Correspondence to:
C. Gerlein-Safdi,
cgerlein@umich.edu

A CYGNSS-Based Algorithm for the Detection of Inland Waterbodies

Cynthia Gerlein-Safdi[1] (iD) and Christopher S. Ruf[1] (iD)

[1]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

Abstract The Cyclone Global Navigation Satellite System (CYGNSS) is a new constellation of eight low Earth orbiting spacecrafts that receive both direct and reflected signals from GPS satellites. Coherent reflection of the GPS signal from standing water over land results in a high surface reflectivity signal in the CYGNSS data. An image processing algorithm is presented, which leverages the surface reflectivity signal to produce a watermask of inland waterbodies at $0.01° \times 0.01°$ spatial resolution. The watermask is compared to hand-drawn maps of inland waterbodies, as well as to the MODIS watermask product. We find that the CYGNSS watermask provides accurate, time-varying maps that are able to resolve changes in lake and river position and extent. With CYGNSS' short return time, watermasks can be generated using as little as half a month of data to produce near-real-time maps of flooding events.

1. Introduction

The position of waterbodies is a defining characteristic of any landscape, yet the inherent changes in waterbody extent and position are key to the ecosystems they are part of (Leira & Cantonati, 2008). Mapping waterbodies at the global scale presents a unique challenge because of the variety of terrains and water surface characteristics found around the world (Alsdorf et al., 2007). Two main watermask products are currently available. The Terra Moderate Resolution Imaging Spectroradiometer (MODIS) Land Water Mask (MOD44W, https://lpdaac.usgs.gov/products/mod44wv006/) Version 6 data product provides a yearly global map of surface water at 250-m spatial resolution between 2000 and 2015 (Carroll et al., 2017). This product is widely used to mask waterbodies in other remote sensing products. Based on Landsat data, the "Pekel" watermask (https://global-surface-water.appspot.com/download) provides yearly masks at 30-m resolution from 1982 to 2018 (Pekel et al., 2016). Both products are based on optical remote sensing and are not able to detect water under vegetation or clouds. This is a significant limitation in tropical areas such as the Amazon, because many small streams are fully covered by vegetation. Clouds can also be an issue in these areas, since cloud cover can approach 100% during the rainy season (Martins et al., 2018), leading to watermasks biased toward dry season water levels. Finally, the slow return time and the need to assimilate multiple overpasses to generate cloud-free maps limit the frequency of these maps to a yearly timescale.

Launched in December 2016, the Cyclone Global Navigation Satellite System (CYGNSS) is a constellation of eight low Earth orbiting spacecraft receiving both direct and reflected signals from Global Positioning System (GPS) satellites (Ruf et al., 2016). GPS satellites operate at a frequency of 1.575 GHz (L-band), allowing the CYGNSS satellites to see through cloud, rain, and all but the densest of vegetation canopies (Li et al., 2019). Designed to estimate wind speed over oceans, CYGNSS also proved to be highly sensitive to both standing water over land (Chew et al., 2018; Ruf et al., 2018) and soil moisture (Chew et al., 2016; Chew & Small, 2018; Ruf et al., 2018). Standing water has a particularly strong signal because flat surfaces such as the calm waters of a lake or a slow-flowing river will produce coherent specular scattering that generates a much larger signal than the diffuse scattering coming from surrounding land or oceans (Ruf et al., 2018). In addition, because of the GPS' operating frequency, the signal is only partially attenuated by vegetation, and rivers or lake under canopies can still be detected by CYGNSS (Ruf et al., 2018).

The footprint of a single CYGNSS return will depend on the incidence angle and the relative contribution from diffuse and specular scattering. Over land, the footprint cross track is estimated to be between 0.5 and 1 km depending on the incidence angle and the along-track resolution is estimated to be 7 km (3.5 km since July 2019). The ability to resolve smooth, small water features is enhanced by the large difference in reflected signal strength between coherent and incoherent scatterers. Coherent scattering from smooth surfaces is typically ~30 dB stronger than incoherent scattering from rough surfaces, provided that

the smooth surface extends over the complete first Fresnel zone (~0.5 × 0.5 km², Geremia-Nievinski et al., 2016; Camps, 2019). Waterbodies that are significantly smaller in size can also be resolved since their scattered signal strength is reduced by the fractional aerial extent of the water body relative to the first Fresnel zone. For example, the signal scattered from a 0.1 × 0.1 km² water body would still be ~16 dB stronger than that scattered incoherently from the rough surface surrounding the small water body. Because the CYGNSS signal over land is affected by surface type and roughness, soil and vegetation water content, and vegetation density, the signal-to-noise ratio (SNR) of the reflected signal that is actually measured by CYGNSS is highly variable, making it difficult to capture waterbodies globally, especially small ones, using only a thresholding method. Here, we propose an algorithm to clean and process CYGNSS SNR data and produce a reliable global watermask over the full latitudinal extent of CYGNSS coverage from about 40°N to 40°S.

