-\tau \sec\theta)$. For these reasons, the SM is treated as a function of surface roughness, Γ, and τ. In this work, τ is adopted from SMAP dataset as auxiliary data while the surface roughness effects and reflectivity Γ are resolved from the CYGNSS data.

3.1. Derivation of reflectivity from CYGNSS

By following the assumption of coherent reflections over smooth land for Eq. (1) (that is also made by Chew and Small (2018); Clarizia et al. (2019); Eroglu et al. (2019)), the surface reflectivity Γ can be readily derived from CYGNSS BRCS σ, through (Eroglu et al., 2019; Rodriguez-Alvarez et al., 2019):

$$\Gamma = \frac{\sigma (R_t + R_r)^2}{4\pi (R_t R_r)^2}, \quad (2)$$

where R_t and R_r are the distances from the transmitter and receiver to SP, respectively. As mentioned in Section 2.1, these parameters are accessible from the CYGNSS data. Instead of taking only the peak value of Γ as done by previous studies (Chew and Small, 2018; Clarizia et al., 2019; Eroglu et al., 2019; Kim and Lakshmi, 2018; Rodriguez-Alvarez et al., 2019), the whole frame of Γ results (17 × 11) is investigated to further exploit the CYGNSS data. Here, the statistical moments of Γ are employed, including the maximum, mean, variance, skewness, and kurtosis, which are denoted by Γ_{max}, Γ_{mean}, Γ_{var}, Γ_{skew}, and Γ_{kur}, respectively. Γ_{max} is first extracted and the rest moments that describe the shape of data distribution are calculated from the Γ divided by Γ_{max}. This process aims at: 1) obtaining the CYGNSS reflectivity as Γ_{max} and 2) deriving other CYGNSS observables, specifically, Γ_{mean}, Γ_{var}, Γ_{skew}, and Γ_{kur} to account for the effects of surface roughness. The effectiveness of the latter step is supported by the fact that surface roughness can be interpreted from the shape of GNSS-R data (e.g., Clarizia et al., 2014; Rivas et al., 2010; Rodriguez-Alvarez et al., 2013; Yan et al., 2018; Yan et al., 2017; Yan and Huang, 2016; Zavorotny et al., 2014; Zavorotny and Voronovich, 2000b). It should be noted that Γ_{max} is regarded as the contribution of the first Fresnel zone roughly with a spatial resolution of 7 × 0.5 km^2, while the remaining statistical moments are derived from the full-size BRSC data, whose spatial resolution (also known as the glistening zone) is about several dozens of km. In the present work, the roughness within the glistening zone is assumed to be uniform so that Γ_{mean}, Γ_{var}, Γ_{skew}, and Γ_{kur} can be representatives for the first Fresnel zone.

As such, the proposed scheme for estimating SM adopts the CYGNSS Γ_{max}, Γ_{mean}, Γ_{var}, Γ_{skew}, and Γ_{kur} as well as the SMAP τ. The linear regression approach (that is also investigated by Chew and Small (2018); Clarizia et al. (2019)) is employed to derive CYGNSS SM (or, SM_{CYGNSS}), through

$$SM_{CYGNSS} = a\cdot\Gamma_{max} + b\cdot\Gamma_{mean} + c\cdot\Gamma_{var} + d\cdot\Gamma_{skew} + e\cdot\Gamma_{kur} + f\cdot\tau + g, \quad (3)$$

where coefficients from a to g are to be determined via a training process that is described in Section 4.2.

4. Results and analysis

In this section, the SM_{CYGNSS} results are presented and evaluated by comparing with SMAP SM data i.e. SM_{SMAP}. In particular, the feasibility of daily SM_{CYGNSS} for filling the data gaps of SM_{SMAP} is demonstrated.

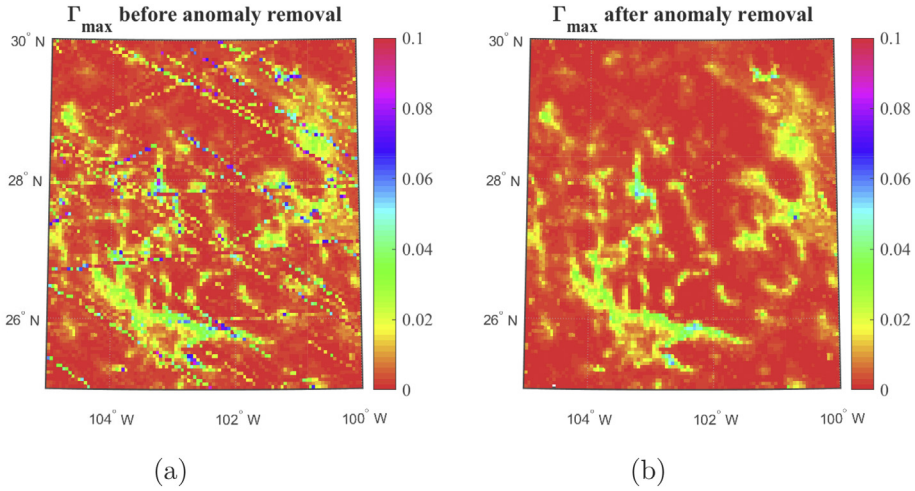

Fig. 2. Removing CYGNSS anomalies: (a) before and (b) after. Examples are shown for regions within [25, 30]° N and [100, 105]° W during the first two months in 2018.

Table 2
Numbers (million) of employed CYGNSS data and resulting samples in the EASE-Grid.

	CYGNSS data	Grided samples
1-day-based	115.93	10.93
3-day-based	169.06	16.33

4.1. Data anomaly removal

As noted in Section 2, the daily CYGNSS observables are gridded over the EASE-Grid with a spatial resolution of 36 km. During this process, some extremely high values of Γ_{max} are found, which appear consecutively within a same ground track (see Fig. 2a). This phenomenon is also mentioned by Gerlein-Safdi and Ruf (2019), which can be resulted from variations in GNSS power (Wang et al., 2019). Here, these anomalies are removed by filtering out Γ_{max} greater than 0.1. Although this threshold is empirically determined, it can effectively remove those anomalies and preserve useful data at the same time (see Fig. 2b).

4.2. Training and testing

After removing the anomalies among CYGNSS observables through evaluating Γ_{max}, these observables along with corresponding global SMAP τ and SM_{SMAP} are derived on the daily basis throughout the year of 2018. The employed CYGNSS data are mapped into the EASE-Grid, generating millions of samples in total; and the detailed data/sample numbers are summarized in Table 2. Each sample represents one pixel in the EASE-Grid, corresponding to an area of 36 × 36 km^2. For illustration, the yearly averaged values of the CYGNSS observables and SMAP τ are displayed in Fig. 3.

