Use of Cyclone Global Navigation Satellite System (CyGNSS) Observations for Estimation of Soil Moisture

Hyunglok Kim[1] and Venkat Lakshmi[1]

[1]School of Earth, Ocean and the Environment, University of South Carolina, Columbia, SC 29208, USA

Corresponding author: Hyunglok Kim (hkim@geol.sc.edu)

Key Points:

- CyGNSS-derived signal-to-noise ratio data was utilized for soil moisture estimation
- CyGNSS data can fill the gap of missing spatial and temporal values in existing satellite-based soil moisture retrieval systems
- By combining CyGNSS and SMAP datasets, reliable daily soil moisture estimates from space can be achieved

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2018GL078923

Abstract

Using the first full annual cycle of Cyclone Global Navigation Satellite System (CyGNSS) observations, we investigated the limitations and capabilities of CyGNSS observations for soil moisture (SM) estimates (0-5 cm). A relative signal-to-noise ratio (rSNR) value from a CyGNSS-derived delay-Doppler map is introduced to improve the temporal resolution of SM derived from Soil Moisture Active Passive (SMAP) data. We then evaluated the CyGNSS-derived rSNR using ground-based SM measurements and the triple collocation method with SMAP and modeled SM products. We found that CyGNSS can provide useful SM estimates over moderately vegetated regions (correlation coefficient of the individual data (R_i): 0.77) but shows degraded performance over arid and densely vegetated regions (R_i: 0.68 and 0.67). However, when rSNR data is combined with SM data from SMAP, daily SM estimates can be achieved. These results show that synergistic use of CyGNSS observations can improve on SM estimates from current satellite systems.

Keywords: surface soil moisture, CyGNSS, Global Navigation Satellite Systems Reflectometry (GNSS-R), SMAP, triple collocation, signal-to-noise ratio (SNR)

1 Introduction

Over the last several decades, researchers have proposed various methods for estimating near-surface soil moisture (SM) values via satellite microwave instruments (Karthikeyan et al., 2017). This work is critical since SM estimates at regional scales are necessary for operational applications such as water resource and irrigation management, near-real-time numerical weather prediction, hydrological modeling, and many other surface processes (Seneviratne et al., 2010; Brocca et al., 2017). In particular, the active and passive satellite-based microwave SM estimation missions, including Soil Moisture and Ocean Salinity (SMOS) and Soil Moisture Active Passive (SMAP), have provided global-scale SM estimates (Entekhabi et al., 2010; Kerr et al., 2010).

In addition to satellite-based microwave SM estimation missions, researchers have also tested the viability of surface-reflected Global Navigation Satellite System (GNSS) signals in estimating surface SM (Masters, 2004; Camps et al., 2016). The first dedicated space-borne Global Positioning System (GPS) reflectometry (GPS-R) receiver on board the UK-Disaster Monitoring Constellation satellite (also known as BNSCSAT-1, launched in September 2003) proved that GNSS signals can reliably describe surface conditions including ocean, snow, and land surface properties (Gleason et al., 2005). GNSS is an umbrella term for satellite navigation systems, encompassing all global satellite positioning systems that provide autonomous geo-spatial positioning; thus, the GPS is one component of GNSS and GPS is now the most widely used GNSS in the world.

The major operational distinction between the GPS-R receiver and the active/passive microwave system is the source of the signals each uses to observe the earth's surface. While a passive sensor observes emissions coming from objects on Earth, an active sensor, also called radar, actively emits microwaves toward the earth's surface and observes the reflected signal.

The GPS-R receivers on board CyGNSS are passive sensors that receive surface-reflected GPS signals. In other words, the GPS-R receivers on board CyGNSS observe signals of opportunity from GNSS. The term *signals of opportunity* refers to the signals which enable CyGNSS to observe the earth's surface; that is, CyGNSS uses its GPS-R receivers to take advantage of signals from existing transmitter systems (i.e., GPS satellites) intended to observe objects on Earth.

Chew et al. (2016) and Camps et al. (2016) demonstrated the sensitivity of the GPS-R dataset on SM and vegetation cover using data from the UK TechDemoSat-1 (TDS-1), which was launched in July 2014. However, TDS-1-dervied data has important limitations in data acquisition, both spatially and temporally, because the TDS-1 GPS-R payload is only active for two out of every eight days (Clarizia et al., 2016). This limitation prevents us from estimating SM values with TDS-1 in a timely manner, as we cannot obtain sufficient datasets for daily SM variability. Lack of access to daily SM variability can produce a substantial bias in simulated surface water and energy fluxes, infiltration and surface runoff, etc. (Crow and Wood, 2002).

Researchers expected to overcome this crucial limitation with NASA's new weather prediction project, the Cyclone Global Navigation Satellite System (CyGNSS). In December 2016, eight CyGNSS micro-satellites were launched from a single launch vehicle. CyGNSS was designed to measure ocean surface wind field using a bistatic scatterometer technique with GPS-R receivers (Ruf et al., 2016a; Ruf et al., 2018). Each micro-satellite has four GPS-R receivers which function as delay-Doppler mapping instruments. A delay-Doppler map (DDM) is a 2-D image generated by diffuse scatter power from the earth's surface in the surroundings of the specular point. Consequently, the CyGNSS constellation generates 32 simultaneous DDMs per second (Clarizia & Ruf, 2016).

The most important feature of CyGNSS's observations is that in contrast to other well-known microwave-based SM retrieval satellite systems in sun-synchronous orbit (SSO) such as SMAP, SMOS, MetOp-A and B, and GCOM-W1, all of which have revisiting times of one to three days, the CyGNSS micro-satellites randomly receive surface-reflected GPS signals with revisiting times of 2.8 (median) and 7.2 (mean) hours per day (Ruf et al., 2016b). Utilizing this unprecedentedly high temporal resolution data could add significant value to existing microwave-based satellite SM retrieval systems. For example, a spatial and temporal gap of one to three days in a certain region in SMAP SM data could be filled with CyGNSS-derived SM information. Consequently, predictions of vast-scale water-related natural disasters could be improved since antecedent soil moisture information at lead times of less than one to three days is necessary to predict most water-related hazards (Brocca et al., 2017; Kim et al., 2017). Therefore, information on daily SM variability from space would contribute to improved prediction capabilities of natural disasters. **Figure S1** shows an example of the sampling frequencies of CyGNSS and SMAP over part of the contiguous United States (CONUS) area. The CyGNSS observation shows a clear potential to fill the gap in SSO satellite observations and to make continuous SM estimates possible.

In the present study, we used the first full annual cycle of CyGNSS-derived signal-to-noise ratio (SNR) observations to investigate the sensitivity of SNR to SM variability over the CONUS area (from March 2017 to March 2018). The overall objectives of our study are as follows: 1) to illustrate the possibility of retrieving SM by using the surface-reflected GNSS signals, 2) to evaluate the relationship between CyGNSS-derived SNR and SM using in-situ SM measurements in relation to different vegetation conditions, 3) to apply the CyGNSS-derived SNR in improving the frequency of SM sampling with SMAP data, and 4) to investigate the error patterns of CyGNSS-derived SNR using triple collocation (TC) analysis considering vegetation and land use effects. We believe that this research can provide novel

insight into the use of the recently available CyGNSS dataset in practical applications and data merging processes to better understand surface hydrologic cycles, while we acknowledge limitations in SM estimates from CyGNSS observations.