2. Materials and Methods

The entire analysis described in the following sections was done in Python 3.0. An example code is provided in the Supporting Information.

2.1. CYGNSS Data and SNR Correction

Here we use version 2.1 of the CYGNSS Level 1 data (available online at https://podaac.jpl.nasa.gov/dataset/CYGNSS_L1_V2.1). In particular, we use the DDM SNR derived from the delay-Doppler Maps (DDM) recorded by CYGNSS. The SNR is corrected for transmitted power, receiving, and GPS antenna gains, as well as for transmitter-to-specular-point and specular-point-to-receiver ranges. We tested corrections adapted to both coherent and diffuse scattering and found that the former correction was better at removing the dependence of the data on the different parameters (Figure 1). The expression for corrected surface reflectivity (SR) for the coherent component of the scattered power is (Chew et al., 2018)

$$\text{SR}^{\text{coherent}} = \text{SNR} - P_r^t - G^r - G^t - 20\log_{10}(\lambda) + 20\log_{10}(\text{TxSP} + \text{SPRx}) + 20\log_{10}(4\pi),$$

where P_r^t is the transmitted power (in dBW), G^r and G^t are the receiving and transmitter antenna gains (in dB), respectively, λ is the GPS wavelength (equal to 0.19 m), and TxSP and SPRx are the distances between the transmitter and the specular point and the specular point and the receiver (in meters), respectively. In the case of the diffuse scattering correction, the expression becomes

$$\text{SR}^{\text{diffuse}} = \text{SNR} - P_r^t - G^r - G^t - 20\log_{10}(\lambda) + 20\log_{10}(\text{TxSP}) + 20\log_{10}(\text{SPRx}) + 30\log_{10}(4\pi),$$

where the effect of the ranges is modified according to the bistatic radar equation for diffuse scattering (Ulaby et al., 2014). Figure 1 shows the dependency of the SNR, $\text{SR}^{\text{diffuse}}$, and $\text{SR}^{\text{coherent}}$ on transmitted power, antenna gains, and ranges. The SR shows a smaller dependence on these parameters after correcting for them using the coherent component of the scattered power ($\text{SR}^{\text{coherent}}$) than when using $\text{SR}^{\text{diffuse}}$, and we therefore use the correction $\text{SR}^{\text{coherent}}$ in the following.

As in Chew et al. (2018), the mean of the bottom 5% of SR values is removed to produce maps in a more intuitive range of dB values. Ocean data are removed using quality flags. All the track-based SR CYGNSS data for the year 2018 are then combined and gridded into a 0.01° × 0.01° grid. At this resolution, samples from a single track fall into different grid cells. When multiple samples from different tracks fall within the same grid cell, we use the average of all the samples within that cell. The use of the median, peak, and 90th percentile values was also explored, but because of the low number of data point per cell (maximum 12), we found only small differences between use of the difference statistics. At this resolution, more than 76% of the map's grid cells have data. Cells without data are assigned to Not-A-Number (NaN).