Table 3
Coefficients for Eq. (3), determined through the pan-tropical training dataset.

Coefficients	Value	*P*-value
a	2.3864	0
b	0.3532	0
c	−0.0409	1.8×10^{-9}
d	−0.0048	6.6×10^{-30}
e	0.0026	0
f	0.2560	0
g	0.0229	0

The overall daily samples are randomly divided into two separate groups for training and test that have 5% and 95% of the total samples, respectively. Through the training process, the values of coefficients in Eq. (3) are determined and tabulated into Table 3. Through inspection, skewness and kurtosis are found to be highly correlated (with a correlation coefficient of 0.96), which is also evident from Fig. 3. As a

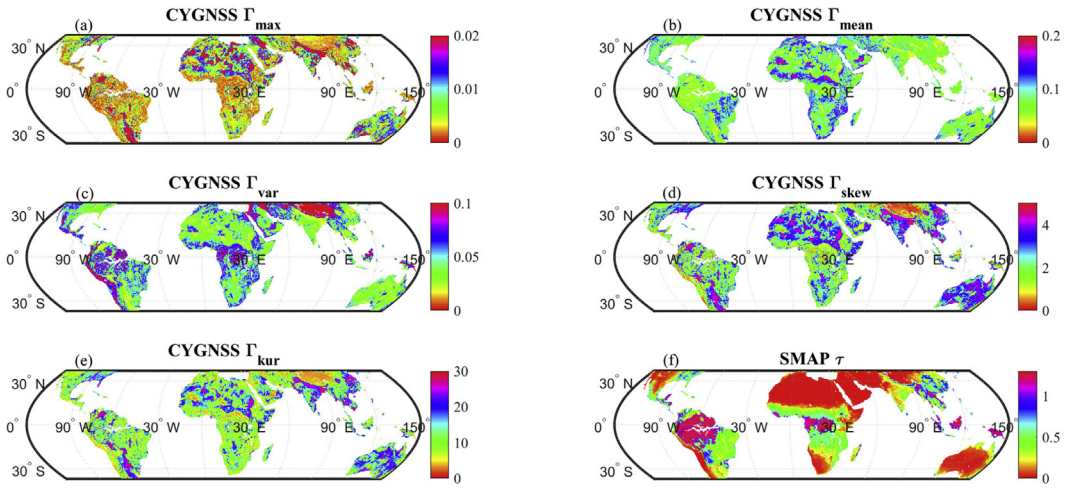

Fig. 3. Annual mean values of: (a) Γ_{max}, (b) Γ_{mean}, (c) Γ_{var}, (d) Γ_{skew}, (e) Γ_{kur}, and (f) SMAP τ. These variables are dimensionless.

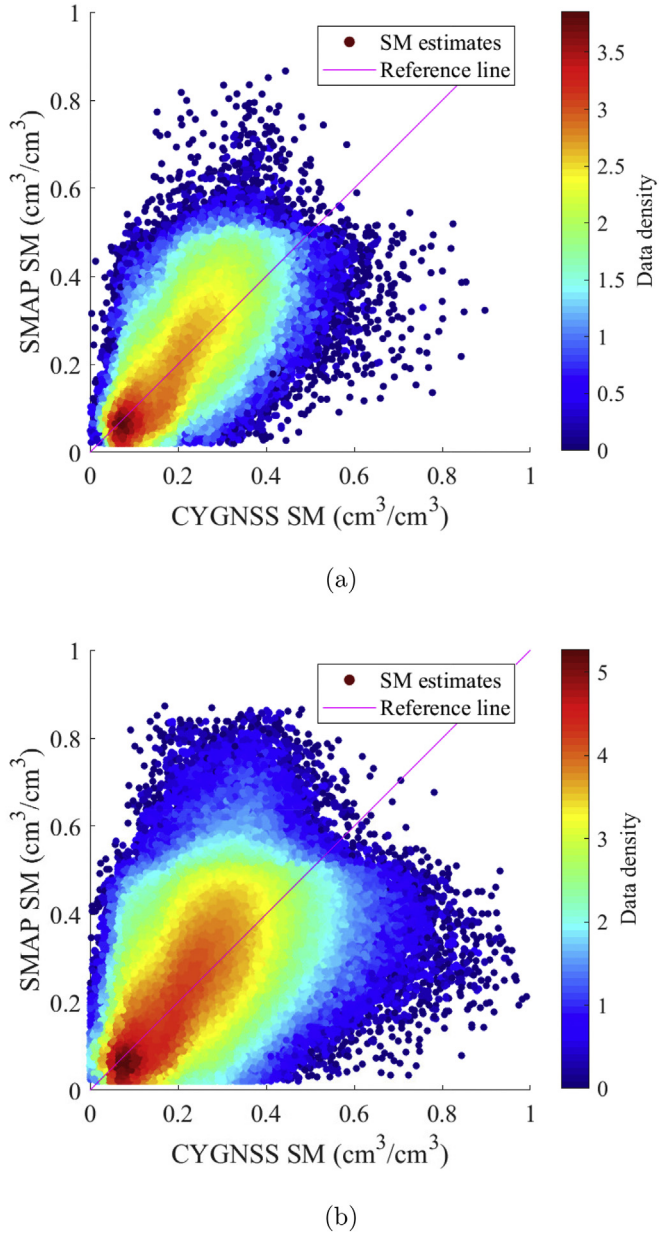

Fig. 4. Density plots (in log-scale) comparing SM_{CYGNSS} and SM_{SMAP} with the 1:1 reference line (magenta): (a) training and (b) test sets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

result, only trivial degradation is caused when developing the regression by eliminating one of these variables. Even so, both variables are adopted in the current regression model to ensure the best precision. Furthermore, different combinations of inputs with/without highly correlated terms are tested, and the achieved correlation coefficients are similar, proving that the proposed model is insensitive to the collinearity problem.

