Extraction of Soil Moisture Based GNSS-R Considering Vegetation Factors

Qinyu Guo[1,2], Shuangcheng Zhang[1,2(✉)], Qi Liu[1], Zhongmin Ma[1], Ning Liu[1], Shengwei Hu[1], Lin Bao[1], Xin Zhou[1], Hebin Zhao[1], Lifu Wang[3], and Tianhe Wan[3]

[1] College of Geology Engineering, Chang'an University, Xi'an 710054, China
shuangcheng369@chd.edu.cn
[2] State Key Laboratory of Geo-Information Engineering, Xi'an 710054, China
[3] Altay National Reference Meteorological Station, China Meteorological Administration, Altay 836500, China

Abstract. One of the keys to retrieve soil moisture (SM) using the Spaceborne Global Navigation Satellite System-Reflectometry (GNSS-R) technique is to correct for the influence of vegetation. In this paper, the surface reflectivity is calculated using the Cyclone Global Navigation Satellite System (CYGNSS) data, and combine the Vegetation Water Content (VWC) provided by Soil Moisture Active Passive (SMAP) data to establish linear regression model to retrieve SM products with a temporal resolution of 3 days and a spatial resolution of 36 km on a pantropical scale, and each of the models is parameterized pixel-by-pixel to allow for tuning in accordance with regional variations. According to the experimental findings, CYGNSS may offer useful SM estimations across regions with moderate vegetation, and the correlation coefficient (R) with SMAP reference data is up to 0.7. However, in the arid and densely vegetated regions, the retrieval performance is degraded, and the R is 0.4 and 0.3 in the forest and bare soil areas, respectively. The overall root mean square error (RMSE) is 0.042 cm^3/cm^3. In addition, a time-series comparison of in-situ data from the International Soil Moisture Network (ISMN) and the CYGNSS SM revealed a good correlation. The study proves the necessity of considering vegetation effect in SM retrieval, which is of positive significance for the promotion of the operational application of Spaceborne GNSS-R SM retrieval.

Keywords: Spaceborne GNSS-R · CYGNSS · Soil moisture · Vegetation water content · Linear regression model

1 Introduction

Soil moisture (SM) is an important parameter in the energy exchange process between land and atmosphere, plays a very important role in the climate system [1]. Traditional SM monitoring techniques, such as drying and time-domain reflectometry, have great accuracy but are difficult to use for large-scale monitoring because of the high time consumption and equipment cost [2]. The growth of satellite-based remote sensing technology provides a new opportunity for continuous acquisition of large-scale SM data [3,

4]. In recent years, Spaceborne GNSS-R technology has gradually become a research hotspot to obtain large-scale and continuous SM data. By processing GNSS(Global Navigation Satellite System) satellite signals reflected from the earth surface, this technology can be used to retrieve geophysical parameters of ocean and land. GNSS-R SM detection has many advantages, such as low cost, wide coverage and a large number of signal sources, which can realize all-day and all-weather monitoring [5–8].

At present, a large number of literatures have studied the feasibility and effectiveness of SM retrieval using satellite-borne GNSS-R. Chew et al. used TDS-1 (TechDemoSat-1) data to prove that space-borne GNSS-R can be used to retrieve SM [9]. However, due to the long revist time of the TDS-1 satellite and the limited space coverage, the amount of data collected cannot achieve daily estimates of SM. By contrast,the CYGNSS (Cyclone Global Navigation Satellite System) satellite constellation launched by NASA in 2016 has short revisit time and large data volume. Previous studies have shown that the correlation between time fluctuation of CYGNSS signal and SM is better than that of TDS-1 data [10], which brings a new opportunity for obtaining SM with high spatial and temporal resolution. Chew et al. found a strong positive correlation between the change amount of CYGNSS reflectivity and SMAP SM, proving that CYGNSS can be used to develop global SM products with high temporal resolution (maybe every 6 h) [10]. Clarizia et al. proposed a triple linear regression algorithm of "Reflectivity—Vegetation—Roughness" to retrieve SM [11]. Yan Qingyun et al. improved Clarizia et al. 's algorithm and used CYGNSS DDM statistical moment to characterize the surface roughness information, reducing the dependence on external auxiliary data [12]. Guo Fei et al. considered the influence of surface temperature on space-borne GNSS-R SM retrieval, and used a linear regression model to retrieve the quasi-global SM based on CYGNSS data, proving the necessity of surface temperature in SM retrieval [13]. In addition, with the rapid development of machine learning, artificial neural network, random forest and support vector regression are also widely used in SM retrieval [14–16].

