

UCAR/CU CYGNSS Soil Moisture Product

User Guide

Developed by:
Clara Chew

University Corporation for Atmospheric Research

&

Eric Small

University of Colorado at Boulder



This handbook describes the generation and use of version 1.0 of the UCAR/CU CYGNSS Soil Moisture Product. Future versions of this handbook will be released as the product is improved. This product was a joint effort by the University Corporation for Atmospheric Research (UCAR) and the University of Colorado at Boulder.

If you plan to use the data or methodology as part of a presentation, publication, or other effort, please cite the following paper:

Chew, C. C., & Small, E. E. (2018). Soil Moisture Sensing Using Spaceborne GNSS Reflections: Comparison of CYGNSS Reflectivity to SMAP Soil Moisture. *Geophysical Research Letters*, 45(9), 4049-4057.
<https://doi.org/10.1029/2018GL077905>

Last revised May 9, 2019.

Table of Contents

<i>CYGNSS Reference Documents</i>	4
<i>Introductory Comments</i>	6
1. GNSS-Reflectometry Background	7
1.1 Reflection Geometry	7
1.2 Revisit Time	8
1.3 Spatial Resolution	8
1.4 Delay-Doppler Maps	9
1.5 CYGNSS Observables	10
2. Soil Moisture Sensing Using GNSS-R	11
2.1 Previous Work	11
2.2 Remote Sensing at L-band	11
2.3 GNSS-R Sensitivity to Soil Moisture	12
3. The UCAR/CU Soil Moisture Retrieval Algorithm	13
3.1 Introduction to the Algorithm	13
3.2 Algorithm Description	14
3.2.1 Processing of delay-Doppler maps (DDMs)	14
3.2.2 Correction of P_r for other effects to derive $P_{r,eff}$	14
3.2.3 Land calibration	15
3.2.4 Incidence angle correction	18
3.2.5 Outlier identification	19
3.2.6 Removal of data affected by open water	20
3.2.7 Transforming $P_{r,eff}$ into soil moisture	21
3.2.8 Daily and sub-daily retrievals	24
3.2.9 Quality control	24
3.2.10 Soil moisture retrieval uncertainty	24
3.3 Thoughts on Gridding	25
4. In situ validation	29
5. Plans for future versions	36
6. File overview and loading the data	36
7. Quality flags – Important	37

CYGNSS Reference Documents

The papers listed here provide further information on CYGNSS (with a bias towards land applications), though it is not a comprehensive list.

- Al-Khaldi, M.M., Johnson, J.T., O'Brien, A.T., Balenzano, A., & Mattia, F. (2019). Time-Series Retrieval of Soil Moisture Using CYGNSS. *IEEE Transactions on Geoscience and Remote Sensing*. doi: 10.1109/TGRS.2018.2890646.
- Camps, A., Park, H., Pablos, M., Foti, G., & Gommenginger, C. P. (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.
- Carreno-Luengo, H., Luzi, G., & Crosetto, M. (2018). Impact of the elevation angle on CYGNSS GNSS-R bistatic reflectivity as a function of effective surface roughness over land surfaces. *Remote Sensing*. 10(11). 1749.
- Carreno-Luengo, H., Luzi, G., & Crosetto, M. (2019). Sensitivity of CYGNSS bistatic reflectivity and SMAP microwave radiometry brightness temperature to geophysical parameters over land surfaces. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 12(1). 107-122.
- Chew, C., Colliander, A., Shah, R., Zuffada, C., & Burgin, M. (2017). The sensitivity of ground-reflected GNSS signals to near-surface soil moisture, as recorded by spaceborne receivers. *International Geoscience and Remote Sensing Symposium (IGARSS) (Vol. 2017–July)*.
<https://doi.org/10.1109/IGARSS.2017.8127544>
- Chew, C. C., Small, E. E., & Podest, E. (2018). Monitoring land surface hydrology using CYGNSS. *IGARSS 2018. IEEE 2018 International Geoscience and Remote Sensing Symposium*.
- Chew, C. C., & Small, E. E. (2018). Soil Moisture Sensing Using Spaceborne GNSS Reflections: Comparison of CYGNSS Reflectivity to SMAP Soil Moisture. *Geophysical Research Letters*, 45(9), 4049–4057.
<https://doi.org/10.1029/2018GL077905>
- Chew, C., Lowe, S., Parazoo, N., Esterhuizen, S., Oveisgharan, S., Podest, E., ... Freedman, A. (2017). SMAP radar receiver measures land surface freeze/thaw state through capture of forward-scattered L-band signals. *Remote Sensing of Environment*, 198. <https://doi.org/10.1016/j.rse.2017.06.020>
- 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).
<https://doi.org/10.1002/2016GL068189>
- Chew, C., Reager, J. T., & Small, E. (2018). CYGNSS data map flood inundation during the 2017 Atlantic hurricane season. *Scientific Reports*, 8(1).
<https://doi.org/10.1038/s41598-018-27673-x>

- Clarizia, M.P., Pierdicca, N., Costantini, F., & Floury, N. (2019). Analysis of CYGNSS Data for Soil Moisture Retrieval. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. doi: 10.1109/JSTARS.2019.2895510
- Foti, G., Gommenginger, C., Jales, P., Unwin, M., Shaw, A., Robertson, C., & Roselló, J. (2015). Spaceborne GNSS reflectometry for ocean winds: First results from the UK TechDemoSat-1 mission. *Geophysical Research Letters*, 42(13), 5435–5441. <https://doi.org/10.1002/2015GL064204>
- Kim, H., and Lakshmi, V. (2018). Use of cyclone global navigation satellite system (CYGNSS) observations for estimation of soil moisture. *Geophysical Research Letters*. 45(16). 8272-8282.
- Murray, J. J., Ruf, C., Baker, N., Lang, T., & Uh. (1993). Report on the NASA CYGNSS Mission Applications Workshop, 14, 541–549.
- Nghiem, S. V., Zuffada, C., Shah, R., Chew, C., Lowe, S. T., Mannucci, A. J., ... Rosenqvist, A. (2017). Wetland monitoring with global navigation satellite system reflectometry. *Earth and Space Science*, 4(1). <https://doi.org/10.1002/2016EA000194>
- Rose, R., Ruf, C., Rose, D., Brummitt, M., & Ridley, A. (2013). The CYGNSS flight segment: A major NASA science mission enabled by micro-satellite technology. In *2013 IEEE Aerospace Conference* (pp. 1–13).
- Ruf, C., Chew, C. C., Lang, T., Morris, M., Nave, K., Ridley, A., & Balasubramaniam, R. (2018). A new paradigm in earth environmental monitoring with the CYGNSS small satellite constellation. *Scientific Reports*, 8(1).
- Ruf, C., Gleason, S., Jelenak, Z., Katzberg, S., Ridley, A., Rose, R., ... Zavorotny, V. (2012). The CYGNSS nanosatellite constellation hurricane mission. In *2012 IEEE International Geoscience and Remote Sensing Symposium* (pp. 214–216).

Introductory Comments

The UCAR/CU Cyclone Global Navigation Satellite System (CYGNSS) Soil Moisture Product is an L-band bistatic radar dataset that provides estimates of 0-5 cm soil moisture at a 6-hour discretization for the majority of the extratropics. CYGNSS is a constellation of eight small satellites that was designed to observe ocean surface wind speed during hurricanes (PI Chris Ruf, University of Michigan); it is a NASA Earth Ventures Mission that was launched in December of 2016. These satellites employ a relatively new remote sensing technique called GNSS-Reflectometry (GNSS-R), which records L-band signals transmitted by navigation satellites that have reflected off of the Earth's surface and back into space.

Traditional radar remote sensing requires a transmitter; by using existing signals from navigation satellites, GNSS-R satellites avoid this requirement. All that is needed is a receiver, which significantly reduces the cost of a satellite mission. Because of this, several receivers can be launched for a fraction of the cost of one traditional remote sensing satellite. The outcome is more data that is collected more frequently, albeit with tradeoffs that will be described in his handbook.

CYGNSS, in effect, is repurposing the existing GNSS signals—using them for ocean surface remote sensing instead of navigation. Here, we repurpose the CYGNSS data to estimate soil moisture over land. This product should be used with caution—there are many known issues with the current version of the data, and the data are not final. Users should keep the following in mind when exploring the data: The CYGNSS mission was not designed for soil moisture remote sensing. Data are calibrated and recorded assuming that the rough ocean surface is the target. Only two people have been wholly responsible for the data provided here: recalibration over land, algorithm development, validation, and code generation, with a small amount of money generously provided by UCAR. In addition, using GNSS-R for remote sensing of the land surface is such a new field that much of the theory behind the signal scattering over the land surface is still being understood. Our algorithm makes assumptions about the scattered signal that at best are simplifications and at worst are incorrect. Keeping this in mind, we hope that users will *not* see these soil moisture retrievals as the best that GNSS-R, or even CYGNSS, can provide, but we do hope they will serve as a launching point for learning about the true capabilities of this new field.

We would like to acknowledge Dr. Chris Ruf and the rest of the CYGNSS team for working hard to provide such high quality GNSS-R to the community. Without their efforts of course there would be no soil moisture product. We would also like to acknowledge Jan Weiss, Maggie Sleziak, and Michael Rousseau at UCAR for helping put the retrievals online.

1. GNSS-Reflectometry Background

1.1 Reflection Geometry

Global Navigation Satellite System-Reflectometry (GNSS-R) is a form of bistatic radar that utilizes transmitted navigation signals as the signal source. GNSS is an umbrella term that encompasses constellations like the United States' GPS, but also the EU's Galileo, Russia's GLONASS, China's BeiDou, India's IRNSS, and Japan's QZSS. In total, there are over 80 GNSS satellites currently in orbit (32 of which are GPS satellites), with more being planned in the coming years.

To date, GNSS-R most commonly utilizes signals transmitted from GPS satellites, which are circularly-polarized, L-band microwave signals. Unlike monostatic radar, which measures backscatter, GNSS-R measures the forward-scattered signal, which has reflected off of the surface of the Earth and back into space. Figure 1 presents a schematic of the signal geometry. A satellite in low Earth orbit, with a GNSS-R receiver onboard, has one or more downward-looking antennas, which record the forward-scattered signals.

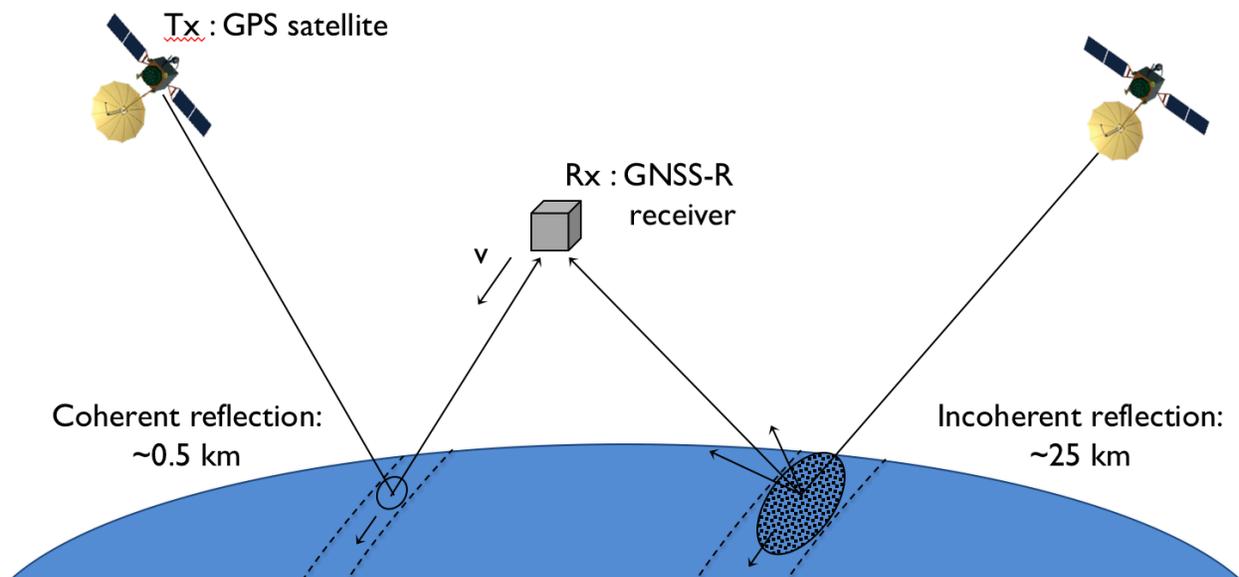


Figure 1. Schematic of the GNSS-R technique. A GNSS satellite transmits (Tx) a signal towards the Earth's surface. Part of this signal reflects in the forward (specular) direction and back into space. A GNSS-R receiver (Rx) onboard a low Earth orbiting satellite, with a downward looking antenna, records this signal. The point on the Earth's surface where the signal reflects depends upon the positions of the transmitting and receiving satellites. The roughness of the surface at the reflection point determines the spatial resolution of the signal, with rougher surfaces producing

larger spatial footprints. Nearly always, the receiver integrates the reflected signal over a period of time, which elongates the spatial footprint in the along-track direction.

1.2 Revisit Time

The point of reflection on the Earth's surface is determined by the positions of the transmitting and receiving satellites. Because these positions are constantly changing, the collecting areas are pseudo-randomly distributed on the Earth's surface. This is different than traditional remote sensing techniques, which collect data in repeatable swaths. The temporal repeat time of GNSS-R is thus statistical. This means that, for a given point of the Earth's surface, observations could be recorded one hour apart, and then there could be no observations for the next several hours, for example. Observations are recorded at all times of day, again, unlike traditional remote sensing techniques, which tend to design their collection strategies to always occur at a particular location at a particular time of day. The pseudo-random distribution of observations, over time, aggregate such that complete maps of the reflected signal can be made (Figure 2).