The density plots (in log-scale) showing the comparison between SM_{CYGNSS} and SM_{SMAP} for both training and test datasets are presented as Fig. 4. From Fig. 4, good consistency between SM_{CYGNSS} and SM_{SMAP} can be seen, especially for the most dense data. Specifically, a correlation coefficient of $r = 0.79$ and an RMSE of 0.07 cm^3/cm^3 are obtained for both training and test sets. Similar performance achieved for both datasets indicates the generalizability of the proposed method. Moreover, these results are generated pan-tropically and throughout the year, proving the general applicability and robustness of the

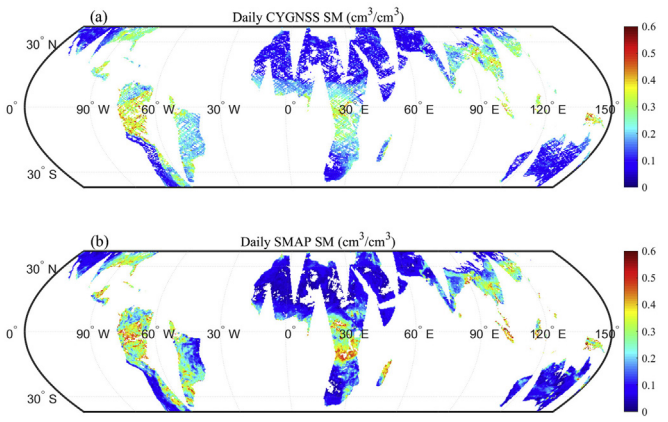

Fig. 5. Pan-tropical daily SM of: (a) SM_{CYGNSS} and (b) SM_{SMAP}. This is an example of January 1, 2018. Increased coverage of CYGNSS over SMAP is demonstrated in Fig. 8.

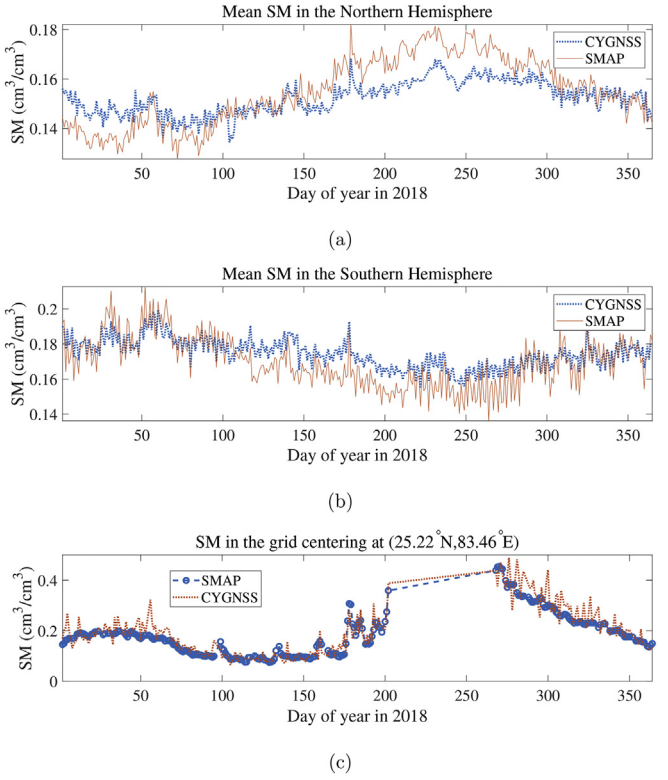

Fig. 6. SM time series between CYGNSS and SMAP throughout the year of 2018 for: (a) the average in the Northern and (b) Southern Hemispheres, and (c) the grid cell centering at (25.22°N,83.46°E).

proposed method. The obtained accuracy (in terms of RMSE) and fundamentals of the proposed method are comparable with those presented by Clarizia et al. (2019), the main differences of these two methods lie in the following aspects: 1) The proposed method employs the statistics of real-time CYGNSS Γ to compensate the surface roughness effect, which reduces the amount of auxiliary data. On the other hand, the method developed by Clarizia et al. (2019) relies on the SMAP roughness coefficient that is essentially based on a lookup table and lacks resistance in temporal variability. 2) An extra step of error correction is required by Clarizia et al. (2019), while the proposed method has no such limitation. 3) The developed algorithm is evaluated with more data that provide longer time span (1 year v.s. 5 months by Clarizia et al. (2019)) and larger spatial coverage. Relevant comparison between these two algorithms are is listed in Table 1.

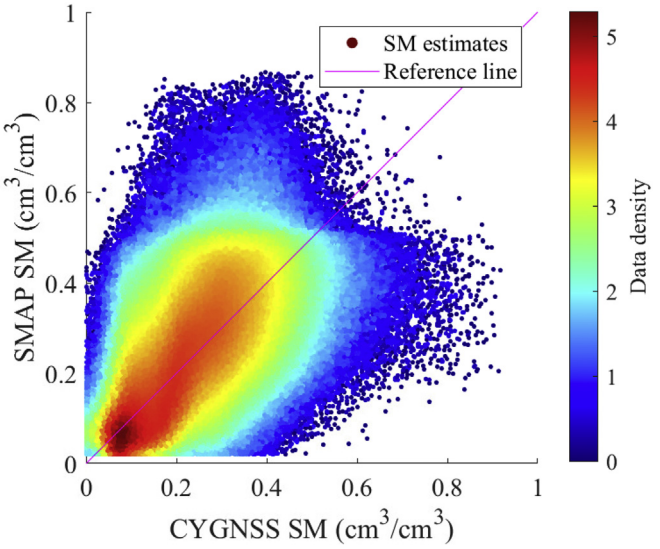

Fig. 7. Density plots for comparing daily SM_{CYGNSS} with 3-day SM_{SMAP}.

For illustration purposes, an example of pan-tropical daily SM_{CYGNSS} and SM_{SMAP} on January 1, 2018 is presented in Fig. 5. In addition, the SM time series of SM_{CYGNSS} and SM_{SMAP} during the year are produced and found to be consistent (see Fig. 6). It is worth mentioning that due to the different seasonal distribution between two hemispheres, the mean SM values are separately considered for both the northern and southern hemispheres and the corresponding r obtained (between the mean values of SM_{SMAP} and SM_{CYGNSS}) for the north and south are 0.87 and 0.86, respectively. In addition, the time series for the grid cell centering at (25.22°N,83.46°E) achieves an $r = 0.92$ (see Fig. 6(c) and the corresponding location is marked by a black circle in Fig. 9). Detailed investigations into the coverage and effectiveness of SM_{CYGNSS} and ultimately, its feasibility of infilling the data gap for SM_{SMAP} are presented in the following subsections.

4.3. SM_{CYGNSS} v.s. 3-day averaged SM_{SMAP}

The results shown above focus on exhibiting the general applicability and efficiency of the proposed method for generating SM_{CYGNSS} estimation. Here, further analyses are provided for studying the pan-tropical coverage and performance of SM_{CYGNSS} during the annual circle, to prove the capability of SM_{CYGNSS} for infilling SM_{SMAP} data gaps.