2.2. Isolated Pixel Removal

Because of the spatial sampling properties of the CYGNSS data (samples taken along one-dimensional swaths that are referred to as "tracks") and the multiple CYGNSS and GPS satellites, we find that some tracks have a very high SR that appears to be track-based and not related to surface properties. These high SR tracks are likely due to variations in GPS power (Wang et al., 2019) and will be corrected for in upcoming versions of the data. To remove these here, we identified small, isolated clusters of very high
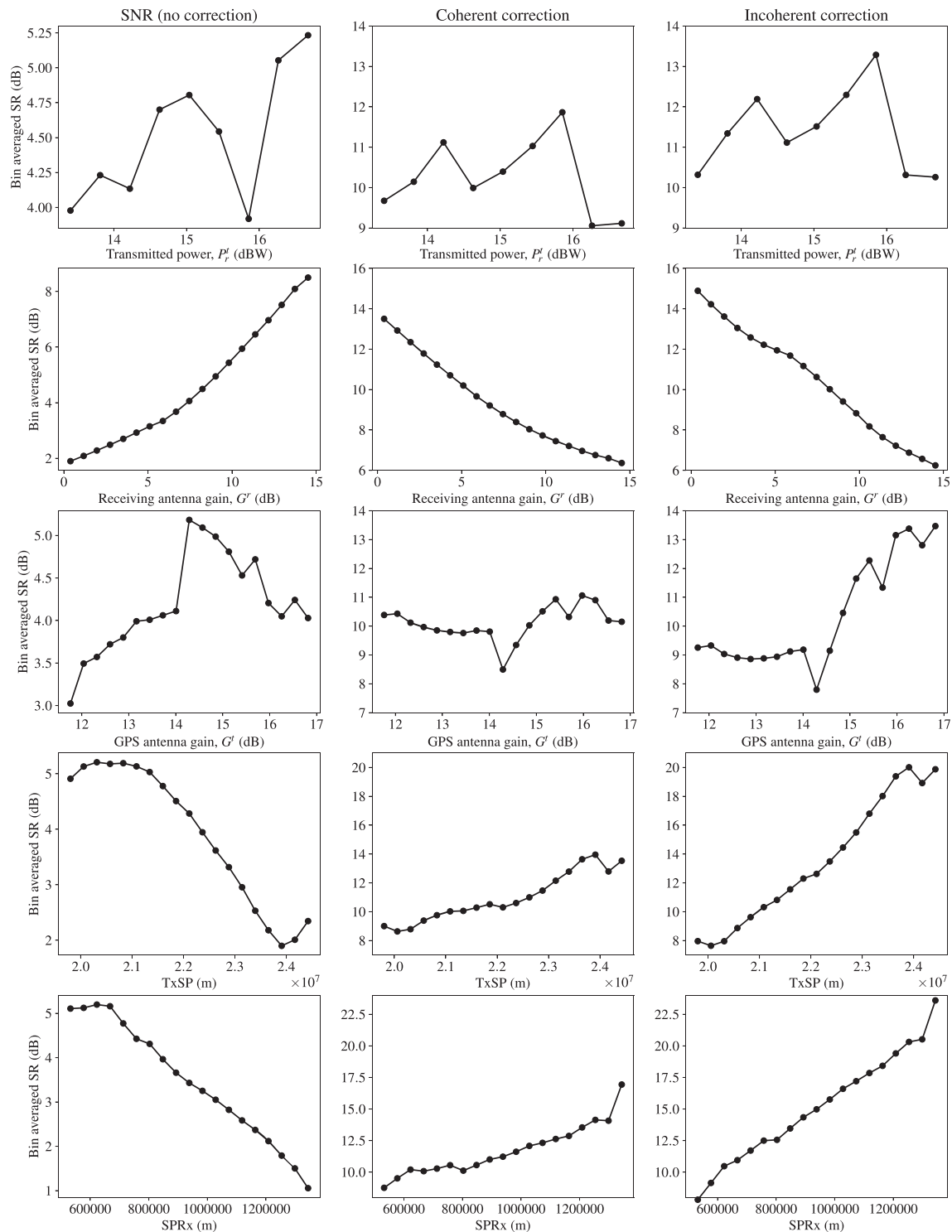
Figure 1. Dependence of the uncorrected SNR data (left column), coherent-correction SR (central column), and diffuse-correction SR (right column) with (from top to bottom): transmitted power, receiving antenna gain, GPS antenna gain, distance between the transmitter and the specular point, and distance between the specular point and the receiver. We find that the coherent correction is better at removing the dependence of the data on these various parameters. Data used here is all CYGNSS land measurements for October through December 2018.

Table 1

Algorithm Parameters Described in Sections 2.2 to 2.5 and the Ranges Tested for Each of Them During the Parameter Estimation Phase Described in Section 2.6

Parameter	Description	Algorithm step	Value range	Value step
T_r	Cluster removal-threshold value	B	10-20	2
C_s	Cluster removal-cluster size	B, D	4-24	2
B_s	STD box size	D	10-150	20
D_s	Segmentation-diffusion parameter	E	0-220	20

Note. See Figure 2 for the corresponding algorithm steps.