2 Datasets

In this study, we utilized five different satellite datasets: 1) Level 0 CyGNSS products, 2) Half-hourly Global Precipitation Measurement (GPM) with 0.1° spatial resolution from late-run Integrated Multi-satellitE Retrievals for GPM (IMERG) products Version 4 (Skofronick-Jackson et al., 2017), 3) Half-orbit SMAP Enhanced L3 Radiometer Global Daily 9 km EASE-Grid SM and 4) Vegetation Water Content (VWC) Version 1 (Chan et al., 2018), and 5) the IGBP Land Cover Types Classification from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors (Friedl et al., 2002). We also used 1) SM (0-10 cm) and 2) surface temperature (0-10 cm) from hourly NLDAS-2 with 0.125° spatial resolution (Xia et al., 2012). For the ground-based SM estimates, we employed hourly in-situ SM estimates from the International Soil Moisture Network (ISMN) (Dorigo et al., 2011). All datasets are masked out when the surface temperature is below zero. A detailed description of the CyGNSS dataset is provided below.

3 Methodology

3.1 CyGNSS-derived SNR values

The frequency transmitted from GPS satellites is a microwave band (L1-band, λ~19cm) with a reflected electromagnetic wave capable of conveying information about how much moisture exists in surface soil (Schmugge, 1986; Ulaby et al., 2014). The DDM generated from signals of opportunity from GNSS are known to have a direct relationship with surface SM dynamics (Voronovich and Zavorotny 2017). Each DDM is characterized by the scattering geometry (i.e., incidence and scattering angles), antenna gains, distance, and dielectric and statistical properties of the surface (Alonso-Arroyo et al., 2016).

In general, the scattering pattern of the land surface consists of both coherent and noncoherent components; thus, the DDM (defined as $\overline{|Y(\hat{\tau},\hat{f})|^2}$) general expression is composed of two terms:

$$\overline{\left|\mathrm{Y}(\hat{\tau},\hat{\mathrm{f}})\right|^2} = \overline{\left|\mathrm{Y}_{\mathrm{coh}}(\hat{\tau},\hat{\mathrm{f}})\right|^2} + \overline{\left|\mathrm{Y}_{\mathrm{ncoh}}(\hat{\tau},\hat{\mathrm{f}})\right|^2} \quad \text{Eq. (1)}$$

where $\hat{\tau}$ is a relative time delay and $\hat{f}$ is a Doppler frequency. In the present study, we assumed a homogenous, smooth land surface with different incidence angles. This means that we assumed that much of the signal originates from coherent reflections and ignored incoherent reflections because strong coherent scattering of GNSS signals has been shown over land (Carreno-Luengo et al., 2016). For the coherent component contribution part of the signal, the DDM can be expressed based on the Friis transmission formula and the Fresnel reflection coefficient of the equivalent smooth surface at $\hat{\tau}$ and $\hat{f}$ bin as follows (Ulaby et al., 2014; Voronovich and Zavorotny 2017):

$$\overline{\left|\mathrm{Y}_{\mathrm{coh}}(\hat{\tau},\hat{\mathrm{f}})\right|^2} = \frac{\mathrm{P_r^T}\lambda^2}{(4\pi)} \cdot \langle\left|\chi(\hat{\tau},\hat{\mathrm{f}})\right|^2\rangle \cdot \frac{120\pi \mathrm{G}^{\mathrm{T}}_{(\hat{\tau},\hat{\mathrm{f}})}, \mathrm{G}^{\mathrm{R}}_{(\hat{\tau},\hat{\mathrm{f}})}}{(4\pi)^2} \cdot \frac{\left|\bar{\mathrm{V}}(\hat{\tau},\hat{\mathrm{f}})\right|^2}{\left(\mathrm{R}^{\mathrm{R}}_{(\hat{\tau},\hat{\mathrm{f}})} + \mathrm{R}^{\mathrm{T}}_{(\hat{\tau},\hat{\mathrm{f}})}\right)^2} \quad \text{Eq. (2)}$$

where P_r^T is the transmitted Right Hand Circular Polarization (RHCP) power, λ is the wavelength (~19 cm) of the L1-band, χ is the Woodward Ambiguity Function (WAF), $\bar{G}^T_{(\hat{\tau},\hat{f})}$ and $G^R_{(\hat{\tau},\hat{f})}$ are the transmitter and the receiver antenna gains, $\bar{V}(\hat{\tau},\hat{f})$ is an average reflection coefficient, and $R^T_{(\hat{\tau},\hat{f})}$ and $R^R_{(\hat{\tau},\hat{f})}$ are the transmitter-to-surface and surface-to-receiver range losses.

Over land surface, the variability of $V(\hat{\tau},\hat{f})$, the reflection coefficient, is a function of soil wetness conditions and incidence angle (θ_i). In a natural land surface, the probability density function (PDF) of measured height of the ground surface is well modeled by a Gaussian

distribution. Under the method of the Kirchoff approximation, $V(\hat{\tau}, \hat{f})$ can be expressed with the Fresnel reflection coefficient and incidence angle as follows (Davies, 1954):

$$V_{F(\hat{\tau},\hat{f})}(s, \theta_i, \varepsilon) = \exp^{-4\Psi(s,\theta_i)^2} \cdot |R_{rl}(\varepsilon, \theta_i)|^2 \quad \text{Eq(3)}$$

$R_{rl}(\varepsilon, \theta_i)$ is a combination of vertical and horizontal polarization Fresnel reflection coefficients, which is a function of the dielectric constant of the land surface (ε) and the incidence angle (θ_i) (Zavorotny & Voronovich, 2000):

$$R_{rl}(\varepsilon, \theta_i) = \frac{R_{vv}(\theta_i) - R_{hh}(\theta_i)}{2} = \frac{(\varepsilon - 1)\sqrt{\varepsilon - \sin^2\theta_i}}{\left(\varepsilon\cos\theta_i + \sqrt{\varepsilon - \sin^2\theta_i}\right)\left(\cos\theta_i + \sqrt{\varepsilon - \sin^2\theta_i}\right)} \quad \text{Eq. (4)}$$