1.3 Spatial Resolution

The spatial resolution of the reflecting signal depends on the roughness of the surface at and near the reflection point. If the surface is relatively rough, then the reflected signal is incoherent and comes from an area called the 'glistening zone,' which is on the order of several kilometers (~25 km in the case of the ocean surface). If the surface is relatively smooth, then the reflected signal is coherent and comes from an area defined by the first Fresnel zone. For a low Earth orbiting GNSS-R satellite, this area is on the order of $\frac{1}{2}$ a kilometer, though this depends slightly (+/- a few hundred meters) on incidence angle. What can be defined as a rough versus smooth surface is still a subject of debate. Theoretically at L-band, once the surface roughness exceeds a few centimeters, then there should be little to no reflected signal coming from the first Fresnel zone. In practical terms, however, surface roughness is an extremely difficult parameter to measure, and surface roughness will vary considerably on scales as large as the first Fresnel zone. In all likelihood, the reflected signal for the coherent case comes from areas within the first Fresnel zone that are smooth, and most signals are probably a combination of incoherent and coherent scattering. Regardless, some studies are beginning to show that a large portion of spaceborne GNSS-R signals collected over the land surface have a coherent component, though these studies have yet to be published.

Due to the fact that CYGNSS was designed to be an ocean sensor, where the reflected signal is relatively weak, the processing software integrates the signal over a period of 1 second for each 'observation.' During that time, the spacecraft has moved approximately 7 km, which means that the smallest along-track

spatial resolution possible over land is 7 km, though the across track could still be the theoretical 0.5 km. This results in the spatial footprint having a minimum size of 7 x 0.5 km, with the signal being smeared out along track (Figure 2). Future GNSS-R missions could be designed such that the integration time is shortened, which would decrease the spatial footprint.

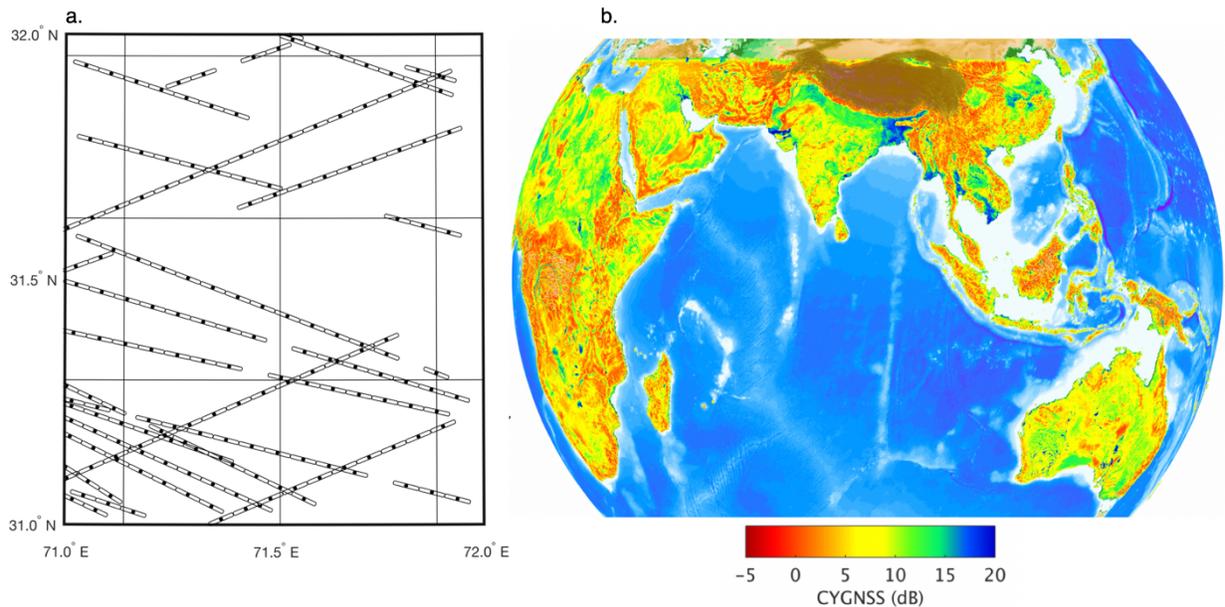


Figure 2: a. Illustration of the pseudo-random surface sampling by CYGNSS (ellipses with dots). Ellipses are approximately 7 x 0.5 km in size, which is the expected footprint if the surface has little surface or topographic roughness. Dots are the location of the specular reflection points recorded by CYGNSS. b. Over time, observations made by CYGNSS completely cover the land surface, producing maps such as this. Here, higher values could indicate a wet surface or a relatively flat surface.

1.4 Delay-Doppler Maps

The reflected GNSS signal is recorded by the receiver in the form of what is called a delay-Doppler map (DDM). A DDM is created by cross correlating the received signal with a locally-generated replica for different path delays (resulting from the path distance between the transmitter, reflecting surface, and receiver) and Doppler shifts (resulting from the relative motions of the transmitter, reflecting surface, and the receiver). Two examples of DDMs are shown in Figure 3. Figure 3a is an example of a DDM recorded by TDS-1 (a precursor to CYGNSS) over the land surface, and Figure 3b is an example of a DDM recorded over the ocean surface. The horseshoe shape of the ocean DDM is an indication that the reflection is incoherent and comes from a large, rough area. The lack of horseshoe in Figure 3a indicates that the reflection is mostly coherent, and comes from a smaller, smoother area. The maximum power of each DDM is affected by both surface roughness and the dielectric constant of the surface, which is explained further in Section 3. Some GNSS-R researchers try to quantify how much

'horseshoe' is present in the DDM as a proxy for the amount of coherence—we do not do this here.

1.5 CYGNSS Observables

DDMs are most commonly used by summarizing them into one metric or observable, though in very rare cases, the entire DDM or waveform (one slice of the DDM along constant doppler) may be used. The observables that are commonly used for soil moisture estimation are the peak cross-correlation of each DDM, or the peak divided by the noise floor (signal to noise ratio, SNR). The value of the peak cross-correlation of each DDM (called $P_{r,eff}$ in this document) is related to surface characteristics at the specular reflection point of the GNSS signal—including the roughness of the surface and the surface dielectric constant. Information about how DDMs are processed is contained in Section 3.2.1.

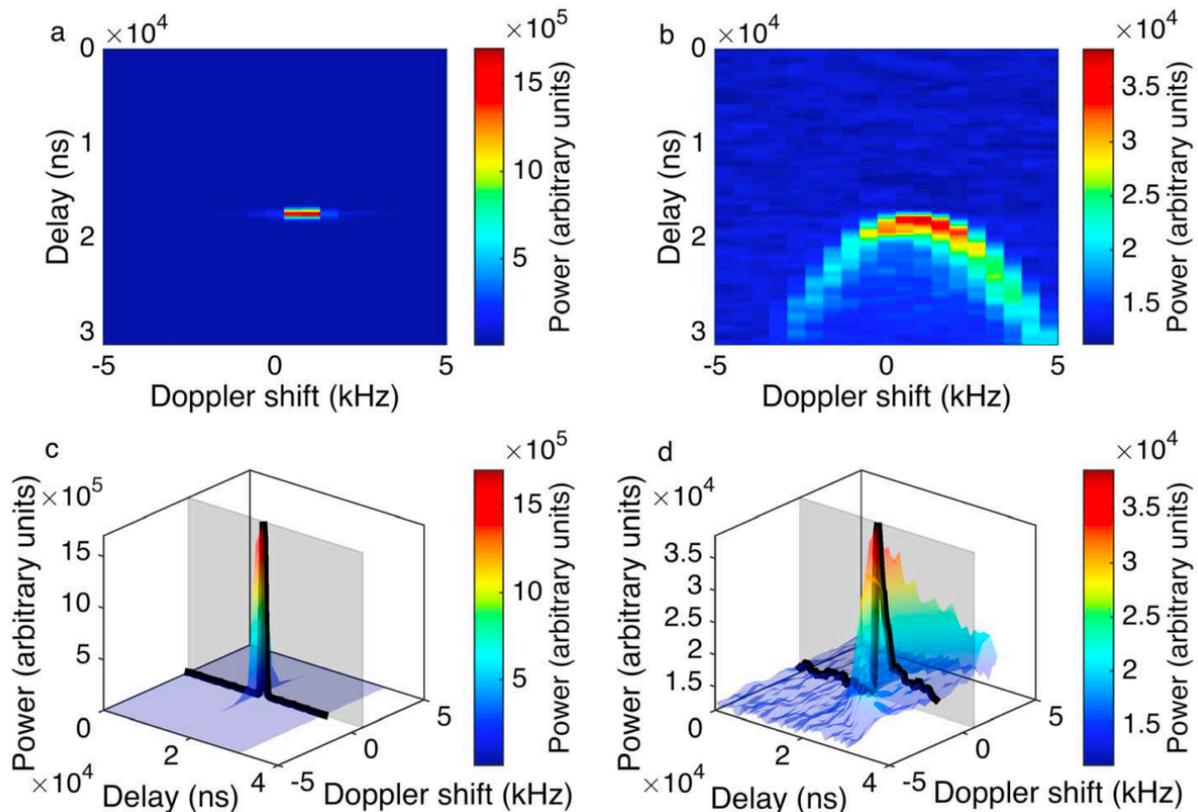


Figure 3. (a) DDM recorded over a rice field in the Ebro Delta, Spain. (b) DDM recorded over the Mediterranean Sea. (c and d) Same as Figures 1a and 1b except that also shown are the waveforms at a constant Doppler shift, indicated by the black traces. Note that the z axis scales in Figures 1b and 1d are an order of magnitude smaller than those in Figures 1a and 1c. Figure reproduced from Chew et al., 2016.

2. Soil moisture sensing using GNSS-R

2.1 Previous Work

Historically, the majority of spaceborne GNSS-R studies focused on signals reflecting from the ocean surface, either for the purpose of relating ocean surface roughness to wind speed, or for altimetric applications. Ground- and aircraft-based experiments had shown success in measuring GNSS-R signals over land and relating them to changes in near-surface (0-5 cm) soil moisture or vegetation water content, but it had generally been assumed that spaceborne GNSS-R signals recorded over the land surface would be too weak to be useful for these kinds of applications.

After the launch of TechDemoSat-1 (TDS-1) in 2014, observational evidence began to mount in favor of developing GNSS-R for land applications. Both Camps et al. (2016) and Chew et al. (2016) analyzed data from TDS-1 for sensitivity to soil moisture and found spatial and temporal variations in the GNSS reflected signal that appeared to be driven by soil moisture. Since then, both TDS-1 and the SMAP radar receiver (adapted to record GNSS-R signals) have shown sensitivity to a variety of land surface variables including wetland extent (Nghiem et al., 2017) and surface freeze/thaw (Chew et al., 2017). Both TDS-1 and SMAP, though garnering the largest spaceborne GNSS-R datasets of their time, do not collect enough data to provide operational products and are mostly limited to proof-of-concept investigations.

NASA's Cyclone GNSS (CYGNSS) constellation, launched in December of 2016, however, does provide enough data. Instead of being a single instrument, CYGNSS is comprised of eight GNSS-R satellites in low Earth orbit around the tropics. This vastly decreases the temporal repeat time. For instance, for the latitudinal band $\sim\pm 38$ degrees, CYGNSS samples approximately 80% of SMAP's 36 km EASE2 grid cells every day, and most of the time CYGNSS will have multiple observations for these grid cells.

2.2 Remote Sensing at L-band

Data collected by CYGNSS are sensitive to near-surface soil moisture for the same reason that all instruments that collect signals at L-band are sensitive to soil moisture. How strongly any signal reflects off of a surface is dependent on the dielectric constant of the surface. At L-band, the dielectric constant of the Earth is mostly controlled by its moisture content, with wetter surfaces producing stronger reflections. There is a secondary dependence on soil texture (i.e. the relative amounts of sand, silt, and clay that comprise a soil), though it is small compared to the effect from soil moisture.

L-band is often quoted as the wavelength of choice when it comes to soil moisture remote sensing. Higher frequencies like X- or C-band cannot penetrate even minimal vegetation canopies, whereas L-band can. L-band can penetrate the soil surface to some extent, and the amount of penetration also depends on soil moisture (Njoku and Entekhabi, 1996). In general, the effective penetration depth of an L-band signal, and thus of GNSS-R signals, is between 0-5 cm. Longer wavelength signals, like P-band, have been studied for their ability to sense rootzone soil moisture, though its penetration depth will also depend on moisture content, which leads to greater uncertainty in knowing at what depth the retrieved soil moisture is actually representing. Restrictions on the transmission of this wavelength have also limited its development.

2.3 GNSS-R Sensitivity to Soil Moisture

Since 2015, there have been several studies investigating the sensitivity of GNSS-R to soil moisture (Camps et al., 2016; C. Chew et al., 2016; C. Chew, Colliander, et al., 2017; C.C. Chew & Small, 2018; Clara C. Chew et al., 2018). Most of these studies have been conducted using empirical observations from CYGNSS or TechDemoSat-1. Observational evidence clearly shows that GNSS-R is very sensitive to surface water from lakes and rivers (Figure 4) even in the presence of an overlying vegetation canopy.

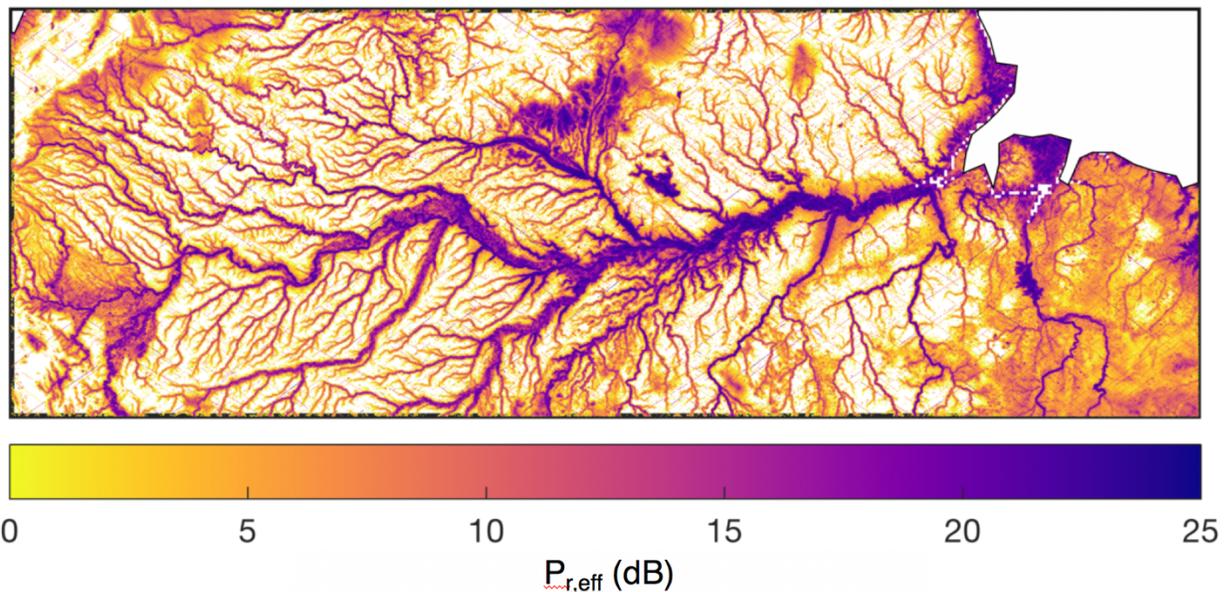


Figure 4. Observations of $P_{r,eff}$ over the Amazon basin.