SR and removed them. To do so, we first choose a threshold value, T_r, and transform the SR map into a binary map with values over or under T_r. We then use the measurements.label function within the SciPy library (https://www.scipy.org/scipylib/index.html) to identify clusters of high SR values. Based on a binary image where pixels with a SR value above T_r are set to 1 and pixels below T_r are set to 0, the function identifies all the individual clusters in the image and labels each of them using a different integer. It is then easy to count the number of elements in a cluster. Clusters with a number of elements below a value C_s are assigned to NaN.

2.3. NaN Removal

To fill in the NaNs left by the incomplete grid cell filling and the removal of high SR clusters, we use the nearest-neighbor interpolation from the SciPy library to assign values to each grid cell containing a NaN.

2.4. Standard Deviation Map

One of the defining features of waterbodies in the CYGNSS data is that they stand out from the background SR values of their surroundings. We therefore transform our map of SR values into a map of standard deviation (STD) values. To do so, we determine the average SR within a square box of size B_s centered around each grid cell. We then estimate how many STDs above or below the average the grid cell is. Each grid cell therefore gets assigned the value -2, -1, 0, 1, or 2, indicating how many STDs above or below the box average the cell is.

Because this step generates a lot of speckle within the image, that is, new clusters of high STD values, we proceed again to apply the steps described in sections 2.2 and 2.3. When removing high SNR clusters, the threshold value is set to 0 in this case, indicating that any small cluster of values, which are 1 or 2 STD above the average, will be removed. The size of the cluster is kept to C_s.

2.5. Random Walker Segmentation

Finally, we segment the STD map into water and dry land using the `random_walker` segmentation function from the `scikit-image` library for Python (van der Walt et al., 2014). Image segmentation is the identification of multiple units within an image (in our case dry land and water), and many approaches exist to do so, for example, using thresholding or edge detection techniques (Shapiro, 1992). The random walker method was chosen because it is especially good at segmenting noisy images, as is the case here. For this step of the algorithm, two values are chosen: a high threshold value *HT* and a low threshold *LT*. Each pixel at or lower than *LT* will be labeled with an *LT* marker. Similarly, each pixel at or higher than *HT* will be labeled with an *HT* marker. Here we choose *HT* to be 1 and *LT* to be 0. The markers are then allowed to diffuse anisotropically, with diffusion being more difficult across strong gradients. The diffusion strength is set by parameter D_s. Each unlabeled pixel is assigned the label of the marker that reaches it first. Oceans are labeled with a third, unique label. A more detailed description can be found in Grady (2006), and example code can be found on the `scikit-image` page (https://scikit-image.org/).

2.6. Parameter Estimation

In order to estimate the optimal set of parameters (see Table 1), we generate hand-drawn watermasks, which are used as the training sets and that we then compare to our generated watermasks using a range of parameter values for T_r, C_s, B_s, and D_s (Table 1). The training masks are generated at 0.01° ×0.01° resolution using CYGNSS SR images and manually identifying features such as rivers and lakes. We estimate false positive (FPR, pixels wrongly identified as water) and false negative rates (FNR, pixels wrongly identified as dry land). We choose the final set of parameters that yields the lowest