Ψ is function of surface rms height (s) and the incidence angle (θ_i) as follows (Ulaby et al., 2014):

$$\Psi = \frac{2\pi}{\lambda} s\cos\theta_i \quad \text{Eq. (5)}$$

The large difference in ε is the basis for estimating SM when a signal is sensitive to ε. The ε of dry and wet soil is approximately 3 and 25 respectively, and the electromagnetic roughness in Eq. (4) (i.e., $\frac{2\pi}{\lambda}s$) for the smooth and rough surfaces is approximately 0.2 and 0.5~1.94 (De Roo & Ulaby,1994). **Figure S2** shows how CyGNSS-derived reflection coefficient values can relate to the dielectric constant and consequently can be applied to soil wetness estimations. This illustrates the simulated reflection coefficient values $V(\hat{\tau}, \hat{f})$ with respect to different SM conditions for four different surface roughness values and different incidence angles; as SM increases, the reflection coefficient increases. Based on **Figure S2** and the equations (3) to (5), we see that CyGNSS-retrieved $\overline{|Y_{coh}(\hat{\tau}, \hat{f})|^2}$ can potentially detect surface SM variability with respect to different θ_i and s. This CyGNSS-derived $\overline{|Y_{coh}(\hat{\tau}, \hat{f})|^2}$ can be directly described using the SNR of the scattered signal--the ratio of the CyGNSS-observed reflect peaks of

$\overline{|Y(\hat{\tau},\hat{f})|^2}$ to the direct peaks of $\overline{|Y(\hat{\tau},\hat{f})|^2}$ power waveforms. A detailed calculation of compensation for the noise power floor and the antenna gains used to calculate SNR can be found in the CyGNSS handbook (Ruf et al., 2016b). In order to compare SNR with SMAP and NLDAS-2 datasets, we re-projected SNR observations into 9 km EASE-Grid cells of equivalently-sized enhanced SMAP grid projections and assumed each grid cell to be homogenous.

As shown in **Figure S2**, the strong variability of the reflection coefficient with the incidence angle θ_i may be a fundamental problem when SNR data is applied to SM estimates. For example, a lower θ_i reflected GNSS signal is related to a lower reflection coefficient value (different line colors) over similar surface wetness and roughness conditions. This indicates that SNR data from different θ_i cannot simply be averaged out when more than one SNR value observed from different θ_i exists in an intra-grid cell. Considering these issues, we propose a method for the normalization of the SNR value with respect to the reference incidence angle. All SNR values from various θ_i at (x_i, y_i) pixels (hereafter SNR(x_i, y_i, θ_i)) are normalized at SNR values associated to θ_i range [35° ± 5°] within the same pixel (hereafter $SNR_{ref}(x_i, y_i)$). A normalized SNR(θ_i) value at (x_i, y_i) pixel (hereafter $nSNR(x_i, y_i, \theta_i)$ can be obtained using equation (6) as follows:

$$nSNR(x_i, y_i, \theta_i) = \frac{(SNR(x_i, y_i, \theta_i) - \mu(SNR(x_i, y_i, [\theta_i \pm 5°])) \times \sigma(SNR_{ref}(x_i, y_i))}{\sigma(SNR(x_i, y_i, [\theta_i \pm 5°]))} + \mu(SNR_{ref}(x_i, y_i)) \quad \text{Eq. (6)}$$

where the μ represents the average over time and the sigma is the standard deviation of SNR datasets during the study period. This normalization process enables us to fully utilize SNR values from different incidence angles by matching the standard deviation of SNR data from various incidence angles to the reference angle's SNR dataset.

In this study, we employed a procedure and two important assumptions similar to Wagner et al. (1999a, 1999b) in order to derive the relative moisture content in the topsoil (0-5 cm). The criteria for the relative values were the historically collected maximum and minimum values of moisture content in the top soil layer (0-5 cm). First, we assumed that the maximum ($nSNR_{wet}$) and minimum ($nSNR_{dry}$) values of nSNR had been obtained within a full year of CyGNSS observations because the temporal resolution of CyGNSS is sufficient to catch most rainfall events. Thus, we were likely to encounter the conditions of maximum and minimum values of nSNR during the time series. Second, we assumed a linear relationship between the surface SM contents and nSNR. Wagner et al. (1999c) assumed that the backscattering coefficient was linearly related to the surface SM content in the presence of vegetation, referencing Dobson and Ulaby (1986); they found that the backscattering coefficient is linearly related to the surface SM content over bare soil. However, no previous research has been conducted to show the linear relationship between CyGNSS-derived nSNR values and SM contents. Therefore, we compared nSNR values with ground-based SM measurements to determine if such a linear relationship should exist. **Figure S3** shows six specific sites with six different vegetation conditions and four land cover types (Cropland/Natural, Cropland, Open Shrublands, and Grass lands). Based on **Figure S3**, we assumed a linear relationship between nSNR and the volumetric SM because we found them to be linearly dependent on ground-based SM measurements and to have linear correlation coefficient values between 0.58-0.73. We calculated the relative value of nSNR (hereafter rSNR) using the following equation (7):

$$\mathrm{rSNR}(x_i, y_i, t) = \frac{\mathrm{nSNR}(x_i, y_i, t) - \mathrm{nSNR}_{\mathrm{dry}}(x_i, y_i)}{\mathrm{nSNR}_{\mathrm{wet}}(x_i, y_i) - \mathrm{nSNR}_{\mathrm{dry}}(x_i, y_i)}\ \phi \quad \text{Eq. (7)}$$

where the x_i and y_i are a location of 9 km EASE-Grid projection at time t and ϕ is the soil porosity. The rSNR value can vary from 0 to 1, where 1 indicates all soil pores are filled with

liquid water and 0 indicates the SM content to be at or near the wilting point. We embrace the same hypothesis as Wagner et al. (1999c), that rSNR is equal to the degree of saturation in the top layers of the soil: the degree of saturation is the ratio of the volume of water contained in the soil and the volume of pores. Using this rationale as a base, we can obtain the volumetric SM by multiplying the rSNR value with the soil porosity value. In this study, we calculated the porosity values by applying the equations of Saxton and Rawls (2006) based on soil texture characteristics from the Harmonized World Soil Database.

3.2 Validation methodologies and statistical metrics

The CyGNSS rSNR and SMAP SM data were evaluated against point-scale ground-based SM estimates. We calculated the Pearson correlation coefficient (R) values and only considered the values at $p < 0.05$. However, this conventional statistical indicator, R-value, has four major limitations in validating satellite-based data against in-situ observations: (1) a measurement depth discrepancy occurs between the CyGNSS-derived reflectivity and in-situ sensors; (2) the in-situ observations cannot be considered as the true values since in-situ observations can have their own measurement uncertainty; (3) the in-situ observations are sparsely located, so they provide only limited regional satellite-derived data performance; and (4) the two pairs of datasets being compared represent quite different levels of spatial resolution (e.g., areal average value vs. point-scale value) (Dorigo et al., 2015; Zohaib et al., 2017).