Measuring the sensitivity of GNSS-R/CYGNSS observations to soil moisture, however, is more challenging. Spatial variations in both land cover and topography, which affect the roughness of the surface, will also affect $P_{r,eff}$. This is exemplified in Figure 5, which shows a satellite image of northern India along with

CYGNSS observations of $P_{r,eff}$. Although higher $P_{r,eff}$ is observed in vegetated areas, which should have higher soil moisture than the surrounding arid regions, one can also see the influence of mountain ranges and other surface features on $P_{r,eff}$.

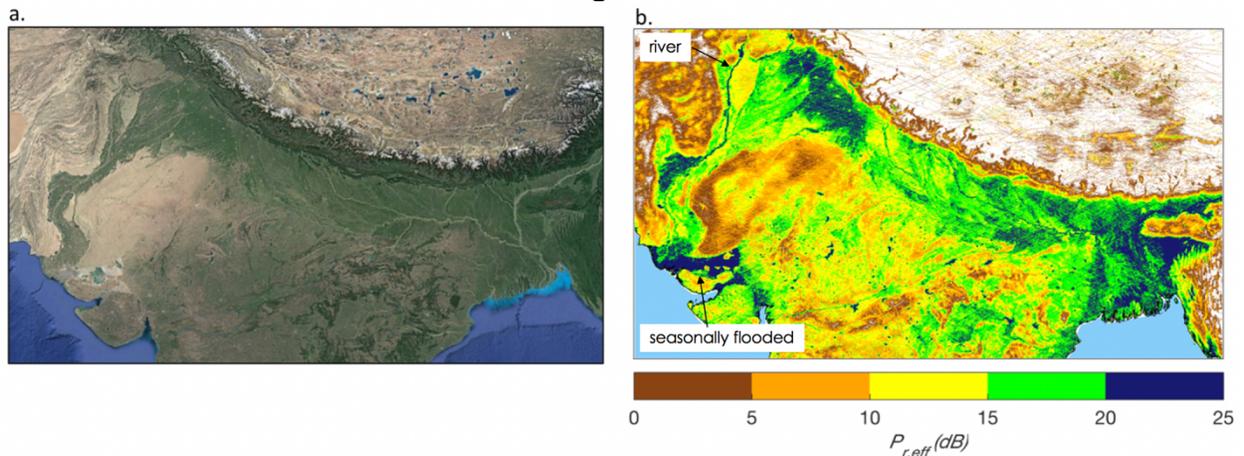


Figure 5. a. Google Earth image of northern India. b. CYGNSS observations of $P_{r,eff}$ over the same region. The color bar is continuous and is only chunked by 5 dB to highlight the response of $P_{r,eff}$ to different land cover types.

In order to untangle the response of $P_{r,eff}$ to both soil moisture and land cover/surface roughness, we assume that over time only soil moisture changes whereas land cover and surface roughness remain largely static. Of course, this approach ignores changes in vegetation water content. By looking at temporal fluctuations in both soil moisture and $P_{r,eff}$, we can quantify the sensitivity of $P_{r,eff}$ to soil moisture. Figure 6 shows an example of this kind of analysis in India, where changes in SMAP soil moisture are compared to gridded changes in $P_{r,eff}$. The correlation between the two is strong ($r = 0.84$).

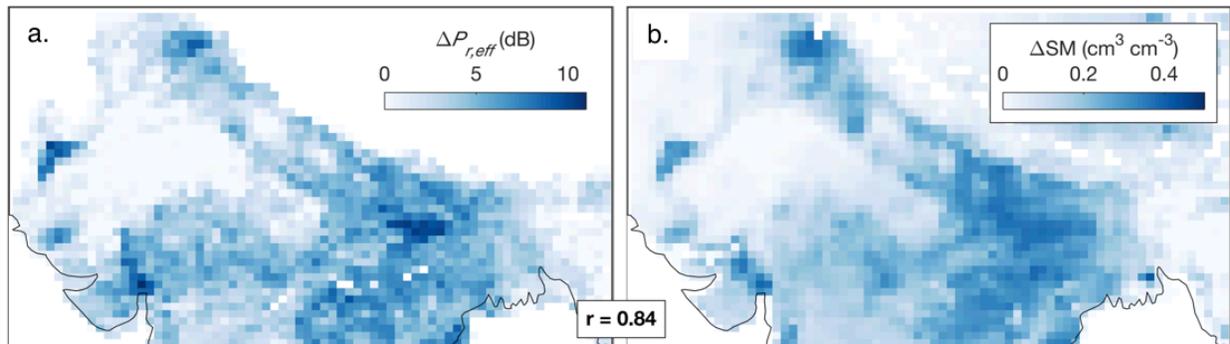


Figure 6. a. Changes in $P_{r,eff}$, gridded to 36 km, between May and August, 2017. b. Changes in soil moisture from SMAP between May and August, 2017. Adapted from Chew and Small, 2018.

3. The UCAR/CU Retrieval Algorithm

3.1 Introduction to the Algorithm

Our algorithm uses collocated soil moisture retrievals from the Soil Moisture Active Passive (SMAP) mission as 'ground truth' to calibrate concurrent (same day) CYGNSS observations. For a given location, a linear relationship between SMAP soil moisture and CYGNSS reflectivity is determined, and the relationship is used to transform all CYGNSS observations into soil moisture, even at times when there are no SMAP match ups.

Using SMAP data as 'ground truth' of course comes with many drawbacks, the major one being that SMAP soil moisture retrievals are not actual ground truth observations and have their own error and uncertainties. One must be careful when using CYGNSS data in areas where it is known that SMAP performs poorly. In addition, SMAP's 40 km spatial resolution is *likely* coarser than that of CYGNSS, though this is still up for debate. Intelligent upscaling of CYGNSS data to the 36 km EASE grid that SMAP uses is necessary. If the resolution of CYGNSS is smaller than 36 km, then we are in effect degrading the CYGNSS data by doing this and not using it to its full potential. However, in the absence of mature or validated GNSS-R scattering models, empirical algorithms must suffice, and SMAP data are considered to be the most accurate of the existing soil moisture products.

3.2 Algorithm Description

This section is a step-by-step guide to the soil moisture retrieval algorithm. It assumes a working knowledge of the CYGNSS Level 1, version 2.1, netcdf files, all of which are available here: <https://podaac.jpl.nasa.gov/CYGNSS>). In general, for every day of the year there will be eight Level 1 files, one for each CYGNSS satellite. Each of the eight files contains information pertaining to the thousands of reflections recorded on that day. The following steps are applied to each reflection in each file.

3.2.1 Processing of delay-Doppler maps (DDMs)

In previous works, we used the signal-to-noise ratio (SNR) as the signal of interest, which in the CYGNSS files is contained in the metadata as a variable called 'ddm_snr.' However, as time went on, we realized we were getting better results if we instead just pulled the peak value of the analog DDMs (variable name: power_analog) themselves and did not worry about the noise floor. We do not know why this is the case, possibly the noise floor itself is too noisy, and we do still utilize the SNR value itself for quality control. The peak value of the analog DDM is found and converted to dB, which we thereafter call P_r . The delay bin at which P_r occurs is also found during this step.

3.2.2 Correction of P_r for other effects to derive $P_{r,eff}$

Pr is not just affected by soil moisture or surface roughness; it is also affected by the gain of the receiving antenna, bistatic range, and the transmitted power of the GPS satellite. Pr is then corrected for antenna gain, range, and the GPS transmit power assuming a coherent reflection:

$$P_{rl}^r = \frac{P_r^t G^t}{4\pi(R_{ts}+R_{sr})^2} \frac{G^r \lambda^2}{4\pi} \Gamma_{rl} \quad (1)$$

Where: P_r^t is the transmitted RHCP power, G^t is the gain of the transmitting antenna, R_{ts} is the distance between the transmitter and the specular reflection point, R_{sr} is the distance between the specular reflection point and the receiver, G^r is the gain of the receiving antenna, λ is the GPS wavelength (0.19 m), and Γ_{rl} is the surface reflectivity. P_{rl}^r is the Pr as explained above. If you want to do this yourself, you'll need the following variables: 'sp_rx_gain' (G^r), 'rx_to_sp_range' (R_{sr}), 'tx_to_sp_range' (R_{ts}), and 'gps_eirp' ($P_r^t G^t$).

What we actually want to do is solve for Γ_{rl} , and we do this by converting all terms to dB (some of them are already in dB in the CYGNSS files). We tend to call Γ_{rl} (in dB) that has been corrected for all of these effects $P_{r,eff}$, which stands for effective reflectivity.

3.2.3 Land calibration

We make additional empirical calibrations for the GPS transmit power, which we have not described in previous papers. While these corrections are sub-optimal, it is much better than doing nothing. It is no secret that v2.1 GPS transmit powers are rough estimates, and we have found some biases in Pr depending on GPS PRN #. However, unlike other researchers, we have not found that removing the Block IIF satellites is necessary, so we keep them in to preserve more than a third of the total observations.

We currently calibrate the CYGNSS data in part of the Sahara Desert where Pr is relatively stable throughout the year, and soil moisture and vegetation changes have a negligible effect. In future versions, we will recalibrate over a longer time period and use data from dedicated CYGNSS cal/val sites, which at this time are still being determined. Figure 7 shows the part of the Sahara where the current calibration was completed (limits are indicated by the pins):



Figure 7. The region used for calibration is outlined by the yellow pins.

Figure 8 shows what $P_{r,eff}$ (labeled as SNR in the plot) looks like for the outlined region. Black dots are limits of what we will call sub-cells, which here we chose to be approximately 7 km x 7 km . In order to calculate PRN biases, we take the mean of P_r observations within each sub-cell and then calculate deviations from the mean (Figure 9). Because soil moisture, vegetation, and roughness, should be expected to minimally affect P_r for each sub-cell, we assume that deviations from the mean are the result of suboptimal PRN corrections (and incidence angle variations, described in the next section).

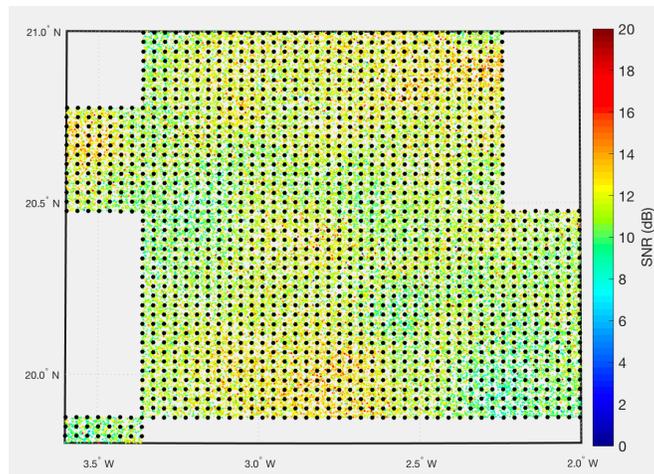


Figure 8. Observations of $P_{r,eff}$ (colored dots). Black dots outline 7 x 7 km sub-cells.

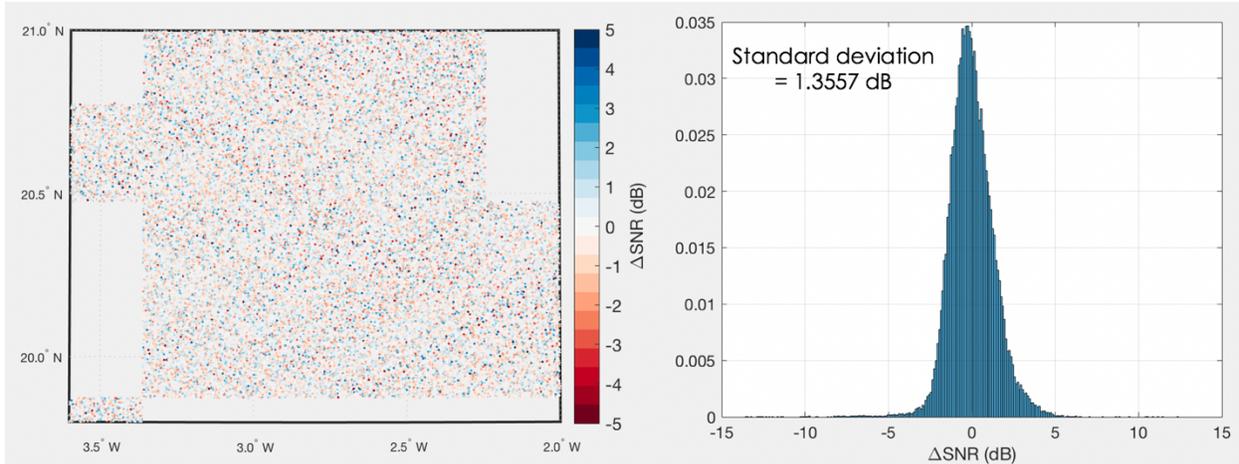


Figure 9. (left) Deviations of $P_{r,eff}$ from the mean for each sub-cell. (right) Histogram of deviations of $P_{r,eff}$. The standard deviation of the distribution is 1.3557 dB.

The goal of this exercise was to decrease the standard deviation of the distribution as much as possible (Figure 9). It turns out that if you bin this distribution by PRN, you see some consistent biases (Figure 10).