The retrieval of SM by machine learning method can flexibly select the input parameters and deal with the nonlinear relationship between the parameters. However, it is often faced with problems such as high dependence on auxiliary data, large amount of training data required by the model and poor generalization ability of the training model. In contrast, the method of SM retrieval based on empirical model is less dependent on auxiliary data, and can show a clearer relationship between CYGNSS observation data and SM. Moreover, it is more convenient to establish the retrieval model grid by grid for local parameterization. In addition, existing studies have shown that there is a significant correlation difference between different vegetation parameters and the attenuation effect of microwave signals. For example, the Vegetation Optical Depth (VOD) data of vegetation provided by SMAP L3 products has a poor correlation with the attenuation of vegetation at microwave frequencies [11]. Based on this, this paper attempted to use the data of Vegetation Water Content (VWC) in SMAP and CYGNSS observations to establish a linear regression model for SM retrieval. Finally, the retrieval results were compared with the reference data.

2   Observation Data

2.1   CYGNSS GNSS-R Data

CYGNSS is an Earth observation mission of NASA, which was launched on December 15, 2016. A total of eight small satellites are equipped with four-channel GNSS-R dual-base radar receivers that collect GPS (Global Position System) signals reflected from the ground and ocean [17]. These small satellites operate over the pan tropics with an orbital inclination of 35 degrees, covering about 38° N to 38° N, and the average revisit period is only 7 h [18]. The study uses CYGNSS Level 1 v3.1 scientific data products. In order to improve the retrieval accuracy, this paper carried out quality control on the CYGNSS data set. In addition to using the quality flag, the following data were also filtering: (1) the incidence angle greater than 65°; (2) SNR less than 0dB; (3) Receiver antenna gain less than 0dB; (4) Peak DDM values outside of delay bins 4 and 15 were filtered out.

2.2   SMAP Data

This paper uses the SMAP L3 SM products, which has a spatial resolution of 36km. The daily data consists of two parts: descending (AM) and ascending (PM) [4]. In this paper, "SM" and "VWC" data are used as auxiliary parameters for model training and prediction. In order to fully cover the research area, the data of three consecutive days are averaged [12]. Also, CYGNSS data are gridded to the same EASE-Grid as SMAP data for later comparison and verification [11]. Figure 1 shows the 2020 mean data of SMAP SM and VWC and CYGNSS reflectivity.
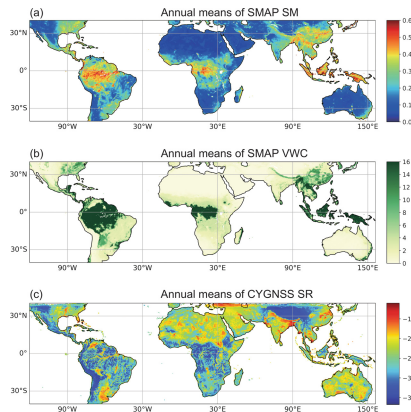

Fig. 1. Annual means of: (a) SMAP SM and (b) SMAP VWC, and (c) CYGNSS reflectivity

2.3   ISMN Data

In this paper, the International Soil Moisture Network (ISMN) dataset was used to independently verify the CYGNSS SM. Due to the limited penetration depth of L-band

in soil, the observation value of 5cm SM was adopted, and only the data labeled "good" of hourly SM was retained, which was processed as three-day average and compared with CYGNSS retrieval results [19].