$$E = \sqrt{\mathrm{FPR}^2 + \mathrm{FNR}^2}$$

in order to take into account both overestimation and underestimation of the amount of water in a scene. To estimate the performance of our watermask, we compare it to the most recent MODIS watermask.
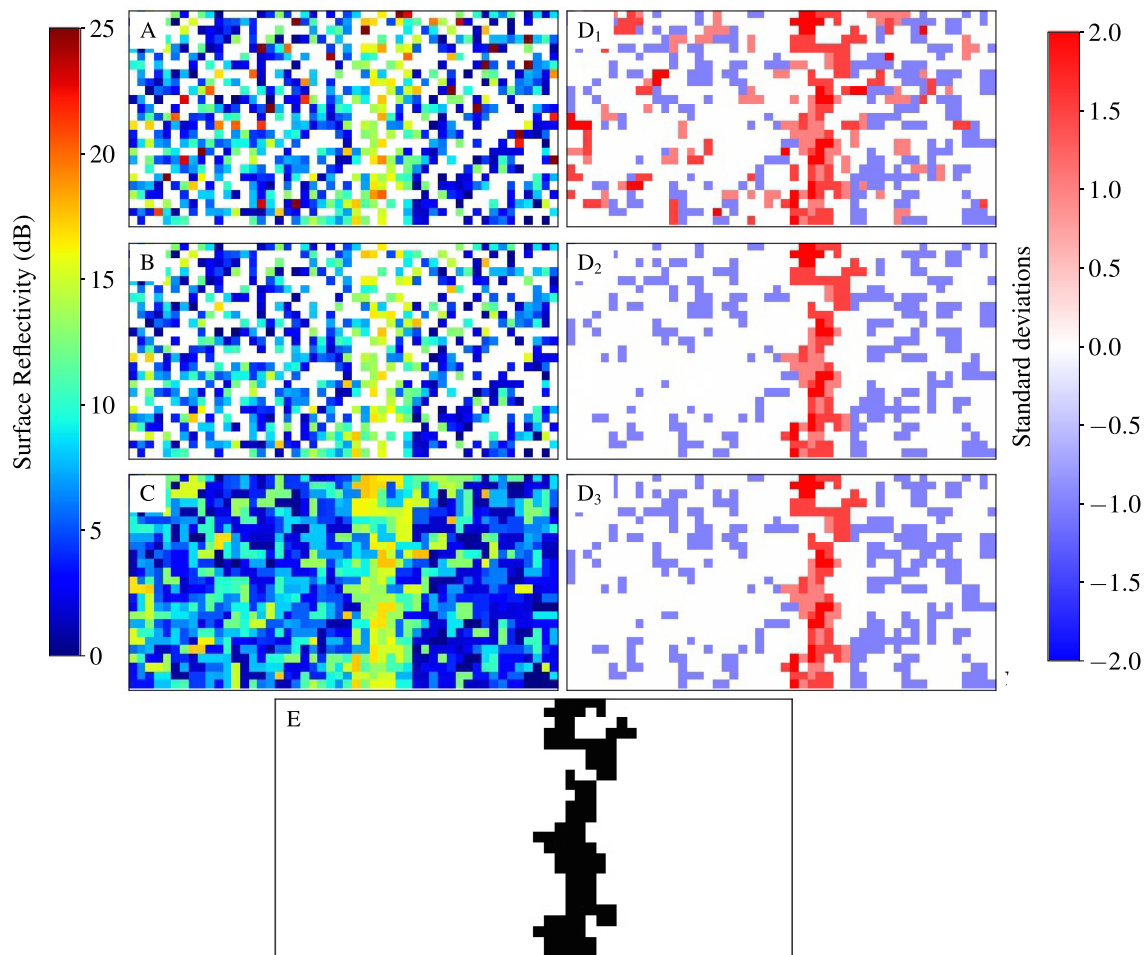
AGU
100
ADVANCING EARTH
AND SPACE SCIENCE

Geophysical Research Letters

10.1029/2019GL085134

Figure 2. Evolution of a map of a section of river in the Congo Basin showing how each stage of the algorithm modifies the initial picture. a) Shows the surface reflectivity SR corrected for coherent scattering. b) After high SR clusters have been removed. c) After NaNs have been filled. d_1) After transforming the SR map into an STD map, removing high value clusters (d_2) and filling in NaNs (d_3). e) Final result of the random walker segmentation performed on d_3, with water showing in black and dry land in white.

3. Results

The intermediate steps of the watermask algorithm for a region in the Congo Basin that includes a river is illustrated in Figure 2. In this example, a river is running from top to bottom on the right half of the image. The background in Figure 2a shows pixels with high SR values intersperse throughout that are effectively eliminated in Figure 2b. Noisy pixels reappear in the background after the image has been filled in (Figure 2c) and transformed into an STD map (Figure 2d_1). After a second round of pixel removal and filling in, the random walker segmentation partitions the image between water and dry land (Figure 2e).