As previously noted in Section 2.2, SMAP can provide global coverage on a three-day interval. For the sake of intensive spatial coverage, each of the SMAP SM and τ is re-processed by averaging data over every three consecutive days, and hereafter, denoted as 3-day data. In addition, they are ideally treated as the daily representative by neglecting data variations within the three-day interval. The daily CYGNSS observables are also re-derived according to the new spatial coverage of their associated 3-day SMAP data and the 3-day SMAP τ is also employed as input. The follow-up investigations are based on the daily CYGNSS and 3-day SMAP data.

After re-producing data with improved spatial coverage, there is a significant increase on the sample number (see Table 2 for comparison). Same as the previous 1-day data, each sample covers an area of 36 × 36 km^2. By setting the covered area of 3-day SM_{SMAP} as the maximum pan-tropical coverage, the percentages of actual daily coverages by SM_{CYGNSS} and SM_{SMAP} are computed, from the 2nd to the 364th days in 2018. On average, SM_{CYGNSS} and SM_{SMAP} provide 81% and 66% of the pan-tropical coverage, respectively.

The derived regression model parameterized by coefficients listed in Table 3 is also utilized here for estimating SM_{CYGNSS}. Overall results in the form of density plot are presented in Fig. 7. Through calculation, a correlation coefficient of $r = 0.80$ and an RMSE of 0.07 cm^3/cm^3 are obtained. The attained results experience no drop in accuracy (instead, a slight improvement on r, from 0.79 to 0.80) and prove the rationality of using 3-day averaged data as the daily representative and again, the generalizability of the proposed method. Additionally, SM time series

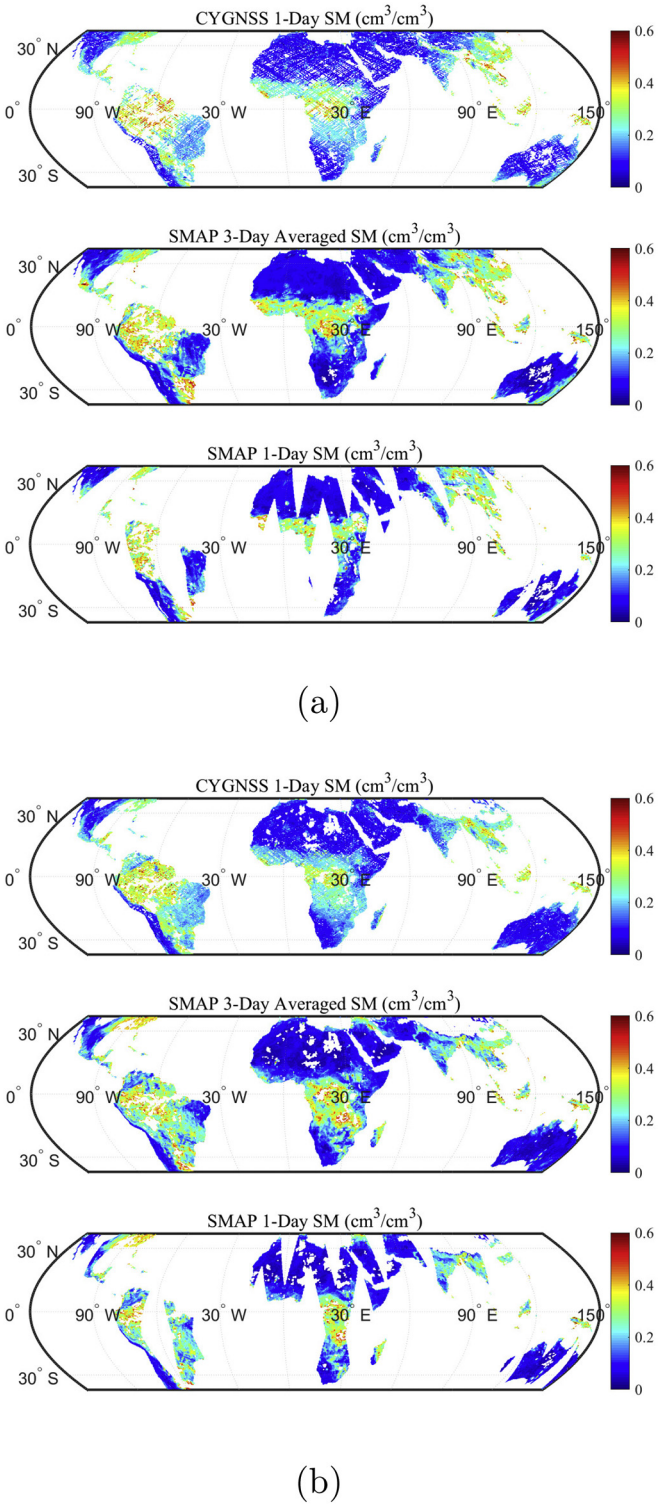

Fig. 8. Comparison of SM values. Form the top to bottom panels are the 1-day SM_{CYGNSS}, and 3-day, and 1-day SM_{SMAP}. Regions without data are in white. The 1-day results are for (a) the 182th day in 2018 (July 1, 2018) and (b) the 364th day in 2018 (December 30, 2018).

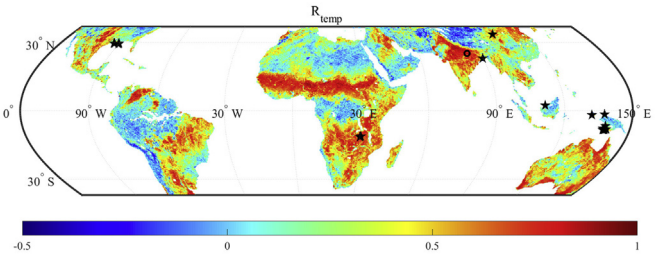

Fig. 9. Temporal correlation between daily SM_{CYGNSS} and SM_{SMAP}.

are found to be similar to those obtained with 1-day data presented in Fig. 6.

Two examples of pan-tropical daily SM_{CYGNSS} are illustrated in Figs. 8, along with the 3-day and 1-day SM_{SMAP} data as reference. The higher spatial coverage of SM_{CYGNSS} than 1-day SM_{SMAP} is clear. However, there exist grid cells not covered with CYGNSS data. This may cause lack of details in SM_{CYGNSS}.

4.4. Influencing factors evaluation

Although an overall good agreement between SM_{CYGNSS} and SM_{SMAP} have been achieved, some discrepancy can be noticed from Figs. 4 and 7. Several influencing factors for the accuracy of SM_{CYGNSS} are studied and evaluated here.