In order to evaluate rSNR against ground-based SM estimates, we applied an exponential filter to rSNR values because the CyGNSS-derived reflectivity originates from the topsoil layer (0-5 cm), while in-situ SM sensors are installed at a certain depth below the surface (~10 cm). The exponential filter allows us to overcome the depth discrepancy between rSNR and ground-based measurements by estimating the average rSNR value over a layer in

the soil profile, also called the Soil Water Index (SWI). We calculated SWI from rSNR (hereafter $rSNR_{SWI}$) using the following recursive equation proposed by Albergel et al. (2008):

$$\mathrm{rSNR}_{\mathrm{SWI}_n} = \mathrm{rSNR}_{\mathrm{SWI}_{(n-1)}} + K_n\left(\mathrm{rSNR}_{\mathrm{SWI}}(t_n) - \mathrm{rSNR}_{\mathrm{SWI}_{(n-1)}}\right) \quad \text{Eq. (8)}$$

where $\mathrm{rSNR}_{\mathrm{SWI}_{(n-1)}}$ is the estimated profile rSNR at a time $t_{(n\text{-}1)}$, and $\mathrm{rSNR}_{\mathrm{SWI}}(t_n)$ is the estimated profile rSNR at a time t_n. The recursive form of gain K_n at a time t_n is calculated by Eq. (9):

$$K_n = \frac{K_{(n-1)}}{K_{(n-1)} + e^{-\left(\frac{t_n - t_{(n-1)}}{T}\right)}} \quad \text{Eq. (9)}$$

where T represents the characteristic time length in days. In previous studies, an optimum T (T_{opt}) value approach was proposed based on the Nash-Sutcliffe score to match the profile SM values at each in-situ station. In the present study, we also employed T_{opt} to match the depth of satellite- and ground-based SM data at each in-situ station. To initialize the exponential filter, $\mathrm{rSNR}_{\mathrm{SWI}_1}$ was set to $\mathrm{rSNR}(t_1)$ and K_1 was set to 1, following Albergel et al. (2008). The range of the gain K is [0, 1]. Similarly, we calculated SWI from SMAP SM (hereafter SM_{SWI}). For details about estimating the profile SM, please refer to Houser et al. (1998) and Walker et al. (2001).

Furthermore, we employed the TC method to evaluate large-scale rSNR values because the TC method does not require additional reference values as conventional metrics (Gruber et al., 2016). The TC method can address limitation (2) through (4), which we mentioned above. TC assumes independent errors, so we selected SM products with derivations as distinct as possible to avoid the chance of similarly retrieved SM values having partially correlated errors. That is, we calculated the TC-based R-value (R_i) with a triplet including radiometer-based SM and model-based SM products along with CyGNSS rSNR data.

3.3 Combined CyGNSS rSNR and SMAP SM

We combined CyGNSS rSNR and SMAP SM products to take advantage of the high sampling frequency (day^{-1}) of CyGNSS observations. To combine these two datasets, we needed to reconcile the systematic differences between SMAP SM and rSNR. These systematic differences between $rSNR_{SWI}$ and SMAP SM_{SWI} can be removed through the normalization of $rSNR_{SWI}$ (hearafter $rSNR_{norm}$) against a SMAP SM_{SWI} (Draper et al., 2009). We combined two datasets (hearafter cSM) after removing their systematic differences with the following Eq. (10):

$$\mathrm{cSM}(x_i, y_i) = \frac{\mathrm{rSNR}_{\mathrm{norm}}(x_i, y_i) + \mathrm{SMAP\ SM}_{\mathrm{SWI}}(x_i, y_i)}{2}$$

$$= \frac{\left[\left(\mathrm{rSNR}_{\mathrm{SWI}}(x_i, y_i) - \mu_{\mathrm{rSNR}_{\mathrm{SWI}}}\right) \times \frac{\sigma_{\mathrm{SMAP\ SM}_{\mathrm{SWI}}}}{\sigma_{\mathrm{rSNR}_{\mathrm{SWI}}}} + \mu_{\mathrm{SMAP\ SM}_{\mathrm{SWI}}}\right] + \mathrm{SMAP\ SM}_{\mathrm{SWI}}(x_i, y_i)}{2} \quad \text{Eq. (10)}$$

where μ represents the mean and σ is the standard deviation of the datasets over time during the study period at (x_i, y_i) pixel.

4 Results and Discussion

4.1 Sensitivity of rSNR to soil moisture and precipitation

Figures 1(a) and **(b)** show the time-averaged values of rSNR and SMAP SM with the longitudinal zonal mean value. The overall time-averaged values of the western CONUS area's rSNR and SMAP SM are relatively low when compared to the eastern CONUS areas. These rSNR and SMAP SM patterns correlate well with the spatial patterns of yearly average precipitation amounts. In **Figure 1(c)**, we only showed CyGNSS observations under ~37°N

since CyGNSS cannot observe latitudes higher than ~37°N. After the quality control processes, we lost a great deal of SMAP SM data over eastern CONUS; thus, it was difficult to make a spatial comparison analysis between CyGNSS rSNR and SMAP SM in some eastern CONUS areas. Most of the data from these areas were masked out because of mountainous terrains and dense vegetation areas since microwaves cannot penetrate high vegetation mass. These limitations also seem to cause erroneously low rSNR, possibly due to decreasing coherent scattering signals from densly vegeated surfaces. For example, we can see the erroneous values of rSNR by comparing maps of the average value of rSNR and GPM-derived precipitation. In **Figure 1(e)**, the average amount of precipitation over the southeast parts of CONUS is relatively high when compared to other areas. However, in **Figure 1(a)**, rSNR values over these areas showed low values of rSNR, as in other low-precipitation areas. In addition, the blue pixels or red pixels in **Figure 1(f)** represent areas where CyGNSS or SMAP show a higher sampling frequency; thus, some parts of CONUS are missing SMAP SM estimates but CyGNSS observations are still present. While this result shows a potential for CyGNSS to gap-fill missing SM values from other SM satellite missions, current rSNR values over some areas may have low data quality; thus, future studies will need to develop a masking process to define low-quality data in order to accurately estimate SM contents.

The time series of $rSNR_{SWI}$ (red dots), SMAP SM_{SWI} (green squares), and cSM (blue stars), as well as the in-situ SM (grey lines), are given in **Figures 2(a)**, **(b)**, and **(c)** for representative pixels with different vegetation conditions: low (VWC ≤1), moderate (1 ≤ VWC < 3), and high (VWC ≥ 3). **Figure S4** shows their locations in CONUS areas with the yearly average map of VWC. The in-situ observations in **Figures 2(a)**, **(b)**, and **(c)** were selected to provide an example of the time series of $rSNR_{SWI}$ values for different vegetation conditions: R-values of $rSNR_{SWI}$ against the in-situ observation are 0.65, 0.75, and 0.41 for

low, moderate, and high VWC, respectively. The $rSNR_{SWI}$ values respond to rainfall events showing peak values around rainfall days, and the $rSNR_{SWI}$ values show similar drying-down patterns with both the in-situ and SMAP SM_{SWI} values. Areas with erroneous responses of $rSNR_{SWI}$ to rainfall events might be caused by spatial resolution mismatch and/or the difference in data acquisition time between the in-situ and CyGNSS overpass times. For example, in **Figures 2(a) and 2(b)**, compared to in-situ SM values (green lines), the $rSNR_{SWI}$ (red dots) overestimates SM values from October to November 2017 in **Figure 2(a)** and from February to March 2018 in **Figure 2(b)**. In especially highly vegetated areas, $rSNR_{SWI}$ showed underestimation of SM patterns during the vegetation growing season (from July to October in **Figure 2(c)**).