3 Methodology

By receiving and processing L-band electromagnetic signals transmitted by GPS satellites, CYGNSS can retrieve the relevant parameters of the scattered surface. The GNSS signal power reflected by the surface consists of two parts: coherent and incoherent scattering component. Based on the assumption that CYGNSS land surveys are primarily specular point reflections, that is, the coherent scattering component is the dominant component in the reflected signal, then the power expression of the coherent component is as follows [20]:

$$P_{rl}^{coh} = \frac{P_r^t G^t}{4\pi (R_t + R_r)^2} \frac{G^r \lambda^2}{4\pi} \Gamma_{rl} \tag{1}$$

where, P_{rl}^{coh} is the coherent component; P_r^t is the transmitting power of the signal; G^t and G^r are the gain of transmitting antenna and receiving antenna respectively; R_t and R_r are respectively the distance between the transmitter and the receiver and the specular point; λ is the wavelength.

Then the surface reflectivity Γ_{rl} can be calculated by CYGNSS BRCS (bistatic radar cross section) σ [14]:

$$\Gamma_{rl} = \frac{\sigma (R_t + R_r)^2}{4\pi (R_t R_r)^2} \tag{2}$$

By correcting the signal frequency attenuation effect caused by surface roughness and vegetation cover, Fresnel reflection coefficient is calculated as follows:

$$\Gamma_{rl}(\theta) = R_{rl}(\theta)^2 \gamma^2 exp\left(-4k^2 s^2 \cos^2(\theta)\right) \tag{3}$$

where, θ is the incident angle; R_{rl} is the Fresnel reflection coefficient; Transmissivity γ indicates the attenuation caused by vegetation on signal transmission. The exponential term, where k is the signal wave number and s is the surface root mean square height, represents the impact of surface roughness.

Then, the relationship between reflection coefficient and dielectric constant was established by Fresnel reflection equation. Finally, the dielectric model was used to retrieve the SM.

Although Eq. (3) establishes an retrieval model of surface reflectivity considering the influence of roughness and vegetation, it is still difficult to use specific coefficient equations for SM retrieval on a global scale due to the absence of auxiliary data with high enough spatial and temporal resolution and precision. Therefore, the linear regression model was used in this paper to estimate SM by grid. In order to improve the spatial coverage of the retrieval results, the CYGNSS observations were grid based on a three-day period. This paper ignores the variation of roughness within the study time range,

but considers the influence of vegetation through the variation of VWC. The retrieval model is as follows:

$$CYGNSS\ SM\ =\ a \cdot \Gamma_{rl} + b \cdot VWC + c \tag{4}$$

where, a, b, and c are the to-be-determined coefficients.

4 Results and Analysis

4.1 Evaluation at Quasi-global Scale

Figure 2 shows the calculation results of the R and RMSE of the CYGNSS SM and SMAP SM for each grid, where the values are 0.55 and 0.04 cm^3/cm^3, respectively. The temporal correlation between CYGNSS SM and SMAP SM varies by region, with generally higher correlations and lower RMSE in medium vegetation coverage and semi-arid regions. For highly vegetated areas the R is significantly reduced, mostly around 0.4, and the root mean square error can even reach 0.07 cm^3/cm^3. This may be because dense vegetation will cause more attenuation of the microwave signal frequencies, making the urface reflectivity less sensitive to soil moisture. Moreover, the data quality of SMAP is also a factor leading to the poor correlation between CYGNSS SM and SMAP SM in these regions.
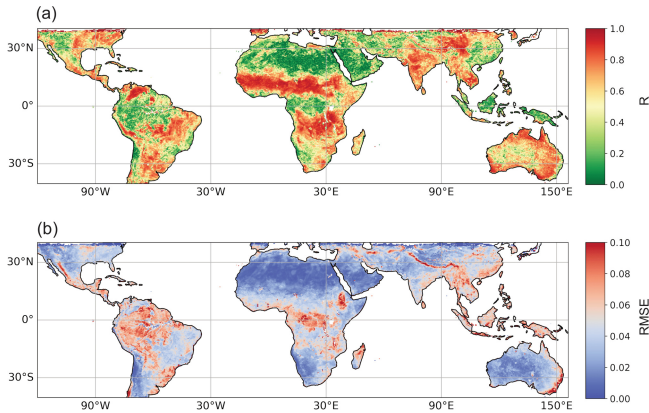