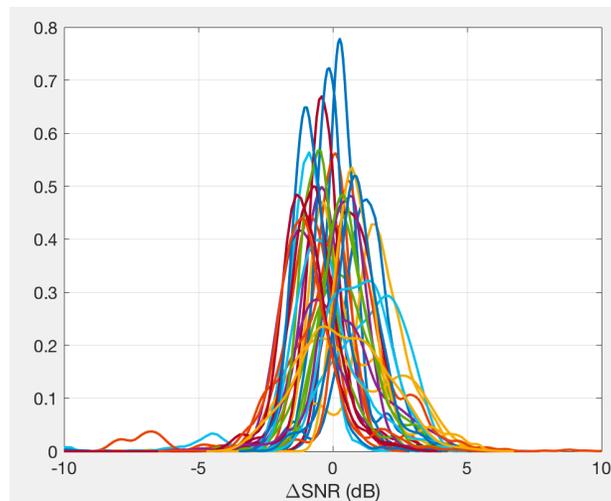


Figure 10. Distributions of $P_{r,eff}$ (labeled as SNR) as a function of PRN (unlabeled colored lines).

These biases are removed from $P_{r,eff}$ as the empirical calibration. It's likely that these biases actually change over time, and in future versions we will update these calibrations. Table 1 shows the biases themselves.

Table 1. Empirical biases in $P_{r,eff}$ found according to PRN.

PRN	Bias (dB)	PRN	Bias (dB)	PRN	Bias (dB)
1	1.017	11	-0.230	21	-0.909
2	0.004	12	-1.021	22	-0.838
3	1.636	13	0.007	23	-0.858
4	NaN	14	-0.730	24	1.140
5	-0.610	15	-0.376	25	0.880
6	0.241	16	-0.481	26	0.163
7	-0.709	17	0.256	27	0.409
8	0.605	18	-0.474	28	-0.712
9	1.498	19	-0.206	29	-1.032
10	-0.783	20	0.345	30	0.877
				31	-0.562
				32	-0.819

3.2.4 Incidence angle correction

Incidence angle is also expected to affect a coherent reflection, though angle only significantly affects the $P_{r,eff}$ when the angle is above 40 or 50 degrees. We modeled how $P_{r,eff}$ should be affected by incidence angle, for several different soil moisture values (Figure 11). If you normalize everything to zero degrees incidence, then you find that soil moisture only slightly changes the relationship between $P_{r,eff}$ and incidence angle. (This normalization is also done in Al-Khaldi et al., 2019.) We compared the mean, modeled relationship to observations of $P_{r,eff}$ and confirmed the overall drop in $P_{r,eff}$ as incidence angle increases beyond 40 degrees. We use the mean, modeled relationship to correct variations in $P_{r,eff}$ due to incidence angle.

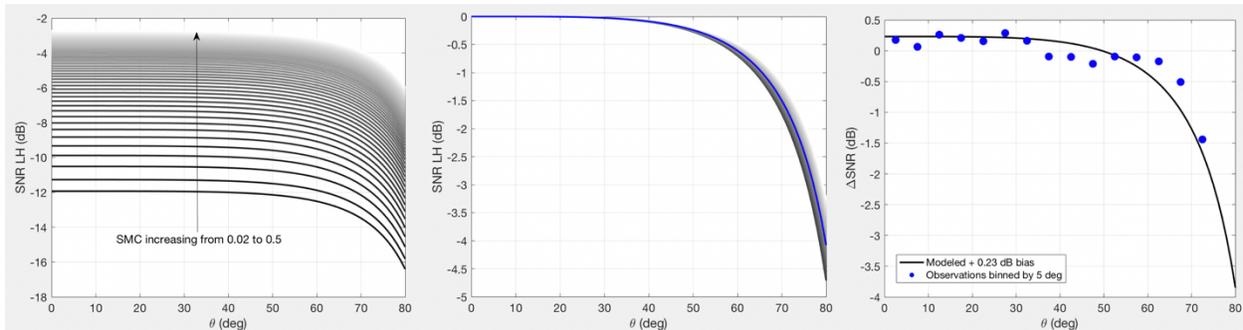


Figure 11. (left) Modeled relationship for how $P_{r,eff}$ should vary depending on incidence angle and soil moisture. (middle) Same as the left-hand panel, though here modeled $P_{r,eff}$ has been normalized to show that soil moisture does not significantly change the relationship between $P_{r,eff}$ and incidence angle. The blue line is the mean of the normalized relationships. (right) We binned observations of $P_{r,eff}$ over the Sahara in 5 degree increments to confirm that the modeled relationship at least loosely resembles what is seen in the observations.

After calibrating $P_{r,eff}$ for PRN biases and incidence angle, we see a significant decrease in the standard deviation of changes in $P_{r,eff}$ over the Sahara (Figure

12). The standard deviation decreased from 1.3 to 1 dB, which significantly improved soil moisture retrievals.

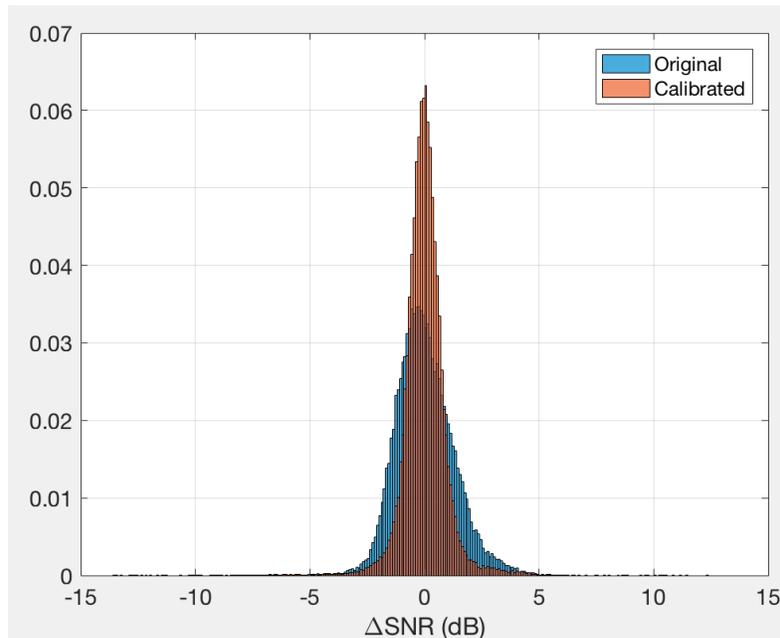


Figure 12. Distribution of changes of $P_{r,eff}$ over the Sahara before (blue) and after (orange) calibration and adjustment for incidence angle.

3.2.5 Outlier identification

Standard quality flags are used in the CYGNSS metadata to remove some outliers—the specific flags we use are 2, 4, 5, 8, 16, and 17, which in order are: S-band transmitter powered up, spacecraft attitude error, black body DDM, DDM is a test pattern, direct signal in DDM, and low confidence in the GPS EIRP estimate.

We perform additional quality control and remove the following: any observations with a (pre-corrected) SNR value less than 2 dB, observations with a receiver antenna gain less than 0, observations with an incidence angle greater than 65 degrees, and any data with a P_r coming in at a delay bin outside of 7-10 pixels (exclusive). In addition, we have found that results are improved if we impose a requirement that (pre-corrected) SNR must be less than or equal to the receiver antenna gain + 14. Lastly, we remove observations if the receiver gain is greater than 13 but still has a corrected $P_{r,eff}$ value less than 0. These are empirical corrections that are not standardized among other researchers using CYGNSS data.

Lastly, any data before ~December 2017 reflecting from a surface elevation above 600 m are removed. Because CYGNSS was optimized for ocean surface

sensing, the satellites did not record DDMs that contained the full surface reflection coming from above about 600 m altitude because they were not looking for data from these heights. The CYGNSS team changed the software after December, 2017, to include these data.

3.2.6 Removal of data affected by open water

The removal of specular reflection points that are affected by open water is a critical step before retrieving soil moisture. Even small water bodies ~25 m wide can significantly affect $P_{r,eff}$, which then means that these $P_{r,eff}$ observations will not be affected by soil moisture as strongly (Figure 13). We have probably tried a dozen different ways to mask open water, and none of them are perfect because no currently available water mask is perfect. Thus far, we have found the best success using the Pekel et al. (2016) dataset (<https://global-surface-water.appspot.com>), which is a 30 m, optically-derived water mask. Because it is derived from optical data, it underestimates the amount of water beneath vegetation.

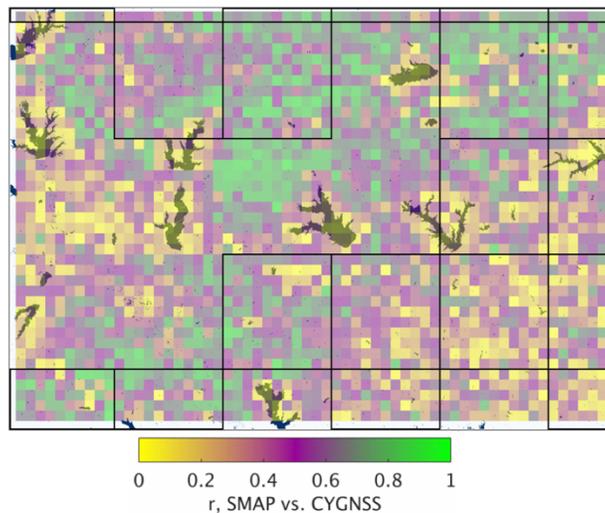


Figure 13. The correlation between CYGNSS observations and SMAP soil moisture for part of the United States (colored pixels). The Pekel et al. (2016) dataset for this region is also shown—the area is dominated by large reservoirs and smaller lakes. The correlation between SMAP and CYGNSS is low when observations fall on or near water bodies, which is expected. For reference, large black outlined boxes are the 36 km EASE-2 grid used by SMAP which normally contain high quality SMAP soil moisture observations.

The current algorithm removes open water using the ‘seasonality’ data product provided by Pekel. This product represents how many months out of a year a pixel is inundated (0-12). For our purposes, we make this product binary by considering any value greater than 1 to be flagged as water, and anything below this to be non-water. We do this because sometimes the permanent water bodies are seasonally covered by vegetation, which makes the Pekel dataset represent them as less than 12 (permanent).

For each specular reflection point, we find the amount of water within a 7 x 7 km region surrounding the point. This is a simplification of the actual footprint, but it is computationally more efficient than rotating axes to form actual ellipses, which themselves are simplifications and not well quantified. If the amount of water in the 7 x 7 km region exceeds 1%, we remove that CYGNSS observation from consideration. Changing these thresholds or region sizes changes the results, though never uniformly increasing or decreasing error across regions.

3.2.7 Transforming $P_{r,eff}$ into soil moisture

Here, we will describe how $P_{r,eff}$ is transformed into soil moisture, using SMAP soil moisture retrievals to calibrate CYGNSS observations. Our calibration period was chosen to be March 17, 2017 – October 1, 2018.

Our algorithm is very simple: it assumes that $P_{r,eff}$ is linearly-related to SMAP soil moisture. This relationship is expected to vary spatially, though in its current form we assume that it does not change over time (future versions will allow for these changes). For a given location, we calculated the slope of the best-fit linear regression between SMAP soil moisture and CYGNSS $P_{r,eff}$, after having removed the mean of each for the entire time series. Before we can describe this in more detail, however, we have to understand what 'a given location' means in this context.

We already described that we assume that $P_{r,eff}$ has a finer spatial resolution than SMAP's 40 km resolution. We have found that we get best results when we grid our $P_{r,eff}$ observations to ~3 x 3 km 'subcells' and then aggregate the gridded observations to the 36 km SMAP EASE-2 grid resolution (Figure 14). Why subcells? If we were to aggregate all observations of $P_{r,eff}$ in one 36 km grid cell and look at how $P_{r,eff}$ varied within that grid cell, we would see variations in $P_{r,eff}$ due to factors like land cover type and topography. By dividing up the grid cell into smaller subcells, we see more consistent relationships between $P_{r,eff}$ and soil moisture. The subcells effectively help to remove the confounding effects of land cover and topography on $P_{r,eff}$. The number of points per subcell in the calibration period are shown in Figure 15—subcells with less than 3 observations were not used for calibration.

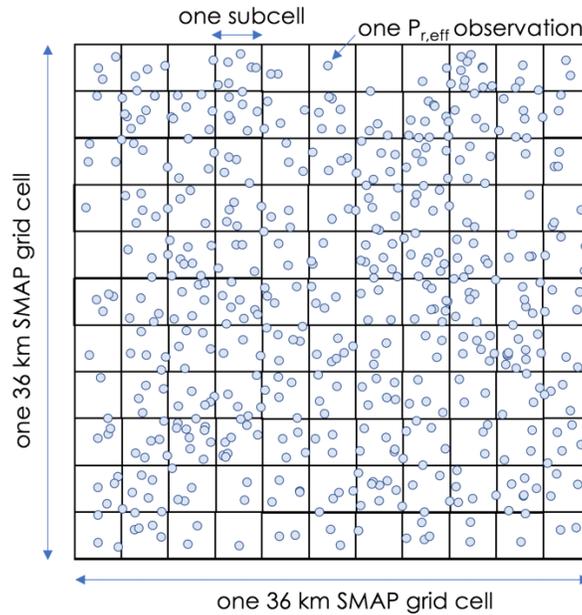


Figure 14. Depiction of how observations of $P_{r,eff}$ are gridded into subcells within a 36 km SMAP cell.

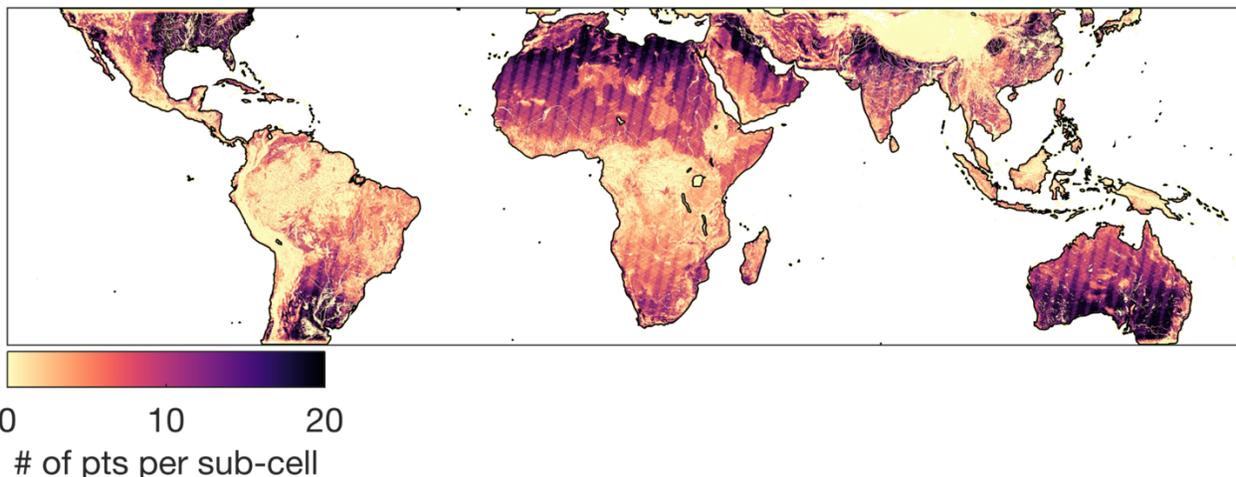


Figure 15. The number of CYGNSS observations for each sub-cell that were used for calibration. Fewer observations are found in higher elevation areas, which only have 'good' data for about $\frac{1}{2}$ the time series, relative to the lower elevation areas. Observations over open water have already been removed.