The results of the average R-values and sampling frequency (day^{-1}) for CyGNSS $rSNR_{SWI}$ and SMAP SM_{SWI}, and for cSM against the in-situ SM, are shown in **Table S1**. In terms of average R-value, SMAP showed better performance than CyGNSS. It has been proven that L-band radiometry provides better sensitivity to SM than other instruments (Kerr et al., 2010). However, SMAP has a lower sampling frequency than CyGNSS, indicating that CyGNSS-derived rSNR has the potential to fill the temporal gap in SMAP SM estimates. When CyGNSS and SMAP are combined, the average R-values are 0.62 with a revisiting time of more than once per day ($f > 1\ day^{-1}$). The results of the combined product's R-values and a sampling frequency of more than once per day are very encouraging. These results show a strong potential for the synergistic use of CyGNSS data with passive or active sensor-based SM data; together, they can improve both the spatial and temporal resolution of SM retrievals.

4.2 Error pattern estimation from TC statistics

Since vegetation is one of the most important parameters to be considered in the SM retrieval algorithm, we evaluated the performance of SM estimation from rSNR using the TC metrics with regard to various ranges of VWC values. **Figure 3(a)** shows a map of the R_i-value of rSNR. The R_i-value is was proposed by McColl et al. (2014) as a means to investigate the linear correlation coefficient of the individual datasets. The average R_i-values are 0.68, 0.77, and 0.67 over the low, moderate, and high vegetation conditions, respectively. We observed better performance of rSNR over moderately vegetated areas than over sparsely and densely vegetated regions. The major land types of moderately vegetated pixels are grasslands and croplands (the middle pie chart in **Figure 3(b)**). The cropland-type regions might be an appropriate environmental condition for microwave-based SM retrievals from CyGNSS signals, e.g., other radiometer sensors (Al-Yaari et al., 2014). In **Figure 3(c)**, over croplands (yellow circle; IGBP code 12 in **Figure 3(b)**), CyGNSS showed the highest R_i-value among land classification types. This result suggests a potential application of CyGNSS observations in agricultural monitoring.

On the other hand, in **Figure 3(c)**, the lowest R_i-values were found over Evergreen Needleleaf Forest (IGBP code 1) and Barren or Sparsely Vegetated regions (IGBP code 16). The average value of VOD for Evergreen Needleleaf pixels is 5.66 kg/m^3 (densely vegetated), and the average value of VOD for Barren or Sparsely Vegetated pixels is 0.09 kg/m^3 (sparsely vegetated). There are several possible error sources for low R_i-value results in these densely and sparsely vegetated regions.

Firstly, the signal reflected from the dry surface originates from deeper layers of soil, causing significant problems in the SM retrieved from microwave-band instruments over arid regions (Holmes et al., 2006). Under extremely dry conditions, a low-frequency microwave signal penetrates more deeply into the soil layer. Decreasing the SM content leads to an exponential-like increase in penetration depth (Ulaby et al., 2014). For example, a GPS signal

with a wavelength of 0.19 m can penetrate more than 1 m of dry soil. This deeply penetrating signal causes volume scattering by subsurface discontinuities such as rock surface or frozen soil beneath shallow soil (Wagner et al., 2013); thus, significant problems can arise when SM is retrieved from passive and active microwave-band instruments (Kim et al., 2018). Secondly, these dry regions have relatively low SM variation; thus, the signal from the surface could be degraded by the background noise of the instrument. Thirdly, the low R_i-value of densely vegetated regions might be associated with the signal attenuation and scattering caused by dense vegetation canopies. To compensate for these errors, since active sensor-based SM retrieval is known to produce better SM information than passive sensor-based SM retrieval over densely vegetated regions, active sensor-based SM estimates could be employed when CyGNSS-derived rSNR is combined with other satellite-based SM products (Kim et al., 2018).

5 Conclusion

This study shows the potential of CyGNSS observations to fill the current spatial and temporal gap in SM estimations provided by existing satellite-based SM retrieval systems. We found a promising application of CyGNSS-derived rSNR in agricultural monitoring systems since rSNR reasonably describes SM variability over croplands. However, over sparsely and densely vegetated regions, the performance of SM estimation from rSNR seems to be degraded by erroneous scattering and attenuation of GPS signals. To overcome these limitations, we combined rSNR value with SMAP data. In this way we were able to achieve daily estimations of SM with reasonable data quality.

These study results were estimated based on one complete year of CyGNSS observations. However, this study length might be not enough to gain the minimum or maximum values of nSNR required for rSNR calculation. Future studies should consider both

growing and non-growing seasons for the sensitivity analysis to precisely investigate the impact of vegetation effects on rSNR values.

Acknowledgements

We gratefully acknowledge funding from the NASA Terrestrial Hydrology Program (Program Manager Dr. Jared Entin, Grant # NNX12AP75G). The authors also thank the teams from NASA and USGS, ISMN, and the CyGNSS teams for making their datasets publicly available. The SMAP, NLDAS, GPM, and IGBP land types classification datasets can be downloaded directly from https://search.earthdata.nasa.gov. CyGNSS data products are available via the following link: http://clasp-research.engin.umich.edu/missions/cygnss/data-products.php.

References

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J. C., Fritz, N., Froissard, F., et al (2008). From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations. *Hydrology and Earth System Sciences Discussions*, *12*, 1323-1337. http://dx.doi.org/10.5194/hessd-5-1603-2008

Alonso-Arroyo, A., Camps, A., Monerris, A., Rüdiger, C., Walker, J. P., Onrubia, R., et al. (2016). On the correlation between GNSS-R reflectivity and L-band microwave radiometry. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(12), 5862-5879. https://doi.org/10.1109/JSTARS.2016.2588281

Al-Yaari, A., Wigneron, J. P., Ducharne, A., Kerr, Y. H., Wagner, W., De Lannoy, G., Reichle, R., et al. (2014). Global-scale comparison of passive (SMOS) and active (ASCAT)

satellite based microwave soil moisture retrievals with soil moisture simulations (MERRA-Land). *Remote Sensing of Environment*, 152, 614-626. https://doi.org/10.1016/j.rse.2014.07.013

Brocca, L., Ciabatta, L., Massari, C., Camici, S., & Tarpanelli, A. (2017). Soil Moisture for Hydrological Applications: Open Questions and New Opportunities. *Water*, 9(2), 140. https://doi.org/10.3390/w9020140

Carreno-Luengo, H., Camps, A., Querol, J., & Forte, G. (2016). First results of a GNSS-R experiment from a stratospheric balloon over boreal forests. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5), 2652-2663. https://doi.org/10.3390/rs71013120

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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(10), 4730-4742. https://doi.org/10.1109/JSTARS.2016.2588467

Chan, S. K., Bindlish, R., O'Neill, P., Jackson, T., Njoku, E., Dunbar, S., et al. (2018). Development and assessment of the SMAP enhanced passive soil moisture product. *Remote Sensing of Environment, 204*, 931-941. https://doi.org/10.1016/j.rse.2017.08.025