Fig. 2.　(a) Temporal correlation and (b) RMSE between CYGNSS SM and SMAP SM

Figure 3 shows the data of land types in the pan-global area. To further demonstrate how vegetation coverage affects the accuracy of CYGNSS SM data, the R between CYGNSS SM and SMAP SM of different land types were aggregated to obtain the mean value, as shown in Fig. 4. For the regions such as savanna, grassland and farmland, the R is relatively high, generally above 0.65. However, due to the influence of vegetation and roughness, the coherent scattering component in the reflected signal decreases, while

the incoherent scattering component increases. Therefore, for the regions with dense or low vegetation coverage, the R decreased significantly, generally below 0.5, further indicates that the retrieval performance of CYGNSS SM is improved when the vegetation coverage changes from low/high to medium.
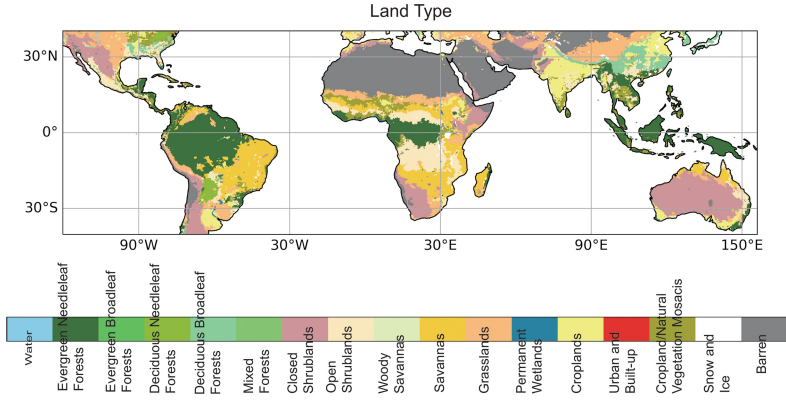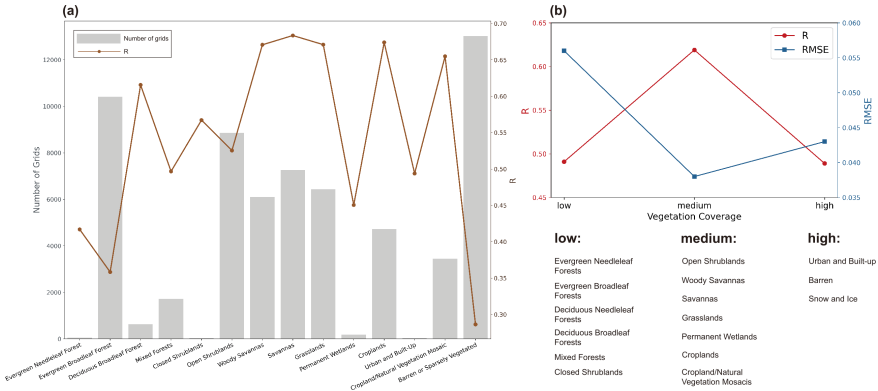

Fig. 3. Land type data

Fig. 4. (a) Means of R based on different land types. Each lans type's number of grids is indicated by a gray bar. (b) Means of R and RMSE based on different vegetation coverage.