The algorithm also performs well when tested on large areas, such as the Congo or Amazon Basins (Figure 3). Overall, we find that the rate of false positives is very low (0.32%), but smaller tributaries are not captured as well by the algorithm, leading to a higher rate of false negatives (0.67%). For the examples shown here, we found that the best set of parameters is $T_r = 10$, $C_s = 8$, $B_s = 150$, and $D_s = 140$ with $E = 0.75$. This set of parameters was obtained by testing all the parameters for the ranges described in Table 1, comparing them to hand-drawn masks of the Amazon and Congo Basins, and minimizing parameter E as described in section 2.6. These parameters were then used to produce masks in other areas, as shown in Figure 4.
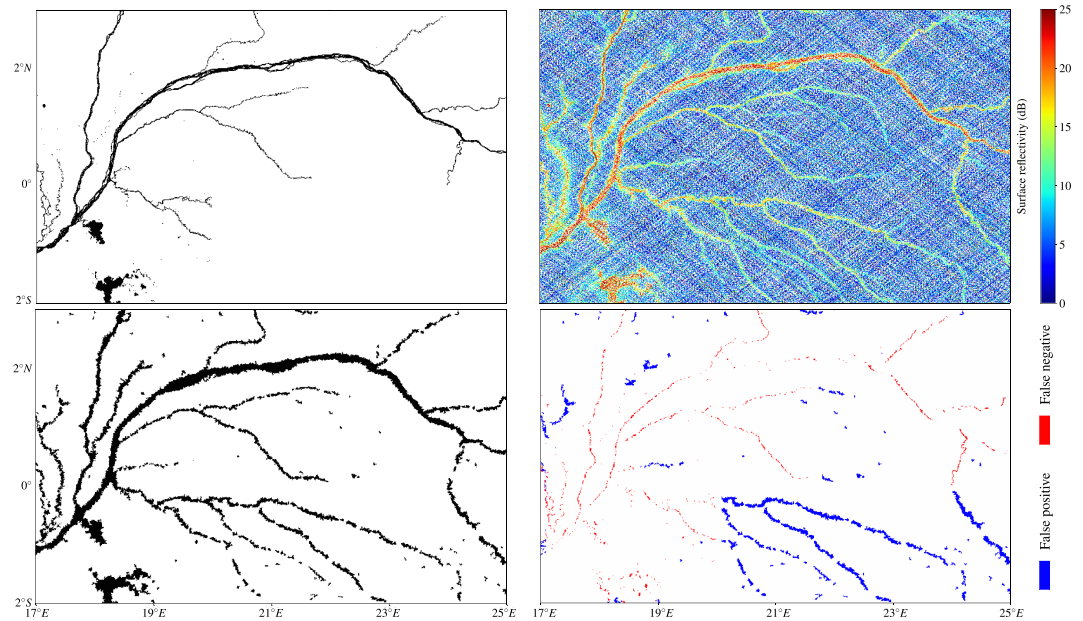
Figure 3. Multiple views of the Congo Basin. Top left: MODIS watermask for the year 2015, with water in black and dry land in white. Top right: corrected CYGNSS SR map. Bottom left: CYGNSS watermask, with water in black and dry land in white. Bottom right: map showing the comparison between the MODIS and the CYGNSS watermasks where pixels identified as water in CYGNSS but not in MODIS (false positive) are shown in blue, and pixels identified as dry land in CYGNSS but not in MODIS(false negative) are shown in red.

4. Discussion

4.1. Comparison to MODIS Watermask Product

One of the main objectives of this work is to demonstrate the ability of the new algorithm to provide watermasks that can be updated on a timely basis using CYGNSS data with its frequent revisit time. Currently, the most commonly used product to mask waterbodies for remote sensing data analysis is the MODIS
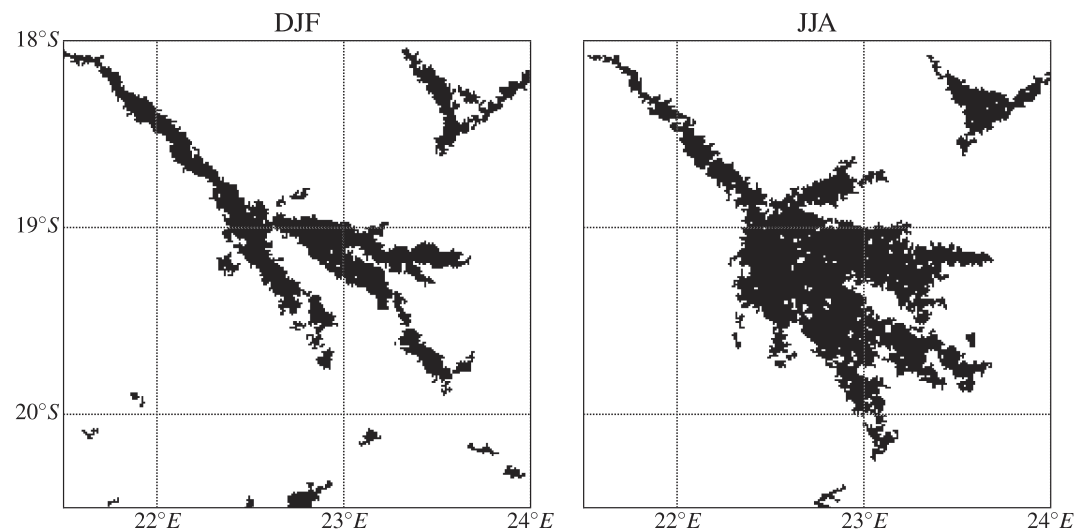