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. *Geophysical Research Letters*, 43(7), 3317-3324. https://doi.org/10.1002/2016GL068189

Clarizia, M. P., Ruf, C., Cipollini, P., & Zuffada, C. (2016). First spaceborne observation of sea surface height using GPS-Reflectometry. *Geophysical Research Letters*, 43(2), 767-774. https://doi.org/10.1002/2015GL066624

Clarizia, M. P., & Ruf, C. S. (2016). Wind speed retrieval algorithm for the Cyclone Global Navigation Satellite System (CYGNSS) mission. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8), 4419-4432. https://doi.org/10.1109/TGRS.2016.2541343

Crow, W. T., & Wood, E. F. (2002). The value of coarse-scale SM observations for regional surface energy balance moeling. *Journal of Hydrometeorology*, 3(4), 467-482. https://doi.org/10.1175/1525-7541(2002)003<0467:TVOCSS>2.0.CO;2

Davies, H. (1954). The reflection of electromagnetic waves from a rough surface. *Proceedings of the IEE-Part IV: Institution Monographs*, *101*(7), 209-214. https://doi.org/10.1049/pi-4.1954.0025

De Roo, R. D., & Ulaby, F. T. (1994). Bistatic specular scattering from rough dielectric surfaces. *IEEE Transactions on Antennas and Propagation*, *42*(2), 220-231. https://doi.org/10.1109/8.277216

Dobson, M. C., & Ulaby, F. T. (1986). Active microwave soil moisture research. *IEEE Transactions on Geoscience and Remote Sensing*, (1), 23-36. https://doi.org/10.1109/TGRS.1986.289585

Dorigo, W., Oevelen, P., Wagner, W., Drusch, M., Mecklenburg, S., Robock, A., & Jackson, T. (2011). A new international network for in situ soil moisture data. *Eos, Transactions American Geophysical Union*, *92*(17), 141-142. https://doi.org/10.1029/2011EO170001

Dorigo, W. A., Gruber, A., De Jeu, R. A. M., Wagner, W., Stacke, T., Loew, A., et al. (2015). Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sensing of Environment*, *162*, 380-395. https://doi.org/10.1016/j.rse.2014.07.023

Draper, C. S., Walker, J. P., Steinle, P. J., De Jeu, R. A., & Holmes, T. R. (2009). An evaluation of AMSR–E derived soil moisture over Australia. *Remote Sensing of Environment*, 113(4), 703-710. https://doi.org/10.1016/j.rse.2008.11.011

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., et al. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, 98(5), 704-716. https://doi.org/10.1109/jproc.2010.2043918

Friedl, M. A., McIver, D. K., Hodges, J. C., Zhang, X. Y., Muchoney, D., Strahler, A. H., et al. (2002). Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, *83*(1-2), 287-302. https://doi.org/10.1016/S0034-4257(02)00078-0

Gleason, S., Hodgart, S., Sun, Y., Gommenginger, C., Mackin, S., Adjrad, M., & Unwin, M. (2005). Detection and processing of bistatically reflected GPS signals from low earth orbit for the purpose of ocean remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6), 1229-1241. https://doi.org/10.1109/TGRS.2005.845643

Gruber, A., Su, C. H., Zwieback, S., Crow, W., Dorigo, W., & Wagner, W. (2016). Recent advances in (soil moisture) triple collocation analysis. *International Journal of Applied Earth Observation and Geoinformation*, *45*, 200-211. https://doi.org/10.1016/j.jag.2015.09.002

Holmes, T. R. H., De Rosnay, P., De Jeu, R., Wigneron, R. P., Kerr, Y., Calvet, J. C., et al. (2006). A new parameterization of the effective temperature for L band radiometry. *Geophysical Research Letters*, 33(7). https://doi.org/10.1029/2006GL025724

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., & Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling

using data assimilation. *Water Resources Research*, 34(12), 3405-3420. https://doi.org/10.1029/1998WR900001

Karthikeyan, L., Pan, M., Wanders, N., Kumar, D. N., & Wood, E. F. (2017). Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms. *Advances in Water Resources*, 109, 106-120. https://doi.org/10.1016/j.advwatres.2017.09.006

Kerr, Y. H., Waldteufel, P., Wigneron, J. P., Delwart, S., Cabot, F., Boutin, J., et al. (2010). The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proceedings of the IEEE*, 98(5), 666-687. https://doi.org/10.1109/jproc.2010.2043032

Kim, H., Parinussa, R., Konings, A. G., Wagner, W., Cosh, M. H., Lakshmi, V., et al. (2018). Global-scale assessment and combination of SMAP with ASCAT (active) and AMSR2 (passive) soil moisture products. *Remote Sensing of Environment*, 204, 260-275. https://doi.org/10.1016/j.rse.2017.10.026

Kim, H., Zohaib, M., Cho, E., Kerr, Y. H., & Choi, M. (2017). Development and Assessment of the Sand Dust Prediction Model by Utilizing Microwave-Based Satellite Soil Moisture and Reanalysis Datasets in East Asian Desert Areas. *Advances in Meteorology*, 2017, 13. https://doi.org/10.1155/2017/1917372

Masters, D. S. (2004). Surface remote sensing applications of GNSS bistatic radar: Soil moisture and aircraft altimetry (Doctoral dissertation). Retrieved from Citeseerx. (http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.652.9553). Boulder, CO: University of Colorado.

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., & Stoffelen, A. (2014). Extended triple collocation: Estimating errors and correlation coefficients with

respect to an unknown target. *Geophysical Research Letters*, *41*(17), 6229-6236. https://doi.org/10.1002/2014GL061322

Ruf, C. S., Atlas, R., Chang, P. S., Clarizia, M. P., Garrison, J. L., Gleason, S., et al. (2016a). New ocean winds satellite mission to probe hurricanes and tropical convection. *Bulletin of the American Meteorological Society*, 97(3), 385-395. https://doi.org/10.1175/BAMS-D-14-00218.1

Ruf, C., Chang, P., Clarizia, M. P., Gleason, S., Jelenak, Z., Murray, J., et al. (2016b). CYGNSS Handbook. Ann Arbor, MI: Michigan Publishing.