4.4.1. Temporal and spatial correlation

In order to further validate the capability of SM_{CYGNSS} for capturing the temporal and spatial variability of SM_{SMAP}, the temporal (r_{temp}) and spatial (r_{spa}) correlations between the SM_{CYGNSS} and SM_{SMAP} are evaluated. r_{temp} here is calculated as the correlation coefficient of the SM_{CYGNSS} and SM_{SMAP} time series for each grid cell, and the results are presented in Fig. 9. It is clear that the consistency varies over different regions. It should be noted that the SM_{SMAP} data in the Amazon and Central Africa that are covered by evergreen broadleaf forests (see Fig. 10 and later discussion) are always flagged by SMAP as not recommended for retrieval. However, such reference data are retained in this work and their low quality for retrieval may be the cause of high discrepancy. In addition, the variance of the SM_{SMAP} time series for each grid cell is also calculated and presented as Fig. 11. Similarities between the variance and r_{temp} maps are obvious, indicating that good agreement usually occurs for the region with high variance in local SM time series. This outcome agrees with the findings by Rodríguez-Fernández et al. (2015). Moreover, the regions where CYGNSS is showing SM less than 0.25 cm^3/cm^3, whereas SM_{SMAP} is greater than 0.75 cm^3/cm^3 are marked by black pentagrams. Moreover, r_{spa}, i.e., the time series of pan-tropical daily correlations, are calculated as 1-D correlation coefficients between the values of SM_{CYGNSS} and SM_{SMAP} for all grid cells and presented in Fig. 12. The results are relatively constant, ranging from 0.75 to 0.84 and with a mean value of 0.80.

4.4.2. Land type and τ

It is found that except regions covered with water bodies (water and wetlands), areas with high values of τ usually result in high discrepancies in SM retrieval. Errors in densely-forested regions (e.g., broadleaf forests) are generally greater than in regions without vegetation (e.g., shrublands and barren or sparsely vegetated regions), which are quantified and listed in Table 4. The land classification is based on the most dominant type with more than 50% coverage within each grid that is provided in SMAP data and an illustration is presented in Fig. 10. The types are in accordance with the International Geosphere Biosphere Program (IGBP) classes. The RMSEs of SM retrieval and mean values of τ in Table 4 are calculated with respect to the dominant land types, correspondingly. Furthermore, the dependence of r_{temp} on land types is clear when visually comparing the similarity between Figs. 9 and 10.

4.4.3. CYGNSS sample number

The sample number of CYGNSS observables within each pixel of EASE-Grid has a direct impact on the accuracy of SM_{CYGNSS}. It is found that the higher the sample number is the less discrepancy between SM_{SMAP} and SM_{CYGNSS} is (see Fig. 13). For one thing, with more CYGNSS observables being averaged together, some adverse effects, for example, noise, can be significantly mitigated. For another, with more samples in one grid, it is more possible for CYGNSS data to cover a larger spatial area within the grid in order to match better with the reference SMAP data.

4.4.4. Clay fraction

The accuracy of SM_{CYGNSS} is sensitive to different soils that are characterized by varying clay fraction values. Through analyses, it is found that with increasing clay fraction (that is provided together with SMAP) the relative difference between SM_{SMAP} and SM_{CYGNSS} decreases, which is evident from Fig. 14. This phenomenon may be attributed to the reason that the increase in clay content enhances the capacity of soil in holding moisture for certain soil types (Ismail, 1991).

Some other issues responsible for uncertainty in SM_{CYGNSS}, such as errors in localization and calibration of CYGNSS data, modeling, and the use of ancillary data can be found in the work by Clarizia et al. (2019).

5. Conclusions

In this paper, a scheme for pan-tropical daily soil moisture

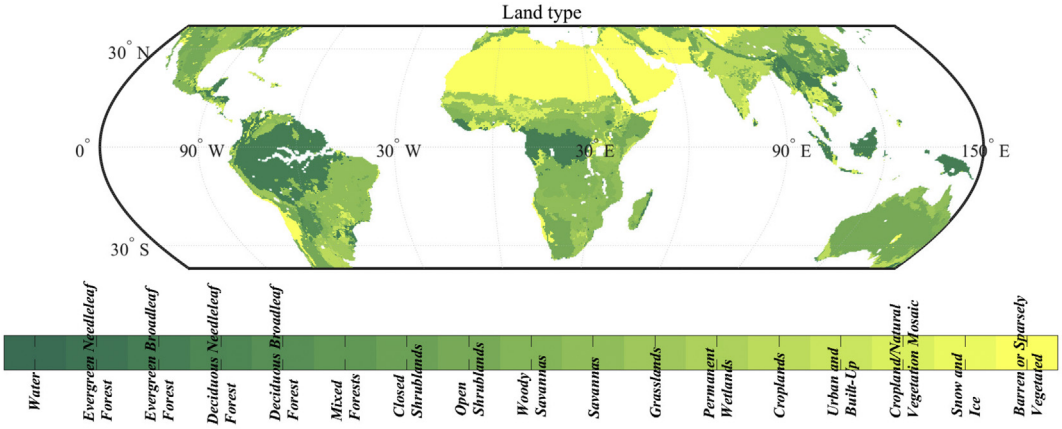

Fig. 10. The dominant land type of each grid cell.

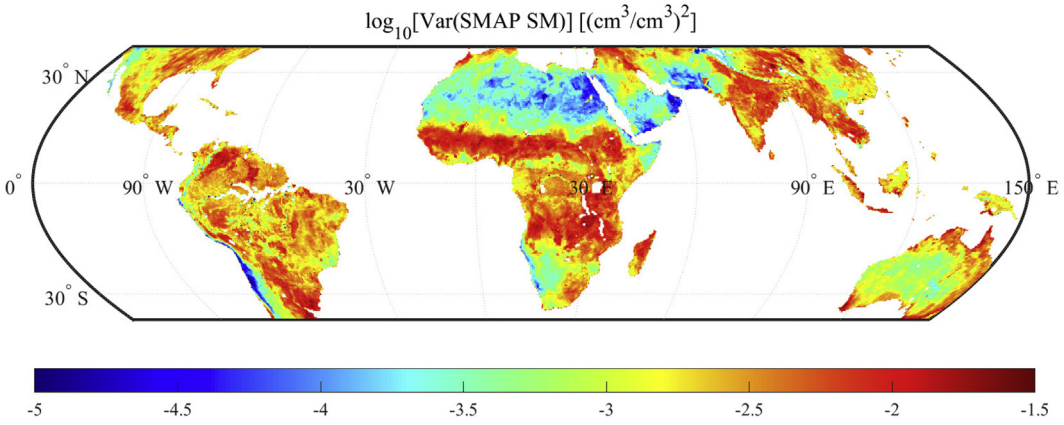

Fig. 11. Variance of SM_{SMAP} time series.