Several grids were randomly selected for time series analysis. As shown in Fig. 5, CYGNSS SM can well reflect the dynamic variation trend of SMAP SM over time, but the measurement accuracy needs to be improved. During certain periods, CYGNSS SM was unable to capture detailed change information. In general, CYGNSS SM and SMAP SM have good consistency. In the selected grids, the mean R is 0.8, and the mean RMSE is 0.049 cm^3/cm^3.
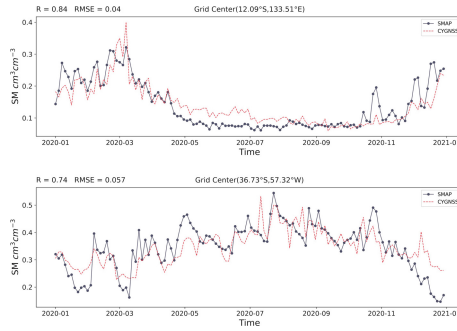
Fig. 5. SM time series between CYGNSS and SMAP in random grids

4.2 Validation at In-situ Sites

Comparison is made between the CYGNSS retrieval results and the in-situ measurements carried out at ISMN sites. Figure 6 shows the time series analysis data of SM at six representative sites. In the stable SM period, the fitting degree between CYGNSS SM and ISMN SM was better. However, when SM changes greatly, CYGNSS SM will give higher or lower estimates. In contrast, the correlation between SMAP SM and ISMN SM is better, but in general, CYGNSS SM and ISMN SM have a good consistency. At all sites, the mean R between the two was 0.68, and the mean RMSE was 0.073 cm^3/cm^3.
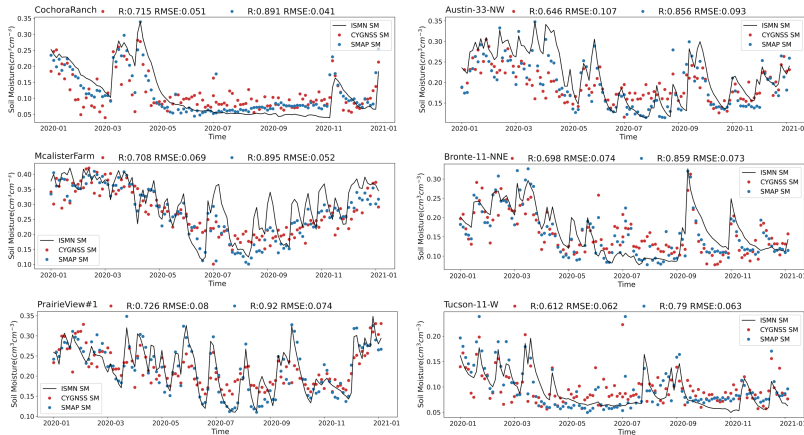

Fig. 6. SM time series between CYGNSS, SMAP, and ISMN at six sites

5 Conclusion and Discussion

In this paper, based on the surface reflectivity of CYGNSS and the VWC data provided by SMAP, the linear regression modele were established to retrieve the pan-global SM products in 2020. The correlation between the retrieval results and the SMAP reference

data is closely related to the vegetation density in the region, and the retrieval performance is the best in the region with the medium vegetation coverage. In addition, a time series analysis was conducted between the retrieval results and SM from ISMN measured sites, which showed a good consistency between them in the selected sites.

CYGNSS has an average revisit period of only 7.2 h. In this paper, in order to improve the spatial coverage of data, three consecutive days of CYGNSS observation values were grid, which resulted in the waste of high time resolution of CYGNSS data to a certain extent. However, high temporal resolution and high spatial resolution are often incompatible. Therefore, it is necessary to balance the temporal resolution and spatial resolution of data according to the demand in practical application.

There is a correlation difference between different vegetation parameters and the attenuation effect of microwave signal frequency. Previous studies mostly focused on the correction of the retrieval effect by the VOD. This paper attempts to establish a model using the VWC parameter for the first time, although good results have been achieved in the area of moderate vegetation coverage, the retrieval accuracy in other areas needs to be improved. In the follow-up study, more vegetation parameters such as leaf area index (LAI) and aboveground biomass can be considered to analyze the influence of vegetation on the retrieval process.