Within each subcell, we calculated the linear regression between SMAP soil moisture and $P_{r,eff}$ match-ups (occurring on the same day), after having removed the mean from both SMAP and $P_{r,eff}$ in that cell (correlation coefficients for this relationship are shown in Figure 16). The mean values of both SMAP and CYGNSS during the calibration period serve as our reference values, in order to return an absolute value of soil moisture from CYGNSS. In our algorithm, the reference value is the mean soil moisture for the entire calibration period. We call the slope of the best fit line β , which is conceptualized in Figure 17.

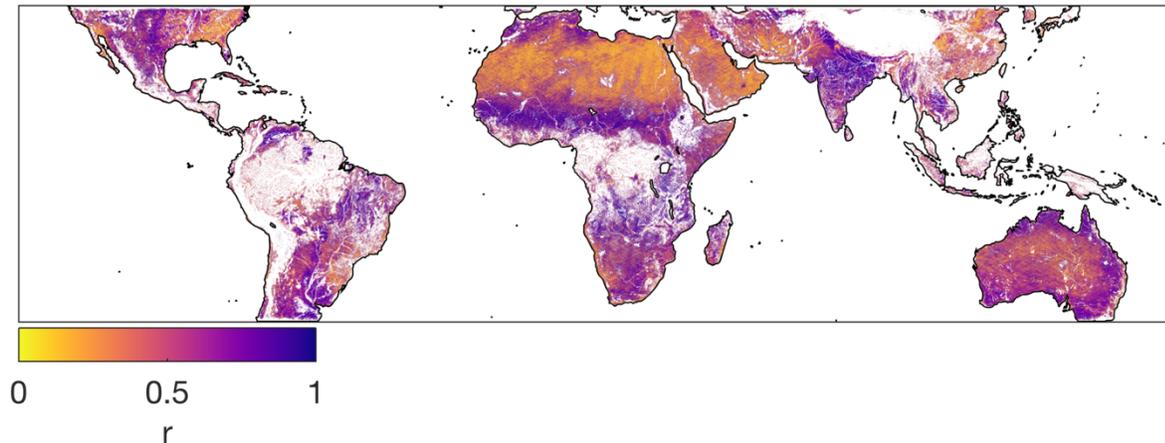


Figure 16. The correlation coefficient between SMAP soil moisture and CYGNSS reflectivity observations. Open water points have been removed. It is 'easier' to get a higher correlation coefficient when there is significant soil moisture variability throughout the year.

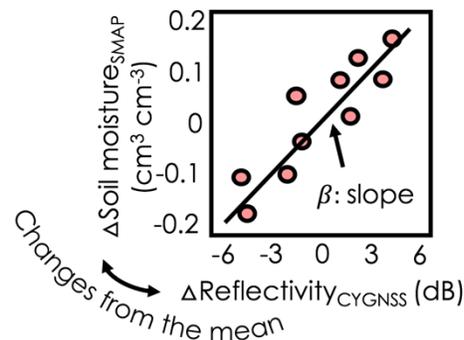


Figure 17. The slope of the best-fit line between SMAP soil moisture and $P_{r,eff}$ (labeled as Reflectivity) match ups is called β and is used to calculate soil moisture from CYGNSS.

β is used to estimate soil moisture from CYGNSS for data falling outside the calibration period as well as data within the calibration period when there are no SMAP match-ups (since SMAP has a 2-3 day overpass period):

$$Soil\ moisture_{CYGNSS} = \beta \times \Delta P_{r,eff} + \overline{Soil\ moisture_{SMAP}}$$

β varies spatially (Figure 18). Unfortunately, sometimes it looks like β is influenced by noise in regions where soil moisture shows little or no variability throughout the year. We are looking into parameter regionalization to minimize these effects.

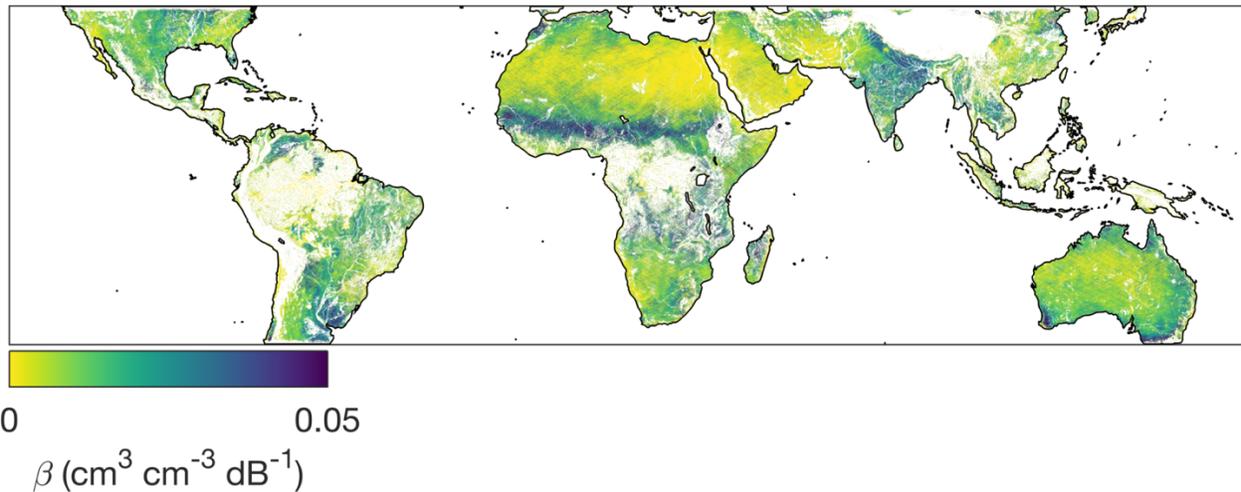


Figure 18. The slope of the linear regression between CYGNSS reflectivity observations and SMAP soil moisture (β). This represents the sensitivity of CYGNSS to soil moisture, with lower values indicating a higher sensitivity—though low values are also found in regions where soil moisture does not tend to vary. Higher values of β mean that CYGNSS is not as affected by increases or decreases in soil moisture. Be careful in interpreting this, as imperfect open water masking will cause an apparent insensitivity to soil moisture.

We then combine the subcell soil moisture retrievals by taking the average for a selected time period (either every 6 hours or every day) to upscale them to the EASE-2 36 km resolution. We are currently investigating whether or not the ~3 km retrievals are valid on their own—if so, we will release them in a future version.

3.2.8 Daily and sub-daily retrievals

We currently provide soil moisture retrievals on daily and sub-daily (6 hourly) time steps. For the daily retrievals, we average all observations within a particular grid cell that fall within the 24-hour time period. For the sub-daily retrievals, we average all observations for a particular grid cell in 6-hour intervals, which are currently midnight – 6 am, 6 am – noon, noon – 6 pm, and 6 pm – midnight (UTC).

3.2.9 Quality control

Currently, quality control is minimal—we remove soil moisture retrievals that indicate soil moisture being less than 0.01 or greater than 0.65 cm³ cm⁻³.

3.2.10 Soil moisture retrieval uncertainty

Figure 19 shows the unbiased root mean square difference (ubRMSD) between CYGNSS and SMAP soil moisture retrievals for the calibration period (March 18, 2017 – October 1, 2018). Semi-transparent regions are those frequently flagged by SMAP as being poor quality. Note that we tend to get higher ubRMSD in areas that flood seasonally.

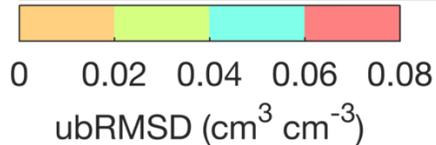
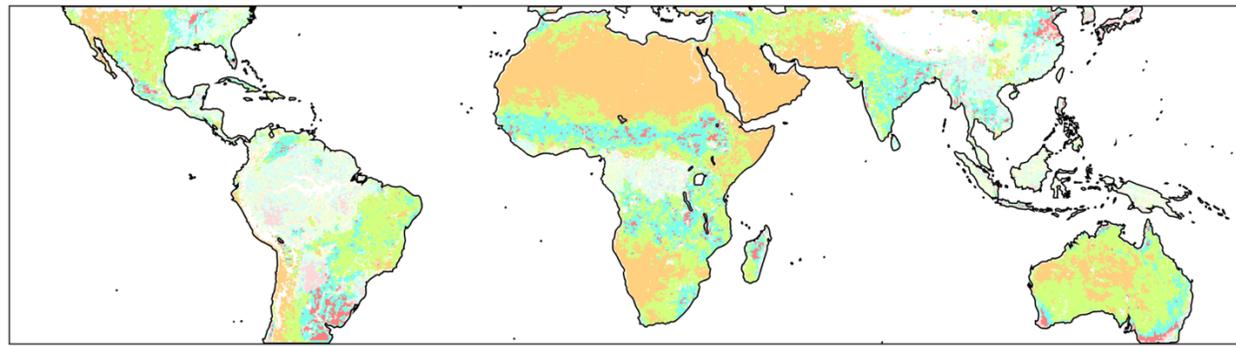


Figure 19. Unbiased root mean square difference between SMAP and CYGNSS soil moisture retrievals. Regions where SMAP always flags the data as being 'poor quality' are semi-transparent, such as the Amazon, central Africa, Indonesia, Japan, southeast Asia, and the majority of the eastern United States. In these regions, you should be careful when using either SMAP or CYGNSS soil moisture data. Higher ubRMSD in regions with 'good quality' SMAP data tend to be found in regions that are seasonally flooded or near coastlines. It is possible that in these areas, the seasonal water influence on CYGNSS reflectivity may overwhelm the soil moisture signal. Or, it is also possible that the soil moisture signal in SMAP data is a red herring, and the brightness temperatures are actually responding to the increase in flooded area instead of soil moisture. Answering this question will be the subject of future research.

3.3 Thoughts on gridding

The gridding scheme described in Section 3.2.7 that utilize ~3.3 km subcells in which to aggregate the CYGNSS data will come as a surprise to some—the majority of researchers analyzing CYGNSS data aggregate the observations to a much larger grid size (say, 25 km). However, we have found that we get the best results when we grid to a much smaller grid size and then upscale the retrievals afterwards.

The following represents a small, and possibly inappropriate, attempt to show why we believe the CYGNSS data respond to land surface characteristics on these scales. We wanted to find an area where we could quantify how much 'blurring' of the CYGNSS signal there is as the soil transitions from dry to wet. These examples are actually not that easy to find (we were trying to avoid looking at transitions between water/dry land), and we settled on looking at the transition between desert and agricultural land in the Punjab region of Pakistan (Figure 20).

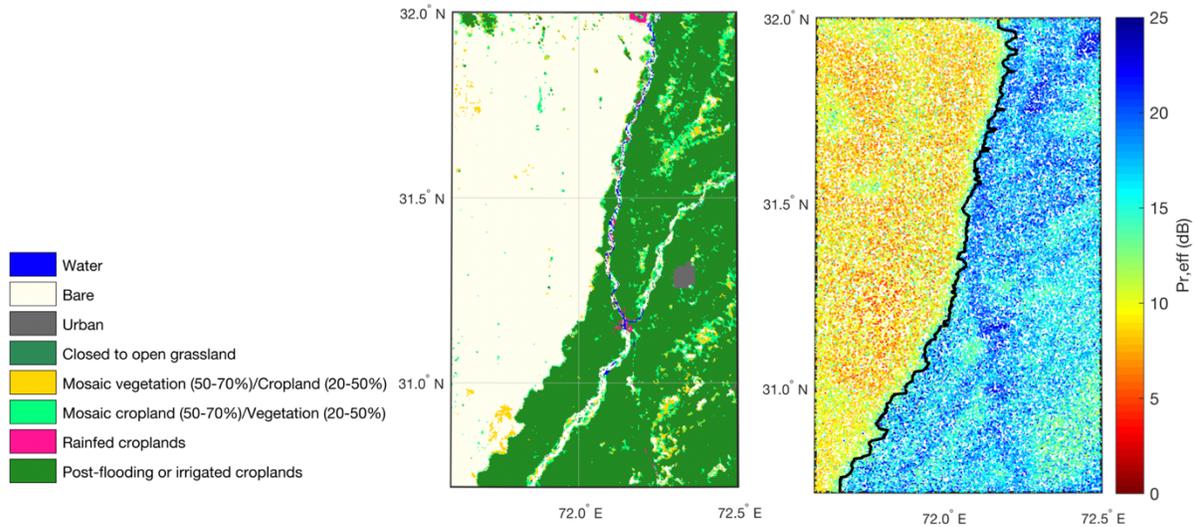


Figure 20. (left) GlobCover2009 landcover map of the transitional region between desert and cropland in Punjab, Pakistan. (right) CYGNSS observations for the same region. The black line is our delineation between bare/croplands, which we gleaned from the landcover map.

As a reminder, we currently think of the spatial footprint of CYGNSS to be an elongated ellipse, as shown in Figure 2a. One might think that, if the patch of land surface contained within this ellipse were completely dry, that $P_{r,eff}$ would be low, and that if the patch of land were wet, that $P_{r,eff}$ would be high. If the ellipse were centered on the transition between wet and dry, such that half the ellipse was wet, and half was dry, that the resulting value of $P_{r,eff}$ would be in between the low and high values (Figure 21). The transition distance between high and low values of $P_{r,eff}$ as it moves across the landscape could be thought of as the blurring of an image.