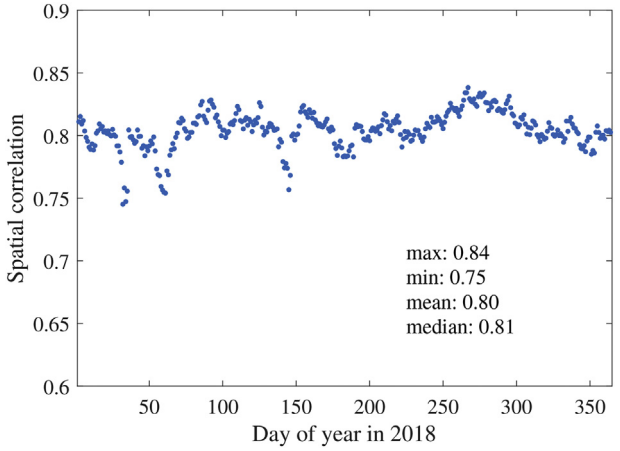

Fig. 12. Spatial correlation between daily SM_{CYGNSS} and 3-day SM_{SMAP}.

Table 4
Error analysis: Impact of land type and τ.

Land type	Mean τ	RMSE (cm^3/cm^3)
Water	0.0664	0.2973
Permanent Wetlands	0.0756	0.1537
Evergreen Broadleaf Forest	1.1222	0.0909
Deciduous Broadleaf Forest	0.7807	0.1223
Mixed Forests	0.7506	0.0790
Woody Savannas	0.5247	0.0795
Evergreen Needleleaf Forest	0.4654	0.0769
Cropland/Natural Vegetation Mosaic	0.3799	0.0791
Savannas	0.3520	0.0727
Urban and Built-Up	0.3311	0.0999
Croplands	0.2287	0.0753
Closed Shrublands	0.1877	0.0489
Grasslands	0.1075	0.0776
Open Shrublands	0.0582	0.0460
Barren or Sparsely Vegetated	0.0009	0.0454

estimation from CYGNSS data is presented. In particular, this method does not require ongoing knowledge of SM and employs the least supplementary data. Here, the surface reflectivity Γ is constructed as a function of the vegetation opacity, SM, and the effect of surface roughness. SM estimation is performed through a linear regression technique. A simple and effective approach is developed to filter out CYGNSS data anomalies. By comparing the CYGNSS SM with the reference SMAP SM data, good consistency is obtained with a correlation coefficient of $r = 0.80$ and an RMSE of 0.07 cm^3/cm^3. The feasibility of SM_{CYGNSS} for infilling data gaps in SM_{SMAP} is demonstrated. On average, daily SM_{CYGNSS} is able to cover 81% of the land surfaces. However, only

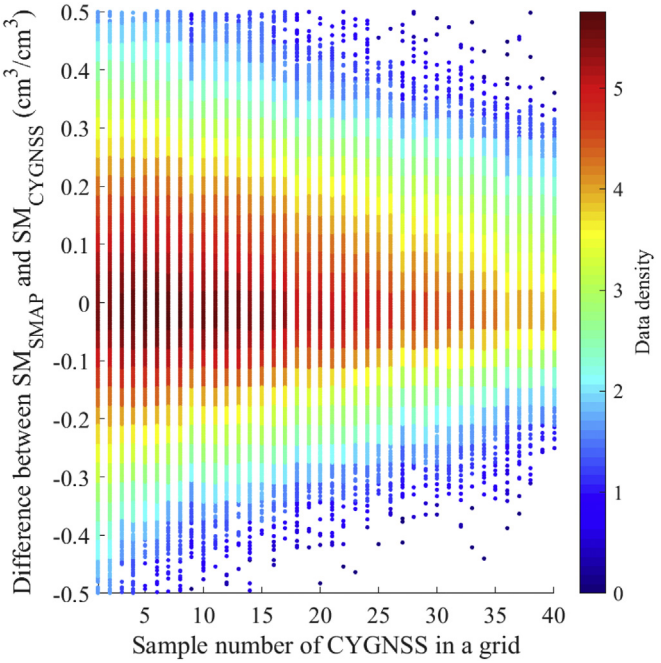

Fig. 13. Impact of CYGNSS sample number on the difference between SM_{SMAP} and SM_{CYGNSS}.

66% land is covered by SMAP.

Two sources impacting the performance of the proposed method are discussed. It is worth investigating more influencing factors and quantifying their effects on the estimation accuracy in the future. By doing so, the results can be calibrated by considering those error sources. In addition, a robust data assimilation scheme can be developed to merge SM_{SMAP} and SM_{CYGNSS}. This work here focuses on evaluating CYGNSS as a complementary tool for SMAP to get denser data. Nonetheless, validation with in situ SM is meaningful and will be carried out once such data are available in the future. More importantly, an enhanced SM estimation method without relying on ancillary data should be developed in the future so that CYGNSS can produce SM data independently.

Declaration of Competing Interest

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

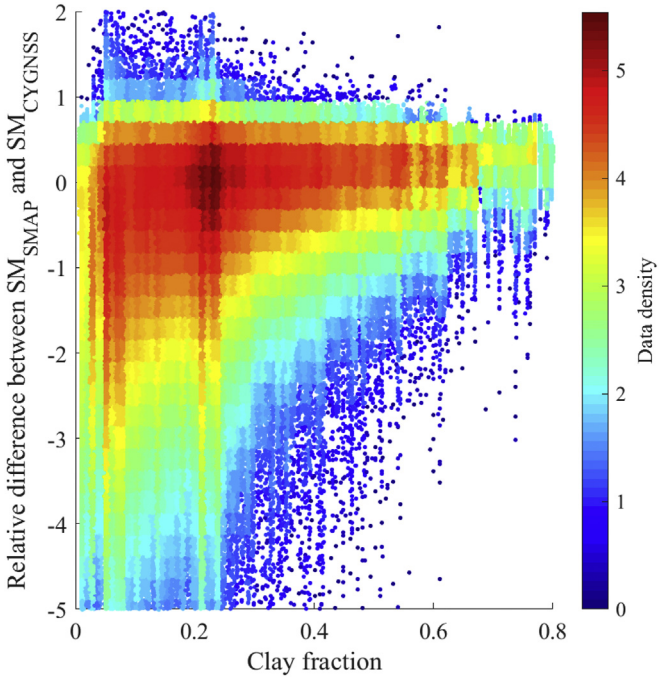

Fig. 14. Impact of clay fraction on the accuracy of SM_{CYGNSS}.