Figure 4. Seasonal watermasks of the Okavango Delta in Botswana. This seasonal delta is dry with rivers running through it during most of the year (December-January-February, left panel) but floods in June-July-August (right panel) after the spring rainfall upstream reaches the delta.

watermask (Carroll et al., 2017). Figure 3 (top left) presents the MODIS watermask for the same region of the Congo Basin considered in the other panels of Figure 3 with the CYGNSS watermask. While the MODIS watermask has finer details (the resolution is 250 m) about the rivers it captures, we find that most of the smaller tributaries clearly identified by our algorithm are missing in the MODIS watermask. Since the latter is based on optical remote sensing, one possible reason is that vegetation is blocking the view to these tributaries, making it impossible for MODIS to "see" them. Thanks to its long wavelength (19 cm), CYGNSS is capable of penetrating even dense vegetation to provide information about standing water and soil moisture beneath thecanopy. Another possible source of divergence is that the MODIS watermask shown here is for the year 2015 (most recent year available in the data set), whereas the CYGNSS mask was generated using data from 2018. A comparison with the Pekel watermask (Pekel et al., 2016; not shown) indicates that it is more detailed than both the MODIS and the CYGNSS watermasks. However, the Pekel watermask format and large file size make it difficult to use at a global scale for remote sensing applications and are more adapted for long-term surface hydrology monitoring.

4.2. Potential Applications

One of the key characteristics of the CYGNSS data is its short return time, with an average revisit time of 3 (median) and 7 hr (mean) for a 25-km grid (Ruf et al., 2018). Hydrologic phenomena happening at short timescales can therefore be monitored using CYGNSS data. For example, the seasonality of the flooding in the Okavango Delta in Botswana (McCarthy et al., 2000) can be seen in the CYGNSS watermask, with the winter map (Figure 4) showing only the main tributaries running through the delta, whereas the summer map displays widespread flooding from upstream spring rainfall. Shorter-lived events, such as floods or drought, can in principle also be monitored using CYGNSS data combined with the present algorithm. However, the CYGNSS track-based sampling system means that reducing the sampling temporal scale implies increasing the spatial scale in order to obtain a sufficient sampling density (Bussy-Virat et al., 2019).

4.3. Future Directions

While the algorithm performs well in natural systems, we should note that it has trouble identifying waterbodies in areaswith a large range of soil water content within a small area, such as a lake surrounded by irrigated croplands in an otherwise arid area with low soil moisture and vegetation. A possible approach to help resolve this issue is the generation of long-term maps of inland waterbodies based on multiple years of data. These maps could then serve as the baseline for shorter timescale "anomaly maps" to look for deviations from the long-term position of the waterbodies. In the case of a lake with irrigated croplands, this method would help identify the actual position of the lake, since the croplands are only irrigated on a seasonal basis and the position of the lake within the cropland could be inferred from the analysis of a longer data set.

The present algorithm can be also applied to data on shorter timescales in order to provide watermasks of events at yearly, seasonal, monthly, and even biweekly timescales. However, additional work is needed to determine the exact spatial scale of each of these products. In particular, it will depend on the algorithm's ability to identify features with a higher number of unsampled pixels. Finally, the method presented here has the potential to be used to develop a watermask product that could be used to mask waterbodies in other remote sensing data sets or monitor temporal changes in a waterbody's position and extent. However, more validation work will be needed to achieve the standards of an official data product.