Figure 21 also shows larger example footprints with diameters of 36 km. In the case of the yellow footprints, the western-most one does not overlap with the cropland, and the resulting signal will not be contaminated from the cropland. In the case of the blue footprints, the western-most one still overlaps with the cropland, so in this case we would expect to see a higher signal than if we also had data from the western-most yellow footprint. In this case, we would expect the blue footprints to show more blurring than the yellow ones.

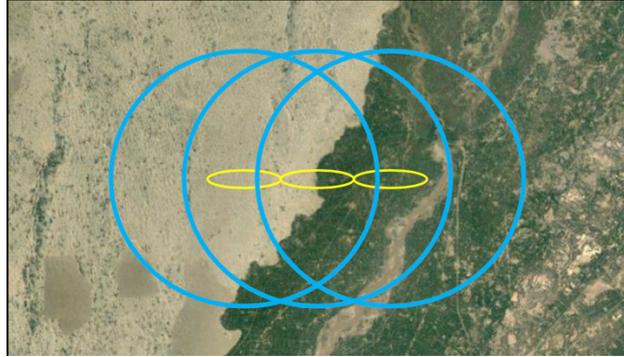


Figure 21. Simplified footprints to illustrate how a remote sensing image will appear more or less blurred depending on footprint size. Yellow footprints are approximately 7 x 0.5 km in size, which is the smallest theoretical footprint for CYGNSS, given its current integration time of 1 second. Blue footprints have a diameter of approximately 36 km.

We attempted to quantify the 'blurriness' of both CYGNSS $P_{r,eff}$ observations and ungridded, Level 1, SMAP brightness temperature observations across this transition zone (Figure 20) by identifying the delineation between desert and croplands using the GlobCover 2009 land cover map (300 m resolution). Our fundamental goal in this exercise was to quantify how 'long' (distance-wise) it took for $P_{r,eff}$ and brightness temperature observations to transition from their mean values over the desert to their mean values over the croplands. The easiest way to do this was to grid the CYGNSS and SMAP observations to the 300 m GlobCover 2009 resolution and then quantify how many grid cells it took for CYGNSS and SMAP to transition across the entire region shown in Figure 20. Of course, the transition line is not directly oriented N-S, so we had to reference all pixels with respect to the line in order to collapse the dependence on latitude (Figure 22).

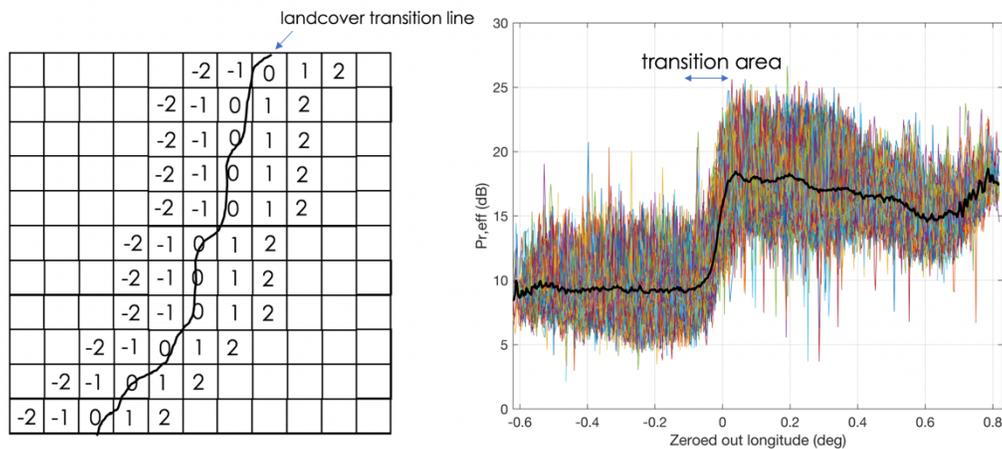


Figure 22. (left) A depiction of how the Punjab was gridded and referenced to the transition line between desert and cropland. (right) Colored lines are gridded $P_{r,eff}$, with each line representing one 300 m strip of grid cells, going N-S. The black line is the mean. The transition area is the distance it takes for $P_{r,eff}$ to change from its mean value over the desert to its mean value over croplands.

We then quantified the transition distance as being the distance that it took for $P_{r,eff}$ to increase from its mean value over the desert, to its mean value over croplands (Figure 23). We found this distance to be 8.65 km, which is only slightly larger than the theoretical smallest along-track resolution of 7 km.

We repeated this exercise for the Level 1 SMAP V-pol brightness temperatures and found the transition distance to be approximately 38 km, which is pretty close to its actual resolution of 40 km; however, these could all be coincidences and need to be analyzed further before any blanket statements can be made.

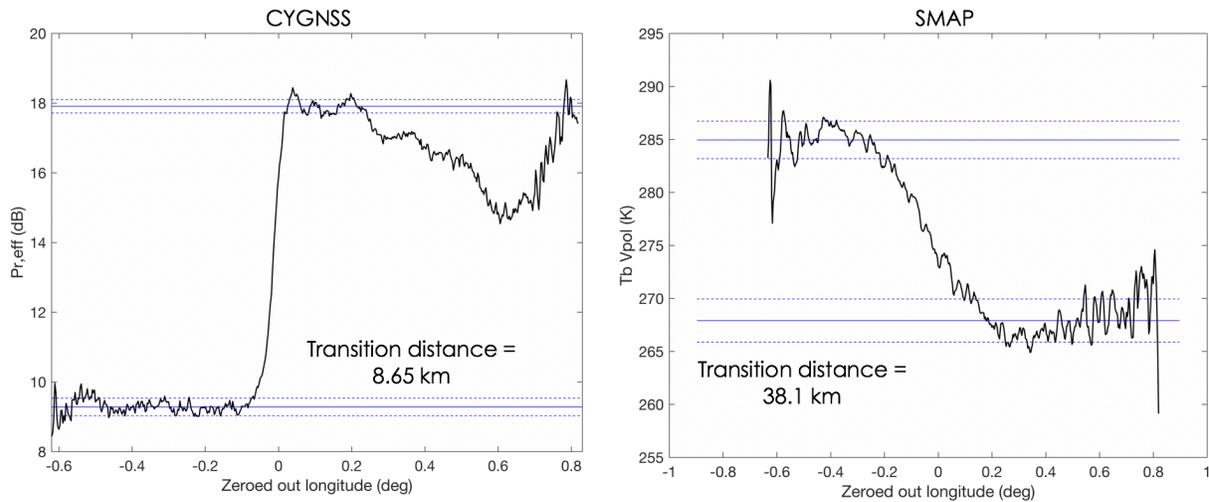


Figure 23. The transition distances across the Punjab for CYGNSS (left) and SMAP (right).

Finally, we have tested various gridding schemes on the CYGNSS data, and we find that both ubRMSD and the correlation between SMAP soil moisture and CYGNSS decrease and increase, respectively, when smaller grid sizes are used (Figure 24).

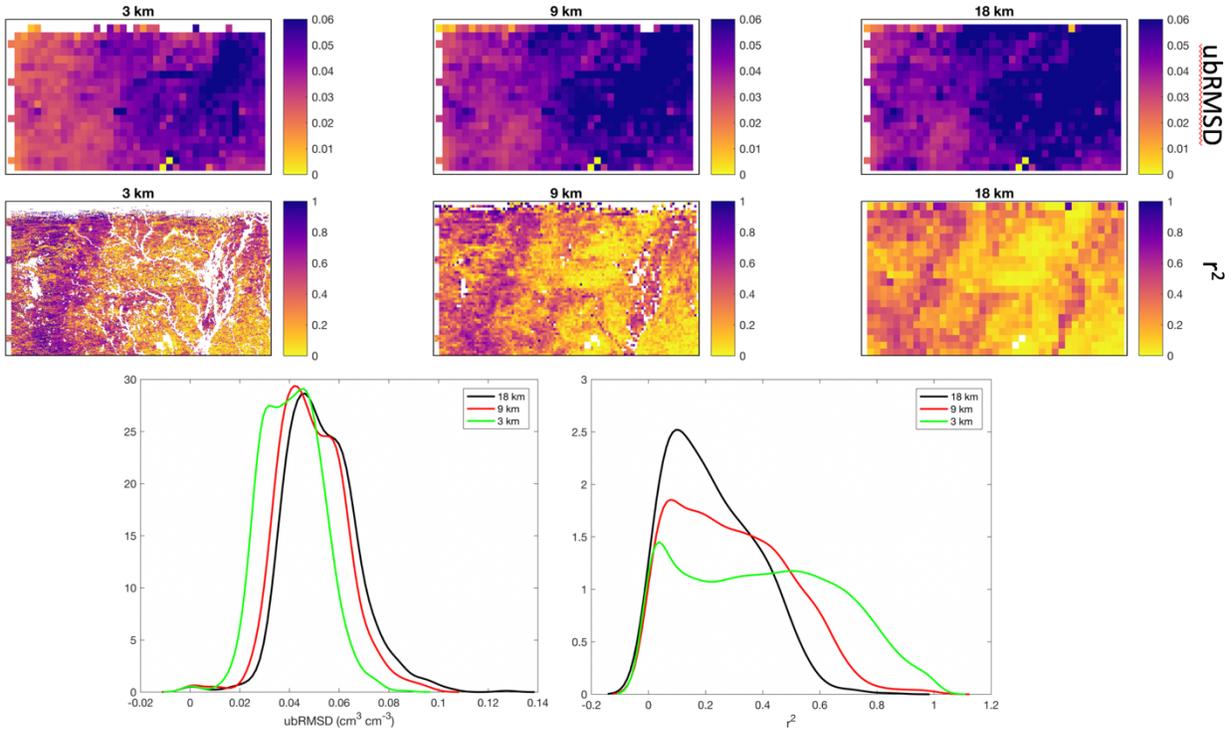


Figure 24. This figure uses a region in Oklahoma to exemplify the effect of the different sub-cell grid sizes on resulting CYGNSS soil moisture retrievals. Longitude and latitude were erroneously not labeled in these figures. (top row) The unbiased root mean square difference ($\text{cm}^3 \text{cm}^{-3}$) between SMAP and CYGNSS soil moisture retrievals when different sub-cell sizes are used—ubRMSD decreases when the sub-cell size decreases. (middle row) The r^2 value between SMAP soil moisture and $P_{r,eff}$ when different sub-cell sizes are used— r^2 increases when sub-cell size decreases. (bottom row) Distributions of ubRMSD ($\text{cm}^3 \text{cm}^{-3}$) and r^2 for the region shown in the top two rows.

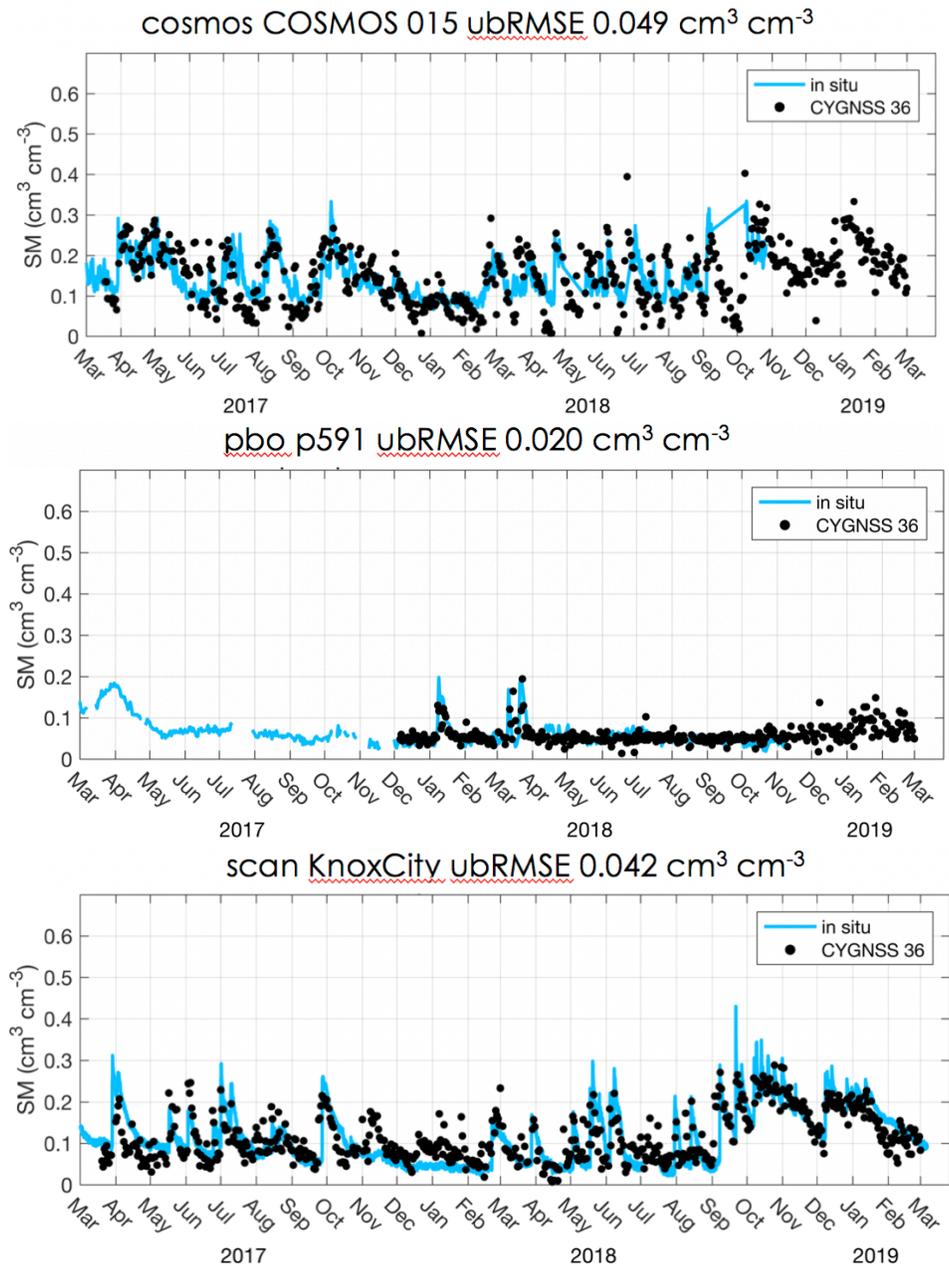
4. In situ validation

We are currently validating the CYGNSS soil moisture retrievals against in situ observations, for the time period March 17, 2017 – March 1, 2019. The networks we have chosen for validation are the following: COSMOS, PBOH2O, SCAN, SNOTEL, and USCRN, though not all networks have data for the entire validation time period. Although other networks exist (like iRON, and SOILSCAPE), we found there to be little to no data that were useful for validation. We also removed some stations from our chosen validation networks that had long periods of non-sensical soil moisture data. In total, we used 203 different sites for validation.