Acknowledgments

This work was supported by 1) Strategic Priority Research Program Project of the Chinese Academy of Sciences (Grant No. XDA23040100), Shanghai Leading Talent Project, Jiangsu Province Distinguished Professor Project (Grant No. R2018T20), and the Startup Foundation for Introducing Talent of NUIST (Grant No. 2018R037) to S. Jin, 2) the Natural Sciences and Engineering Research Council of Canada Discovery Grants (NSERC RGPIN-2017-04508 and RGPAS-2017-507962) and the Canadian Space Agency CubeSat Grant (17CCPNFL11) to W. Huang, and 3) the Natural Science Foundation of Jiangsu Province under Grant BK20180765 to Y. Jia.

.

References

Al-Khaldi, M.M., Johnson, J.T., O'Brien, A.J., Balenzano, A., Mattia, F., 2019. Time-series retrieval of soil moisture using CYGNSS. IEEE Trans. Geosci. Remote Sens. 57, 4322–4331. https://doi.org/10.1109/TGRS.2018.2890646.

Alonso-Arroyo, A., Camps, A., Monerris, A., Rüdiger, C., Walker, J.P., On-rubia, R., Querol, J., Park, H., Pascual, D., 2016. On the correlation between GNSS-R reflectivity and L-band microwave radiometry. IEEE J. Select.Top. Appl. Earth Observ. Remote Sens. 9, 5862–5879. https://doi.org/10.1109/JSTARS.2016.2588281.

Aubert, M., Baghdadi, N., Zribi, M., Douaoui, A., Loumagne, C., Baup, F., El Hajj, M., Garrigues, S., 2011. Analysis of TerraSAR-X data sensitivity to bare soil moisture, roughness, composition and soil crust. Remote Sens. Environ. 115, 1801–1810. https://doi.org/10.1016/j.rse.2011.02.021.

Camps, A., Park, H., Pablos, M., Foti, G., Gommenginger, C.P., Liu, P.W., Judge, J., 2016. Sensitivity of GNSS-R spaceborne observations to soil moisture and vegetation. IEEE J. Select.Top. Appl. Earth Observ. Remote Sens. 9, 4730–4742. https://doi.org/10.1109/JSTARS.2016.2588467.

Chew, C.C., Small, E.E., 2018. Soil moisture sensing using spaceborne GNSS reflections: comparison of CYGNSS reflectivity to SMAP soil moisture. Geophys. Res. Lett. 45, 4049–4057. https://doi.org/10.1029/2018GL077905.

Chew, C., Shah, R., Zuffada, C., Hajj, G., Masters, D., Mannucci, A.J., 2016. Demonstrating soil moisture remote sensing with observations from the UK TechDemoSat-1 satellite mission. Geophys. Res. Lett. 43, 3317–3324. https://doi.org/10.1002/2016GL068189.

Choudhury, B.J., Schmugge, T.J., Chang, A., Newton, R.W., 1979. Effect of surface roughness on the microwave emission from soils. J. Geophys. Res. 89, 5699–5706. https://doi.org/10.1029/jc084ic09p05699.

Clarizia, M.P., Ruf, C.S., Jales, P., Gommenginger, C., 2014. Spaceborne GNSS-R minimum variance wind speed estimator. IEEE Trans. Geosci. Remote Sens. 52, 6829–6843. https://doi.org/10.1109/TGRS.2014.2303831.

Clarizia, M.P., Pierdicca, N., Costantini, F., Floury, N., 2019. Analysis of CYGNSS data for soil moisture retrieval. IEEE J. Select.Top. Appl. Earth Observ. Remote Sens. 12, 2227–2235. https://doi.org/10.1109/JSTARS.2019.2895510.

De Roo, R.D., Ulaby, F.T., 1994. Bistatic specular scattering from rough dielectric surfaces. IEEE Trans. Antennas Propag. 42, 220–231. https://doi.org/10.1109/8.277216.

Dobson, M.C., Ulaby, F.T., Hallikainen, M.T., El-Rayes, M.A., 1985. Microwave dielectric behavior of wet soil-part II: dielectric mixing models. IEEE Trans. Geosci. Remote Sens. GE-23, 35–46. https://doi.org/10.1109/TGRS.1985.289498.

Edokossi, K., Calabia, A., Jin, S., Molina, I., 2020. GNSS-reflectometry and remote sensing of soil moisture: a review of measurement techniques, methods, and applications. Remote Sens. 12. https://doi.org/10.3390/rs12040614.

Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., Kimball, J., Piepmeier, J.R., Koster, R.D., Martin, N., McDonald, K.C., Moghaddam, M., Moran, S., Reichle, R., Shi, J.C., Spencer, M.W., Thurman, S.W., Tsang, L., Van Zyl, J., 2010. The soil moisture active passive (SMAP) mission. Proc. IEEE 98, 704–716. https://doi.org/10.1109/JPROC.2010.2043918.

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

Gerlein-Safdi, C., Ruf, C.S., 2019. A CYGNSS-based algorithm for the detection of inland waterbodies. Geophys. Res. Lett. 46, 12065–12072. https://doi.org/10.1029/2019gl085134.

Ismail, A.M., 1991. Soil properties and moisture characteristics and their relationship with crop mid-day stress in the Sudan Gezira. GeoJournal 23, 233–237. https://doi.org/10.1007/BF00204840.

Jia, Y., Jin, S., Savi, P., Gao, Y., Tang, J., Chen, Y., Li, W., 2019. GNSS-R soil moisture retrieval based on a XGboost machine learning aided method: performance and validation. Remote Sens. 11, 1655. https://doi.org/10.3390/rs11141655.

Jin, H., Henderson, B., 2011. Towards a daily soil moisture product based on incomplete time series observations of two satellites. In: Chan, F., Marinova, D., Anderssen, R. (Eds.), MODSIM. 2011. pp. 1959–1965.

Jin, S., Komjathy, A., 2010. GNSS reflectometry and remote sensing: new objectives and results. Adv. Space Res. 46, 111–117. https://doi.org/10.1016/J.ASR.2010.01.014.