5. Conclusions

The presented algorithm combines existing computer vision tools to clean and enhance CYGNSS maps of SR in order to identify, map, and resolve changes in the position of inland waterbodies such as lakes and rivers. We find that the algorithm performs well over natural ecosystems and is capable of identifying small tributaries that were missing from the MODIS watermask product, possibly due to the presence of vegetation blocking the view of the river at optical wavelengths, an issue that does not exist for CYGNSS thanks to its long radio wavelength. The algorithm can be used to make maps on short timescales using monthly or seasonal data to monitor short-term hydrologic phenomena such as seasonal flooding.

References

Alsdorf, D. E., Rodríguez, E., & Lettenmaier, D. P. (2007). Measuring surface water from space. *Reviews of Geophysics*, *45*, RG2002. https://doi.org/10.1029/2006RG000197

Bussy-Virat, C. D., Ruf, C. S., & Ridley, A. J. (2019). Relationship between temporal and spatial resolution for a constellation of GNSS-R satellites. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(1), 16–25. https://doi.org/10.1109/JSTARS.2018.2833426

Camps, A. (2019). Spatial resolution in GNSS-R under coherent scattering. *IEEE Geoscience and Remote Sensing Letters*, 1–5. https://doi.org/10.1109/LGRS.2019.2916164

Carroll, M. L., DiMiceli, C. M., Wooten, M. R., Hubbard, A. B., Sohlberg, R. A., & Townshend, J. R. G (2017). MOD44W MODIS/Terra land water mask derived from MODIS and SRTM L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD44W.006

Chew, C., Reager, J. T., & Small, E. (2018). CYGNSS data map flood inundation during the 2017 Atlantic hurricane season. *Scientific Reports*, *8*(1), 429. https://doi.org/10.1038/s41598-018-27673-x

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*, 3317–3324. https://doi.org/10.1002/2016GL068189

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*, 4049–4057. https://doi.org/10.1029/2018GL077905

Geremia-Nievinski, F., Silva, M. F. E., Boniface, K., & Monico, J. F. G. (2016). GPS diffractive reflectometry: Footprint of a coherent radio reflection inferred from the sensitivity kernel of multipath SNR. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(10), 4884–4891. https://doi.org/10.1109/JSTARS.2016.2579599

Grady, L. (2006). Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(11), 1768–1783. https://doi.org/10.1109/TPAMI.2006.233

Leira, M., & Cantonati, M. (2008). Effects of water-level fluctuations on lakes: An annotated bibliography. *Hydrobiologia*, *613*(1), 171–184. https://doi.org/10.1007/s10750-008-9465-2

Li, W., Cardellach, E., Fabra, F., Ribó, S., & Rius, A. (2019). Applications of spaceborne GNSS-R over inland waters and wetlands. Presented at the International Geoscience and Remote Sensing Symposium (IGARSS), IEEE 2019.

Martins, V. S., Novo, E. M. L. M., Lyapustin, A., Aragão, L. E. O. C., Freitas, S. R., & Barbosa, C. C. F. (2018). Seasonal and interannual assessment of cloud cover and atmospheric constituents across the Amazon (2000-2015): Insights for remote sensing and climate analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, *145*, 309–327. https://doi.org/10.1016/j.isprsjprs.2018.05.013

McCarthy, T. S., Cooper, G. R. J., Tyson, P. D., & Ellery, W. N. (2000). Seasonal flooding in the Okavango Delta, Botswana—Recent history and future prospects. *South African Journal of Science*, *96*(1), 25–33.

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584

Ruf, C. S., Atlas, R., Chang, P. S., Clarizia, M. P., Garrison, J. L., Gleason, S., et al. (2016). 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. 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

Shapiro, L. (1992). *Computer vision and image processing*. Cambridge, MA: Academic Press.

Ulaby, F. T., Long, D. G., Blackwell, W. J., Elachi, C., Fung, A. K., Ruf, C. S., et al. (2014). *Microwave radar and radiometric remote sensing*. Ann Arbor: University of Michigan Press.

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., et al., & the scikit-image contributors (2014). Scikit-image: Image processing in Python. *PeerJ*, *2*, 453. https://doi.org/10.7717/peerj.453

Wang, T., Ruf, C. S., Block, B., McKague, D. S., & Gleason, S. (2019). Design and performance of a GPS constellation power monitor system for improved CYGNSS L1B calibration. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(1), 26–36. https://doi.org/10.1109/JSTARS.2018.2867773