References

1. Dobriyal, P., Qureshi, A., Badola, R., et al.: A review of the methods available for estimating soil moisture and its implications for water resource management. J. Hydrol. **458**, 110–117 (2012)
2. Western, A.W., Blöschl, G., Grayson, R.B.: Geostatistical characterisation of soil moisture patterns in the Tarrawarra catchment. J. Hydrol. **205**(1–2), 20–37 (1998)
3. Kerr, Y.H., Waldteufel, P., Wigneron, J.-P., et al.: Soil moisture retrieval from space: the soil moisture and ocean salinity (SMOS) mission. IEEE Trans. Geosci. Remote Sens. **39**(8), 1729–1735 (2001)
4. Entekhabi, D., Njoku, E.G., O'Neill, P.E., et al.: The soil moisture active passive (SMAP) mission. Proc. IEEE **98**(5), 704–716 (2010)
5. Tao, T., Li, J., Zhu, Y., et al.: Spaceborne GNSS-R for retrieving soil moisture based on the correction of stage model. Acta Geodaetica et Cartographica Sinica **51**(9), 1942–1950 (2022)

6. Carreno-Luengo, H., Ruf, C.S.: Retrieving freeze/thaw surface state from CYGNSS measurements. IEEE Trans. Geosci. Remote Sens. **60**, 1–13 (2022)
7. Liu, Q., Zhang, S., Nan, Y., et al.: Flood detection of south asia using spaceborne GNSS-R coherent signals. Geomatics Inf. Sci. Wuhan Univ. **46**(11), 1641–1648 (2021)
8. Bu, J., Yu, K., Han, S.: Construction of spaceborne GNSS-R ocean waves significant wave height retrieval model. Acta Geodaetica et Cartographica Sinica **51**(9), 1920–1930 (2022)
9. Chew, C., Rashmi, S., Cinzia, Z.: Demonstrating soil moisture remote sensing with observations from the UK TechDemoSat-1 satellite mission. Geophys. Res. Lett. **43**(7), 3317–3324 (2016)
10. Chew, C.C., Small, E.E.: Soil moisture sensing using spaceborne GNSS reflections: comparison of CYGNSS reflectivity to SMAP soil moisture. Geophys. Res. Lett. **45**(9), 4049–4057 (2018)
11. Clarizia, M.P., Pierdicca, N., Costantini, F., et al.: Analysis of CYGNSS data for soil moisture retrieval. IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens. **12**(7), 2227–2235 (2019)
12. Yan, Q., Huang, W., Jin, S., et al.: Pan-tropical soil moisture mapping based on a three-layer model from CYGNSS GNSS-R data. Remote Sens. Environ. **247**, 111944 (2020)
13. Zhu, Y., Guo, F., Zhang, X.: Effect of surface temperature on soil moisture retrieval using CYGNSS. Int. J. Appl. Earth Obs. Geoinf. **112**, 102929 (2022)
14. Eroglu, O., Kurum, M., Boyd, D., et al.: High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks. Remote Sens. **11**(19), 2272 (2019)
15. Hu, Y., Wang, J., Li, Z., et al.: Land surface soil moisture along sichuan-tibet traffic corridor retrieved by spaceborne global navigation satellite system reflectometry. Earth Sci. **47**(6), 2058–2068 (2022)
16. Lei, F., Senyurek, V., Kurum, M., et al.: Quasi-global machine learning-based soil moisture estimates at high spatio-temporal scales using CYGNSS and SMAP observations. Remote Sens. Environ. **276**, 113041 (2022)
17. Ruf, C.S., Chew, C., Lang, T., et al.: A new paradigm in earth environmental monitoring with the CYGNSS small satellite constellation. Sci. Rep. **8**(1), 8782 (2018)
18. Ruf, C.S., Atlas, R., Chang, P.S., et al.: New ocean winds satellite mission to probe hurricanes and tropical convection. Bull. Am. Meteor. Soc. **97**(3), 385–395 (2016)
19. Vreugdenhil, M., Gruber, A., Hegyiová, A., et al.: Global automated quality control of in situ soil moisture data from the international soil moisture network. Vadose Zone J. **12**(3) (2013)
20. De Roo, R.D., Ulaby, F.T.: Bistatic specular scattering from rough dielectric surfaces. IEEE Trans. Antennas Propag. **42**(2), 220–231 (1994)