Jin, S., Feng, G., Gleason, S., 2011. Remote sensing using GNSS signals: current status and future directions. Adv. Space Res. 47, 1645–1653. https://doi.org/10.1016/J.ASR.2011.01.036.

Katzberg, S.J., Torres, O., Grant, M.S., Masters, D., 2006. Utilizing calibrated GPS reflected signals to estimate soil reflectivity and dielectric constant: results from SMEX02. Remote Sens. Environ. 100, 17–28. https://doi.org/10.1016/J.RSE.2005.09.015.

Kerr, Y.H., Waldteufel, P., Wigneron, J.P., Martinuzzi, J.M., Font, J., Berger, M., 2001. Soil moisture retrieval from space: the soil moisture and ocean salinity (SMOS) mission. IEEE Trans. Geosci. Remote Sens. 39, 1729–1735. https://doi.org/10.1109/36.942551.

Kim, H., Lakshmi, V., 2018. Use of cyclone global navigation satellite system (CyGNSS) observations for estimation of soil moisture. Geophys. Res. Lett. 45, 8272–8282. https://doi.org/10.1029/2018GL078923.

Masters, D., Zavorotny, V., Katzberg, S., Emery, W., 2000. GPS signal scattering from land for moisture content determination. In: International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 3090–3092. https://doi.org/10.1109/igarss.2000.860346.

O'Neill, P.E., Chan, S., Njoku, E.G., Jackson, T., Bindlish, R., 2019. SMAP L3 radiometer global daily 36 km ease-grid soil moisture, Version 6. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/EVYDQ32FNWTH. [14-10-2019].

Paloscia, S., Pettinato, S., Santi, E., Notarnicola, C., Pasolli, L., Reppucci, A., 2013. Soil moisture mapping using Sentinel-1 images: algorithm and preliminary validation. Remote Sens. Environ. 134, 234–248. https://doi.org/10.1016/j.rse.2013.02.027.

Rivas, M., Maslanik, J., Axelrad, P., 2010. Bistatic scattering of GPS signals off Arctic Sea ice. IEEE Trans. Geosci. Remote Sens. 48, 1548–1553. https://doi.org/10.1109/TGRS.2009.2029342.

Rodriguez-Alvarez, N., Bosch-Lluis, X., Camps, A., Vall-Llossera, M., Valencia, E., Marchan-Hernandez, J.F., Ramos-Perez, I., 2009. Soil moisture retrieval using GNSS-R techniques: experimental results over a bare soil field. IEEE Trans. Geosci. Remote Sens. 47, 3616–3624. https://doi.org/10.1109/TGRS.2009.2030672.

Rodriguez-Alvarez, N., Akos, D.M., Zavorotny, V.U., Smith, J.A., Camps, A., Fairall, C.W., 2013. Airborne GNSS-R wind retrievals using delay–Doppler maps. IEEE Trans. Geosci. Remote Sens. 51, 626–641. https://doi.org/10.1109/TGRS.2012.2196437.

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

Rodríguez-Fernández, N.J., Aires, F., Richaume, P., Kerr, Y.H., Prigent, C., Kolassa, J., Cabot, F., Jiménez, C., Mahmoodi, A., Drusch, M., 2015. Soil moisture retrieval using neural networks: application to SMOS. IEEE Trans. Geosci. Remote Sens. 53, 5991–6007.

Ruf, C.S., Atlas, R., Chang, P.S., Clarizia, M.P., Garrison, J.L., Gleason, S., Katzberg, S.J., Jelenak, Z., Johnson, J.T., Majumdar, S.J., O'brien, A., Posselt, D.J., Ridley, A.J., Rose, R.J., Zavorotny, V.U., 2016. New ocean winds satellite mission to probe hurricanes and tropical convection.

Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., Teuling, A.J., 2010. Investigating soil moisture-climate interactions in a changing climate: A review. https://doi.org/10.1016/j.earscirev. 2010.02.004.

Wang, T., Ruf, C.S., Block, B., McKague, D.S., Gleason, S., 2019. Design and performance of a GPS constellation power monitor system for improved CYGNSS L1B calibration. IEEE J. Select.Top. Appl. Earth Observ. Remote Sens. 12, 26–36. https://doi.org/10.

1109/JSTARS.2018.2867773.
Yan, Q., Huang, W., 2016. Spaceborne GNSS-R sea ice detection using delay-Doppler maps: first results from the U.K. TechDemoSat-1 mission. IEEE J. Select.Top. Appl. Earth Observ. Remote Sens. 9, 4795–4801. https://doi.org/10.1109/JSTARS.2016.2582690.
Yan, Q., Huang, W., Moloney, C., 2017. Neural networks based sea ice detection and concentration retrieval from GNSS-R delay-Doppler maps. IEEE J. Select.Top. Appl. Earth Observ. Remote Sens. 10, 3789–3798. https://doi.org/10.1109/JSTARS.2017.2689009.
Yan, Q., Huang, W., Foti, G., 2018. Quantification of the relationship between sea surface roughness and the size of the glistening zone for GNSS-R. IEEE Geosci. Remote Sens. Lett. 15, 237–241. https://doi.org/10.1109/LGRS.2017.2782728.
Zavorotny, V.U., Voronovich, A.G., 2000a. Bistatic GPS Signal Reflections at Various Polarizations from Rough Land Surface with Moisture Content, in: International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 2852–2854. https://doi.org/10.1109/igarss.2000.860269.
Zavorotny, V.U., Voronovich, A.G., 2000b. Scattering of GPS signals from the ocean with wind remote sensing application. IEEE Trans. Geosci. Remote Sens. 38, 951–964. https://doi.org/10.1109/36.841977.
Zavorotny, V.U., Gleason, S., Cardellach, E., Camps, A., 2014. Tutorial on remote sensing using GNSS Bistatic radar of opportunity. IEEE Geosci. Remote Sens. Magaz. 2, 8–45. https://doi.org/10.1109/MGRS.2014.2374220.
Zhan, X., Zheng, W., Fang, L., Liu, J., Hain, C., Yin, J., Ek, M., 2016. A preliminary assessment of the impact of SMAP soil moisture on numerical weather forecasts from GFS and NUWRF models. In: Int. Geosci. Remote Sens. Symp., Institute of Electrical and Electronics Engineers Inc, pp. 5229–5232. https://doi.org/10.1109/IGARSS.2016.7730362.
Zheng, W., Zhan, X., Liu, J., Ek, M., 2018. A preliminary assessment of the impact of assimilating satellite soil moisture data products on NCEP global forecast system. Adv. Meteorol. 2018.