Ruf, C. S., Chew, C., Lang, T., Morris, M. G., Nave, K., Ridley, A., & Balasubramaniam, R. (2018). A New Paradigm in Earth Environmental Monitoring with the CYGNSS Small Satellite Constellation. *Scientific reports*, *8*(1), 8782. https://doi.org/10.1038/s41598-018-27127-4

Ruf, C. S., Chew, C., Lang, T., Morris, M. G., Nave, K., Ridley, A., & Balasubramaniam, R. (2018). A New Paradigm in Earth Environmental Monitoring with the CYGNSS Small Satellite Constellation. Scientific reports, 8(1), 8782. https://doi.org/10.1038/s41598-018-27127-4

Saxton, K. E., & Rawls, W. J. (2006). Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil science society of America Journal*, 70(5), 1569-1578. Soil Sci. Soc. Am. J. 70 (5), 1569–1578. http://dx.doi.org/10.2136/sssaj2005.0117

Schmugge, T., O'Neill, P. E., & Wang, J. R. (1986). Passive microwave soil moisture research. *IEEE Transactions on Geoscience and Remote Sensing*, (1), 12-22. https://doi.org/10.1109/TGRS.1986.289584

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., et al. (2010). Investigating soil moisture–climate interactions in a changing climate: A

review. *Earth-Science Reviews*, 99(3-4), 125-161. https://doi.org/10.1016/j.earscirev.2010.02.004

Skofronick-Jackson, G., Petersen, W. A., Berg, W., Kidd, C., Stocker, E. F., Kirschbaum, D. B., et al. (2017). The global precipitation measurement (GPM) mission for science and society. *Bulletin of the American Meteorological Society*, 98(8), 1679-1695. https://doi.org/10.1175/BAMS-D-15-00306.1

Ulaby, F. T., Moore, R. K., & Fung, A. K. (1986). Microwave remote sensing: Active and passive. (Vol. 3) - From theory to applications. Massachusetts, Norwood, MA: Artech House.

Ulaby, F. T., Long, D. G., Blackwell, W. J., Elachi, C., Fung, A. K., Ruf, C., et al. (2014). Microwave radar and radiometric remote sensing (Vol. 4, No. 5). Ann Arbor, MI: University of Michigan Press.

Voronovich, A. G., & Zavorotny, V. U. (2017). Bistatic Radar Equation for Signals of Opportunity Revisited. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/TGRS.2017.2771253

Walker, J. P., Willgoose, G. R., & Kalma, J. D. (2001). One-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: A simplified soil moisture model and field application. *Journal of Hydrometeorology*, 2(4), 356-373. https://doi.org/10.1175/1525-7541(2001)002<0356:ODSMPR>2.0.CO;2

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., et al. (2013). The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift*, 22(1), 5-33. https://doi.org/10.1127/0941-2948/2013/0399

Wagner, W., Lemoine, G., & Rott, H. (1999a). A method for estimating soil moisture from ERS scatterometer and soil data. *Remote sensing of environment*, 70(2), 191-207. https://doi.org/10.1016/s0034-4257(99)00036-x

Wagner, W., Lemoine, G., Borgeaud, M., & Rott, H. (1999b). A study of vegetation cover effects on ERS scatterometer data. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2), 938-948. https://doi.org/ 10.1109/36.752212

Wagner, W., Noll, J., Borgeaud, M., & Rott, H. (1999c). Monitoring soil moisture over the Canadian Prairies with the ERS scatterometer. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1), 206-216. https://doi.org/ 10.1109/36.739155

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., ... & Livneh, B. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. Journal of Geophysical Research: Atmospheres, 117(D3). https://doi.org/10.1029/2011JD016048

Zavorotny, V. U., & Voronovich, A. G. (2000). Bistatic GPS signal reflections at various polarizations from rough land surface with moisture content. Paper presented at Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International. IEEE, Honolulu, Hawaii.

Zohaib, M., Kim, H., & Choi, M. (2017). Evaluating the Patterns of Spatiotemporal Trends of Root Zone Soil Moisture in Major Climate Regions in East Asia. *Journal of Geophysical Research: Atmospheres*, 122, 7705-7722. https://doi.org/10.1002/2016JD026379

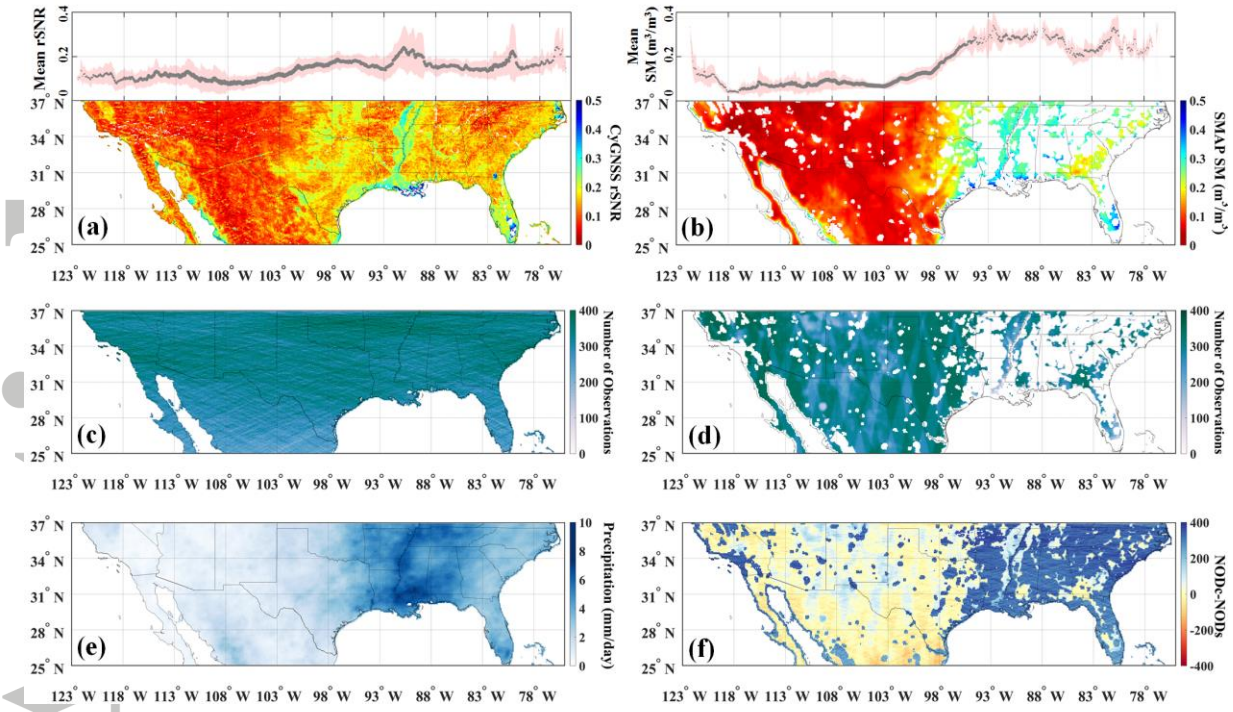

Figure 1. Maps of (a) The average value of CyGNSS-derived rSNR, (b) The average value of SMAP-based soil moisture, (c) The number of observations of CyGNSS-based rSNR, (d) The number of observations of SMAP (both ascending and descending), (f) The average value of GPM-based precipitation, and (f) The difference between (c) and (d) over CONUS from March 2017 to March 2018. The upper panels in (a) and (b) indicate longitude zonal means of each map's variable. Marker sizes in zonal plots illustrate the number of pixels used in zonal mean calculations, and shaded regions show ±1 standard deviation.