In the figures and tables that follow, we show example CYGNSS soil moisture time series and the unbiased root mean square error (ubRMSE) between CYGNSS and in situ soil moisture, as well as the ubRMSE between SMAP and in situ soil moisture for context. In general, SMAP and CYGNSS showed similar ubRMSEs, which one would expect, given that CYGNSS was calibrated from SMAP. CYGNSS sneaked away with a slightly lower ubRMSE overall (Table 2). Of course, ubRMSE is not a

perfect descriptor of how well SMAP or CYGNSS reproduce in situ time series, but it is the most commonly used one.

Keep in mind that stations within these networks often contain only in situ data for a particular point, and that point may not be representative of the 36 km regional soil moisture. For example, many stations are located near water bodies or in agricultural fields. Stations near the ocean are particularly bad, since the SMAP data near coastlines are generally not at all representative of the coastal soil moisture. Given that the *in situ* observations used for validation by SMAP are not available to the public, we had to make do with these.



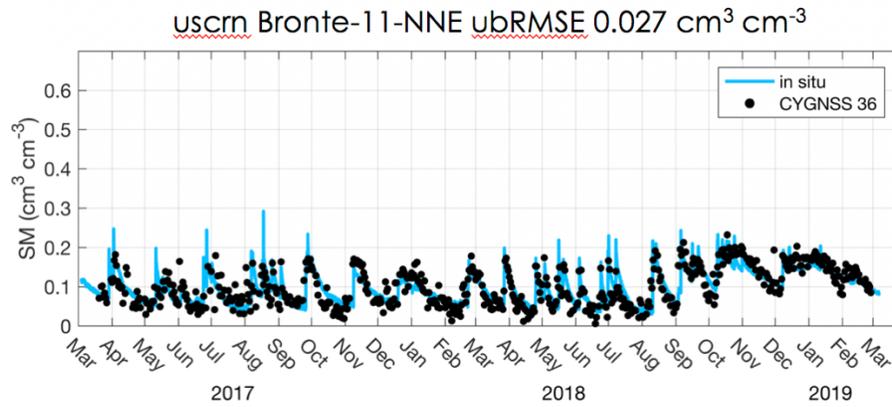


Figure 25. Example time series from various in situ validation networks (blue line) with CYGNSS soil moisture retrievals (black dots).

Table 2. Unbiased root mean square error between CYGNSS soil moisture/in situ and SMAP soil moisture/in situ, for all 203 stations and divided by network.

Network	ubRMSE CYGNSS cm ³ cm ⁻³	ubRMSE SMAP cm ³ cm ⁻³
All	0.0471	0.0502
COSMOS	0.0426	0.0407
PBOH2O	0.0488	0.0471
SCAN	0.0461	0.0477
SNOTEL	0.0692	0.0754
USCRN	0.0438	0.0496

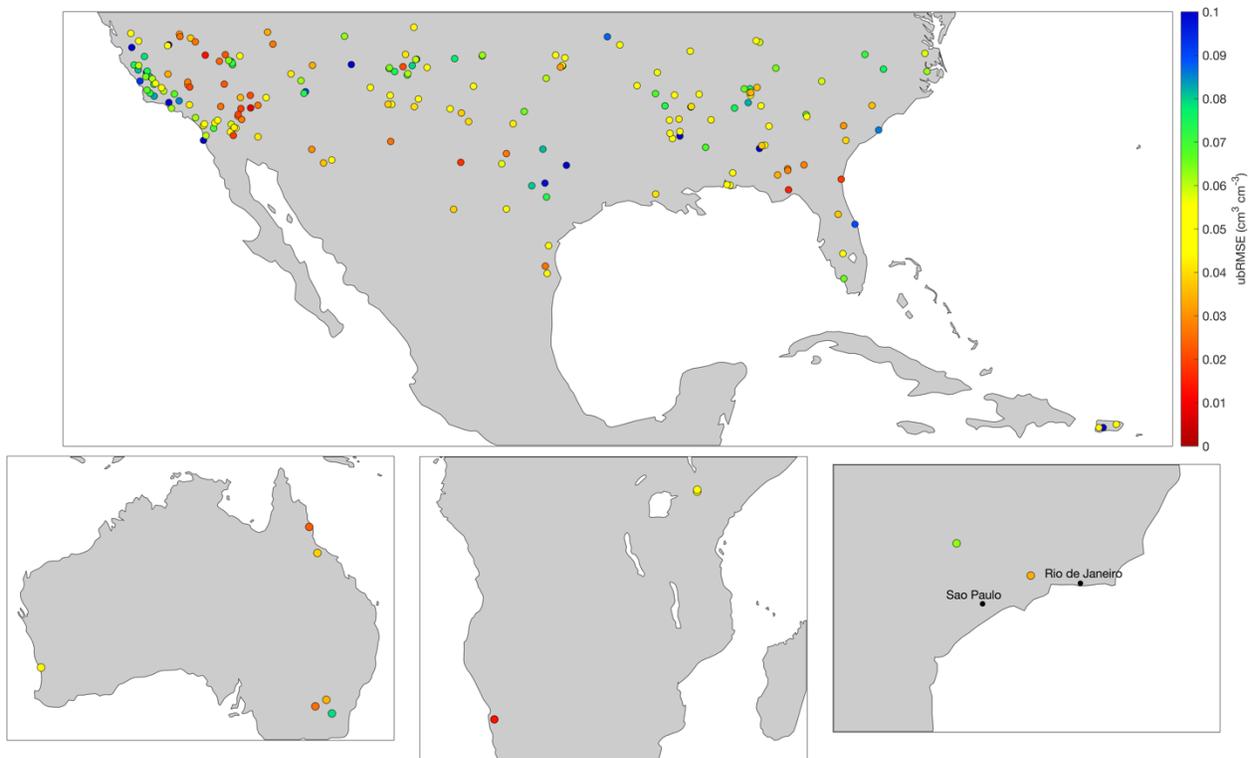


Figure 26. Maps showing the unbiased root mean square error between CYGNSS soil moisture and in situ observations around the world.

Table 3. In situ soil moisture sites used for validation and the unbiased root mean square errors between CYGNSS and in situ/ SMAP and in situ. Also shown are the number of observations used for validation—many in situ sites did not have data for the full time period used for validation (March 17, 2017 – March 1, 2019). Table spans 3 pages. Bias, slope, r values, etc. can be

provided upon request. Tan cells are those where CYGNSS had a smaller ubRMSE than SMAP, though in general they were very similar.

Network	Station	Lat	Lon	ubRMSE		# Obs CYGNSS	# Obs SMAP
				CM3 CM-3	CM3 CM-3		
'cosmos'	'COSMOS_015'	36.61	-97.49	0.049	0.062	441	266
'cosmos'	'COSMOS_064'	35.19	-102.10	0.049	0.044	127	80
'cosmos'	'COSMOS_024'	33.73	-117.70	0.062	0.159	338	209
'cosmos'	'COSMOS_101'	-22.68	-45.00	0.034	0.020	350	247
'cosmos'	'COSMOS_023'	33.61	-116.45	0.045	0.031	542	328
'cosmos'	'COSMOS_061'	35.45	-111.77	0.062	0.027	105	176
'cosmos'	'COSMOS_057'	29.95	-98.00	0.073	0.056	319	140
'cosmos'	'COSMOS_081'	-31.38	115.71	0.043	0.125	132	64
'cosmos'	'COSMOS_089'	-23.55	15.05	0.013	0.013	23	14
'cosmos'	'COSMOS_067'	34.26	-89.87	0.040	0.050	491	293
'cosmos'	'COSMOS_052'	31.24	-84.46	0.028	0.042	45	30
'cosmos'	'COSMOS_055'	0.28	36.87	0.050	0.042	231	239
'cosmos'	'COSMOS_050'	0.49	36.87	0.049	0.048	152	121
'cosmos'	'COSMOS_034'	37.07	-119.19	0.111	0.056	141	311
'cosmos'	'COSMOS_044'	-21.62	-47.63	0.063	0.030	187	188
'cosmos'	'COSMOS_073'	-17.12	145.63	0.023	0.031	31	33
'cosmos'	'COSMOS_014'	36.06	-97.22	0.032	0.034	539	314
'cosmos'	'COSMOS_033'	37.03	-119.26	0.047	0.047	28	31
'cosmos'	'COSMOS_085'	-34.40	147.53	0.034	0.049	153	86
'cosmos'	'COSMOS_077'	-35.66	148.15	0.079	0.082	83	70
'cosmos'	'COSMOS_074'	-19.88	146.54	0.038	0.026	110	76
'cosmos'	'COSMOS_079'	-35.01	146.30	0.025	0.039	145	86
'pbo'	'bkap'	35.29	-116.08	0.023	0.024	437	277
'pbo'	'crrs'	33.07	-115.74	0.046	0.047	476	206
'pbo'	'csci'	34.17	-119.04	0.062	0.046	397	238
'pbo'	'ctdm'	34.52	-118.61	0.085	0.063	230	179
'pbo'	'fgst'	34.73	-120.01	0.083	0.061	335	274
'pbo'	'glrs'	33.27	-115.52	0.043	0.122	437	203
'pbo'	'gnps'	34.31	-114.19	0.026	0.030	371	253
'pbo'	'hunt'	35.88	-120.40	0.074	0.055	297	248
'pbo'	'hvys'	34.44	-119.19	0.101	0.135	404	275
'pbo'	'imps'	34.16	-115.15	0.020	0.022	465	236
'pbo'	'masw'	35.83	-120.44	0.106	0.091	330	257
'pbo'	'ndap'	34.77	-114.62	0.021	0.017	420	274
'pbo'	'ok02'	36.49	-96.96	0.050	0.048	65	43
'pbo'	'p008'	36.14	-111.13	0.034	0.037	127	167
'pbo'	'p010'	34.67	-113.73	0.049	0.047	62	112
'pbo'	'p035'	34.60	-105.18	0.055	0.055	276	273
'pbo'	'p036'	36.42	-105.29	0.100	0.103	99	120
'pbo'	'p038'	34.15	-103.41	0.047	0.034	291	203
'pbo'	'p039'	36.45	-103.15	0.078	0.060	273	267
'pbo'	'p070'	36.04	-104.70	0.049	0.045	282	260
'pbo'	'p094'	37.20	-117.70	0.027	0.025	120	196
'pbo'	'p107'	35.13	-107.88	0.049	0.055	224	247
'pbo'	'p123'	36.64	-105.91	0.041	0.043	236	248
'pbo'	'p250'	36.95	-121.27	0.121	0.090	225	246
'pbo'	'p255'	37.58	-121.32	0.048	0.024	139	265
'pbo'	'p284'	35.93	-120.91	0.081	0.049	320	204
'pbo'	'p288'	36.14	-120.88	0.060	0.043	382	251
'pbo'	'p472'	32.89	-117.10	0.060	0.047	417	252
'pbo'	'p474'	33.36	-117.25	0.057	0.041	442	273
'pbo'	'p475'	32.67	-117.24	0.101	0.158	437	264
'pbo'	'p482'	33.24	-116.67	0.069	0.057	128	274
'pbo'	'p498'	32.90	-115.57	0.018	0.021	472	205
'pbo'	'p505'	33.42	-115.69	0.059	0.124	434	201
'pbo'	'p508'	33.25	-115.43	0.047	0.121	431	199
'pbo'	'p511'	33.89	-115.30	0.020	0.023	457	253
'pbo'	'p514'	35.01	-120.41	0.067	0.045	390	244
'pbo'	'p515'	34.87	-120.24	0.073	0.061	98	194
'pbo'	'p525'	35.43	-120.81	0.093	0.121	382	266
'pbo'	'p530'	35.62	-120.48	0.065	0.050	333	258
'pbo'	'p532'	35.63	-120.27	0.081	0.068	310	244
'pbo'	'p536'	35.28	-120.03	0.075	0.061	287	248
'pbo'	'p537'	35.32	-119.94	0.047	0.032	301	261