*NODc: Number of observation by CyGNSS

*NODs: Number of observation by SMAP

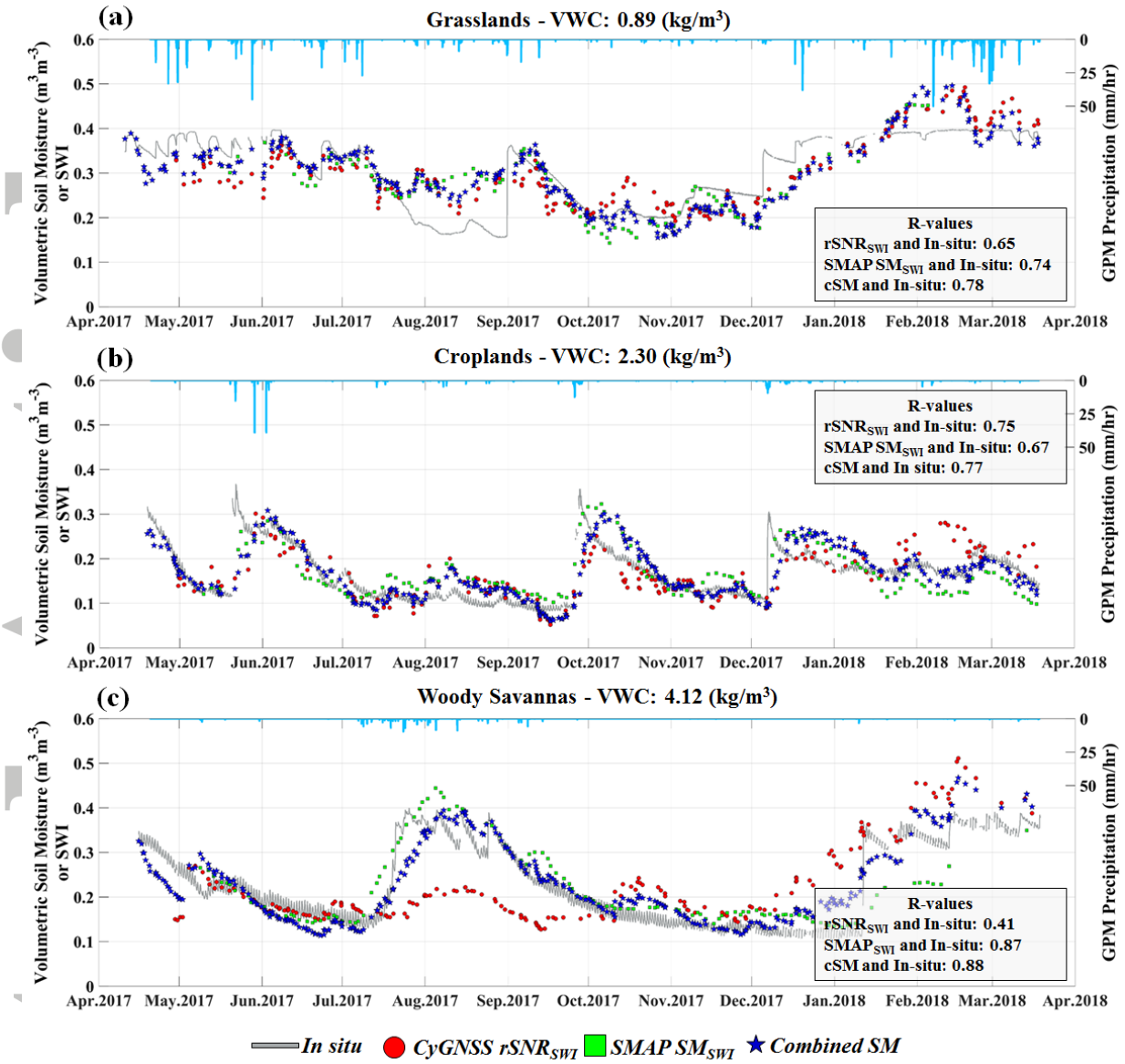

Figure 2. Time series of CyGNSS-derived $rSNR_{SWI}$ (red dots), SMAP SM_{SWI} (green squares), and combined SM (CyGNSS+SMAP combined) (blue stars), in-situ SM (grey lines), and GPM-derived precipitation (blue bars on right y-axis) with respect to different amount of vegetation water contents (VWC) and land cover types (a) 0.89 $kg \cdot m^{-3}$ (Grasslands) (b) 2.30 $kg \cdot m^{-3}$ (Croplands) and (c) 4.12 $kg \cdot m^{-3}$ (Woody Savannas), respectively.

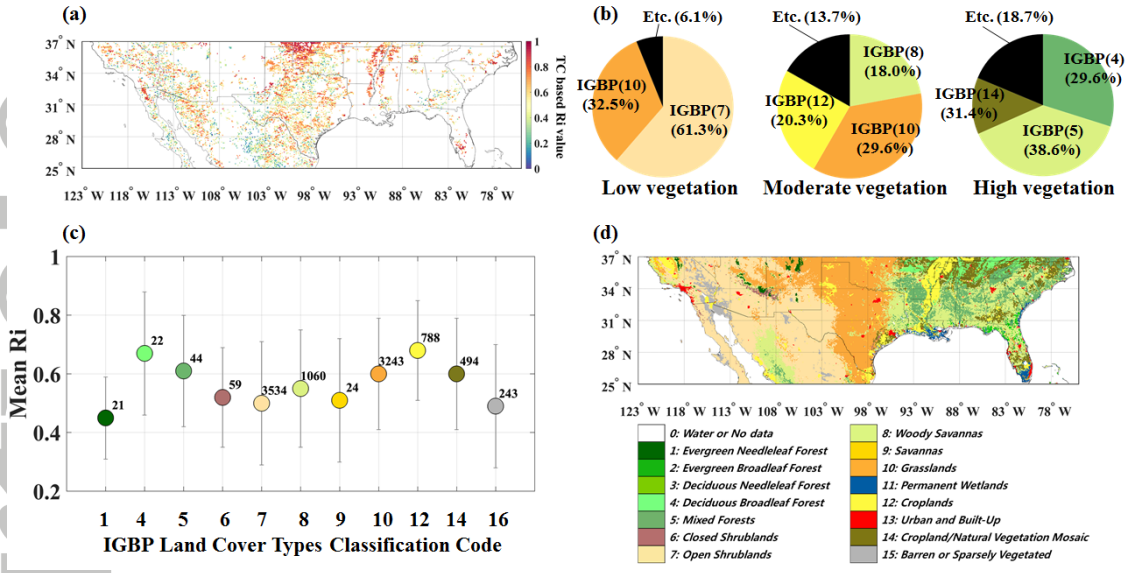

Figure 3. (a) Maps of the statistical result for the CyGNSS-derived rSNR for R_i estimates for the period March 2017 to March 2018. (b) pie charts show the land cover types classification from the IGBP based on VWC ranges. (c) mean R_i for CyGNSS-derived rSNR in terms of different land cover types. The gray line for each circle indicates standard deviation. The number of data used for each land cover types is indicated by a number next to each circle. (d) the IGBP land cover types classification map. 44,185 pixels are used for this analysis in total.