'pbo'	'p538'	35.53	-120.11	0.061	0.054	339	236
'pbo'	'p553'	34.84	-118.88	0.067	0.070	378	271
'pbo'	'p565'	35.74	-119.24	0.034	0.032	442	268
'pbo'	'p568'	35.25	-118.13	0.034	0.029	191	206
'pbo'	'p569'	35.38	-118.12	0.027	0.036	134	161
'pbo'	'p587'	34.33	-118.03	0.042	0.039	257	203
'pbo'	'p591'	35.15	-118.02	0.020	0.034	275	214
'pbo'	'p623'	34.19	-114.60	0.008	0.010	26	19
'pbo'	'p645'	37.54	-118.59	0.033	0.062	11	242
'pbo'	'p724'	37.44	-118.56	0.025	0.023	12	232
'pbo'	'p742'	33.50	-116.60	0.049	0.048	262	272
'pbo'	'p807'	30.49	-98.82	0.080	0.058	393	240
'pbo'	'p811'	35.15	-118.02	0.020	0.030	275	214
'pbo'	'qcy2'	36.16	-121.14	0.075	0.054	397	197
'pbo'	'sdh1'	34.26	-116.28	0.026	0.018	268	272
'scan'	'AAMU-1tg'	34.78	-86.55	0.055	0.059	608	349
'scan'	'AdamsRanch#1'	34.25	-105.42	0.040	0.038	335	343
'scan'	'Alcalde'	36.08	-106.05	0.021	0.032	123	160
'scan'	'AllenFarms'	35.07	-86.90	0.066	0.059	559	328
'scan'	'BraggFarm'	34.90	-86.60	0.050	0.047	607	354
'scan'	'BroadAcres'	32.28	-86.05	0.103	0.082	217	140
'scan'	'Charklin'	36.37	-115.83	0.064	0.052	348	352
'scan'	'CochoraRanch'	35.12	-119.60	0.046	0.033	312	351
'scan'	'Corozal'	18.32	-66.03	0.047	0.054	30	29
'scan'	'DeathValleyJCT'	36.33	-116.35	0.024	0.026	156	175
'scan'	'DeepSprings'	37.37	-117.97	0.038	0.031	15	308
'scan'	'DesertCenter'	33.80	-115.31	0.027	0.026	474	268
'scan'	'Dexter'	36.78	-89.93	0.044	0.075	590	337
'scan'	'EastviewFarm'	35.13	-86.18	0.036	0.040	437	267
'scan'	'Enterprise'	37.63	-113.65	0.034	0.048	5	319
'scan'	'Essex'	34.67	-115.17	0.035	0.022	528	265
'scan'	'FordDryLake'	33.65	-115.10	0.027	0.020	584	341
'scan'	'FortReno#1'	35.55	-98.02	0.061	0.063	625	351
'scan'	'GoodwinCreekTimber'	34.23	-89.90	0.077	0.055	582	354
'scan'	'GuilarteForest'	18.15	-66.77	0.117	0.115	369	263
'scan'	'Kingsville'	27.55	-97.88	0.044	0.040	201	118
'scan'	'KnoxCity'	33.45	-99.87	0.042	0.047	596	335
'scan'	'KoptisFarms'	30.52	-87.70	0.048	0.046	537	315
'scan'	'Levelland'	33.55	-102.37	0.040	0.052	402	351
'scan'	'LittleRiver'	31.50	-83.55	0.028	0.031	381	265
'scan'	'LosLunasPmc'	34.77	-106.77	0.047	0.043	378	335
'scan'	'LovellSummit'	36.17	-115.62	0.072	0.063	273	350
'scan'	'MammothCave'	37.18	-86.03	0.059	0.055	444	347
'scan'	'MaricaoForest'	18.15	-67.00	0.043	0.042	445	263
'scan'	'Mayday'	32.87	-90.52	0.117	0.102	472	353
'scan'	'McalisterFarm'	35.07	-86.58	0.076	0.062	607	354
'scan'	'MccrackenMesa'	37.45	-109.33	0.063	0.052	23	333
'scan'	'MonoclineRidge'	36.54	-120.55	0.079	0.056	535	351
'scan'	'MorrisFarms'	32.42	-85.92	0.037	0.042	579	354
'scan'	'MtVernon'	37.07	-93.88	0.056	0.051	475	343
'scan'	'NorthIssaquena'	33.00	-91.07	0.051	0.078	585	274
'scan'	'Onward'	32.75	-90.93	0.052	0.053	218	305
'scan'	'PeeDee'	34.30	-79.73	0.036	0.042	382	221
'scan'	'PerdidoRivFarms'	31.12	-87.55	0.054	0.040	617	354
'scan'	'PineNut'	36.57	-115.20	0.046	0.039	304	344
'scan'	'ReynoldsHomestead'	36.63	-80.13	0.071	0.050	324	223
'scan'	'Riesel'	31.48	-96.88	0.104	0.089	489	284
'scan'	'RiverRoadFarms'	31.02	-85.03	0.034	0.043	596	349
'scan'	'SanAngelo'	31.55	-100.51	0.057	0.058	482	266
'scan'	'SandHollow'	37.10	-113.35	0.027	0.026	225	345
'scan'	'SandyRidge'	33.67	-90.57	0.042	0.070	542	330
'scan'	'Scott'	33.62	-91.10	0.049	0.074	589	264
'scan'	'SellersLake#1'	29.10	-81.63	0.037	0.056	523	356
'scan'	'Sevilleta'	34.35	-106.68	0.045	0.035	234	128
'scan'	'SilverCity'	33.08	-90.52	0.043	0.063	472	353
'scan'	'StanleyFarm'	34.43	-86.68	0.082	0.059	619	355

'scan'	'Starkville'	33.63	-88.77	0.043	0.049	595	354
'scan'	'Stephenville'	32.25	-98.20	0.082	0.058	626	286
'scan'	'Stubblefield'	34.97	-119.48	0.061	0.042	390	351
'scan'	'SudduthFarms'	34.18	-87.45	0.076	0.058	603	274
'scan'	'Tidewater#1'	35.87	-76.65	0.063	0.064	337	210
'scan'	'TidewaterArec'	36.68	-76.77	0.058	0.046	566	349
'scan'	'Tuskegee'	32.43	-85.75	0.054	0.046	612	282
'scan'	'UAPBDewitt'	34.28	-91.35	0.074	0.066	339	328
'scan'	'UAPBLonokeFarm'	34.85	-91.88	0.069	0.051	565	354
'scan'	'UAPBMarianna'	34.78	-90.82	0.052	0.065	603	266
'scan'	'UAPBPointRemove'	35.22	-92.92	0.047	0.037	595	354
'scan'	'Uvalde'	29.36	-100.25	0.054	0.046	615	353
'scan'	'Vernon'	34.02	-99.25	0.065	NaN	609	NaN
'scan'	'WTARS'	34.90	-86.53	0.035	0.044	396	232
'scan'	'Wakulla#1'	30.30	-84.42	0.016	0.026	424	247
'scan'	'WalnutGulch#1'	31.73	-110.05	0.045	0.032	389	260
'scan'	'Watkinsville#1'	33.88	-83.43	0.068	0.036	259	143
'scan'	'Wedowee'	33.33	-85.52	0.045	0.035	460	214
'scan'	'Weslaco'	26.16	-97.96	0.053	0.050	524	355
'scan'	'YoumansFarm'	32.67	-81.20	0.039	0.045	628	354
'snotel'	'BRISTLECONETRAIL'	36.32	-115.70	0.095	0.099	331	345
'snotel'	'BarM'	34.86	-111.61	0.074	0.068	276	309
'snotel'	'ElkCabin'	35.70	-105.81	0.069	0.052	213	243
'snotel'	'LEECANYON'	36.31	-115.68	0.080	0.082	331	345
'snotel'	'MEDANOPASS'	37.85	-105.44	0.044	0.065	4	261
'snotel'	'MormonMountain'	34.94	-111.52	0.090	0.090	276	309
'snotel'	'NAVAJOWHISKEYCK'	36.18	-108.95	0.100	0.082	230	229
'snotel'	'PALO'	36.41	-105.33	0.060	0.080	318	205
'snotel'	'RAINBOWCANYON'	36.25	-115.63	0.070	0.074	331	345
'snotel'	'SantaFe'	35.77	-105.78	0.062	0.074	156	201
'snotel'	'SenoritaDivide#2'	36.00	-106.83	0.066	0.071	269	303
'snotel'	'TresRitos'	36.13	-105.53	0.079	0.084	98	222
'snotel'	'VacasLocas'	36.03	-106.81	0.064	0.067	269	303
'uscrn'	'Asheville-13-S'	35.42	-82.56	0.059	0.058	370	330
'uscrn'	'Austin-33-NW'	30.62	-98.08	0.098	0.080	619	263
'uscrn'	'Batesville-8-WNW'	35.82	-91.78	0.057	0.054	601	348
'uscrn'	'Blackville-3-W'	33.36	-81.33	0.031	0.032	173	106
'uscrn'	'Bowling-Green-21-NNE'	37.25	-86.23	0.047	0.059	434	343
'uscrn'	'Bronte-11-NNE'	32.04	-100.25	0.027	0.038	596	312
'uscrn'	'Brunswick-23-S'	30.81	-81.46	0.019	0.048	521	353
'uscrn'	'Crossville-7-NW'	36.01	-85.13	0.065	0.055	372	224
'uscrn'	'Durham-11-W'	35.97	-79.09	0.076	0.050	566	347
'uscrn'	'Edinburg-17-NNE'	26.53	-98.06	0.026	0.033	496	304
'uscrn'	'Elgin-5-S'	31.59	-110.51	0.036	0.028	369	254
'uscrn'	'Everglades-City-5-NE'	25.90	-81.32	0.065	0.077	396	264
'uscrn'	'Fairhope-3-NE'	30.55	-87.88	0.047	0.063	284	170
'uscrn'	'Fallbrook-5-NE'	33.44	-117.19	0.053	0.026	554	350
'uscrn'	'Gadsden-19-N'	34.29	-85.96	0.052	0.037	590	352
'uscrn'	'Goodwell-2-E'	36.60	-101.60	0.063	0.059	372	345
'uscrn'	'Goodwell-2-SE'	36.57	-101.61	0.063	0.060	372	345
'uscrn'	'Holly-Springs-4-N'	34.82	-89.43	0.052	0.048	610	348
'uscrn'	'Joplin-24-N'	37.43	-94.58	0.087	0.077	193	340
'uscrn'	'Lafayette-13-SE'	30.09	-91.87	0.041	0.056	556	263
'uscrn'	'Las-Cruces-20-N'	32.61	-106.74	0.026	0.029	386	259
'uscrn'	'Los-Alamos-13-W'	35.86	-106.52	0.075	0.065	18	248
'uscrn'	'McClellanville-7-NE'	33.15	-79.36	0.086	0.085	549	333
'uscrn'	'Merced-23-WSW'	37.24	-120.88	0.046	0.054	430	270
'uscrn'	'Mercury-3-SSW'	36.62	-116.02	0.022	0.020	325	349
'uscrn'	'Monahans-6-ENE'	31.62	-102.81	0.018	0.025	380	348
'uscrn'	'Muleshoe-19-S'	33.96	-102.77	0.038	0.041	380	325
'uscrn'	'Newton-5-ENE'	32.34	-89.07	0.069	0.055	605	349
'uscrn'	'Newton-8-W'	31.31	-84.47	0.029	0.051	608	348
'uscrn'	'Panther-Junction-2-N'	29.35	-103.21	0.038	0.027	358	348
'uscrn'	'Sebring-23-SSE'	27.15	-81.37	0.047	0.047	52	29
'uscrn'	'Socorro-20-N'	34.36	-106.89	0.038	0.034	370	342
'uscrn'	'Stillwater-2-W'	36.12	-97.09	0.077	0.059	610	344
'uscrn'	'Stillwater-5-WNW'	36.13	-97.11	0.055	0.045	596	347
'uscrn'	'Stovepipe-Wells-1-SW'	36.60	-117.14	0.013	0.013	495	350
'uscrn'	'Titusville-7-E'	28.62	-80.69	0.090	0.087	382	294
'uscrn'	'Tucson-11-W'	32.24	-111.17	0.030	0.021	447	349
'uscrn'	'Watkinsville-5-SSE'	33.78	-83.39	0.047	0.034	553	331
'uscrn'	'Williams-35-NNW'	35.76	-112.34	0.041	0.042	351	268
'uscrn'	'Yuma-27-ENE'	32.83	-114.19	0.042	0.034	572	329

5. Plans for future versions

We hope to keep improving our product: a more robust calibration of the signal over land is a priority. We are considering releasing the 3 km soil moisture retrievals themselves, if there is sufficient interest. We would also like to explore 'smart gridding' of the data that allows for a more flexible mesh based on land cover type or topography, to remove imposing arbitrary grid lines on the landscape. In the future, we will also explore parameter regionalization to increase the accuracy of CYGNSS soil moisture retrievals near coastlines, where SMAP has trouble retrieving soil moisture due to its larger footprint.

6. File overview and loading the data

Spatial coverage:	N: 38, S: -38 E: 164, W: -135	Data format:	netCDF4
Spatial resolution:	36 km x 36 km	Platform:	CYGNSS
Temporal coverage:	18 March 2017 to present	Sensor:	CYGNSS GNSS-R receivers
Temporal resolution:	6 hours	Version:	V1.0
Data contributors:	Chew, C.C., E. E. Small		

File naming convention: ucar_cu_cygnss_sm_v1_YYYY_DDD.nc

YYYY: 4 digit year

DDD: 3 digit day of year

Each netCDF file contains the following variables:

latitude: Refers to the latitude of the center of the grid cell. Dimensions: 252 x 802

longitude: Refers to the longitude of the center of the grid cell. Dimensions: 252 x 802

timeintervals: The start and stop time for the subdaily soil moisture retrievals. For example, the first row is (0,6), which means that the first of the reported subdaily soil moisture retrievals were recorded between midnight and 6 am. Dimensions: 4 x 2

SM_daily: The average soil moisture for each grid cell recorded during the full 24 hr period. Dimensions: 252 x 802

SM_subdaily: The average soil moisture for each grid cell recorded during each specified time interval. Dimensions: 252 x 802 x 4

SIGMA_daily: The standard deviation of soil moisture observations for each grid cell for the full 24 hr period. Dimensions: 252 x 802

SIGMA_subdaily: The standard deviation of soil moisture observations for each grid cell during each time interval. Dimensions: 252 x 802 x 4

7. Quality flags—important

In order to keep file size to a minimum, we provide quality flags in a separate file (BoulderCYGSM_static_flags.nc). Just because they are in a separate file does not mean you can ignore them, though. As we've tried to emphasize throughout this handbook, the retrievals we've provided is only version 1 and has known problems.

To encourage the use of the quality flags, we've provided them in simple grids. We don't mean to say that you shouldn't use data that are flagged, but you should use them with caution and not be surprised if retrievals aren't what you would expect.

Quality flag meanings and how we derived them. For all flags, a value of 1 = true, a value of 0 = false:

latitude: Refers to the latitude of the center of the grid cell. Dimensions: 252 x 802

longitude: Refers to the longitude of the center of the grid cell. Dimensions: 252 x 802

flag_poor_SMAP: Indicates that CYGNSS was calibrated to SMAP data where a large portion (>90%) of the SMAP soil moisture retrievals were flagged as 'not recommended for retrieval.' Dimensions: 252 x 802

flag_small_SM_range: Indicates that CYGNSS was calibrated to SMAP data with a small range of soil moisture values ($< 0.1 \text{ cm}^3 \text{ cm}^{-3}$), which means the uncertainty in β is large. Dimensions: 252 x 802

flag_high_ubrmsd: Indicates a high unbiased root mean square difference between CYGNSS and SMAP retrievals ($> 0.08 \text{ cm}^3 \text{ cm}^{-3}$). Dimensions: 252 x 802

flag_few_obs: Indicates a small number of observations in the grid cell for calibration, leading to a less certain β ($n < 100$). Dimensions: 252 x 802

flag_low_signal: Indicates low mean $P_{r,eff}$ after water point removal in the cell, which likely means that roughness or vegetation effects are dominate (mean $P_{r,eff} < 5 \text{ dB}$). Dimensions: 252 x 802