

A new algorithm for mapping large inland water bodies using CYGNSS

Maoxiang Chang, Peng Li, Yufeng Hu, Yue Sun, Houjie Wang & Zhenhong Li

To cite this article: Maoxiang Chang, Peng Li, Yufeng Hu, Yue Sun, Houjie Wang & Zhenhong Li (2024) A new algorithm for mapping large inland water bodies using CYGNSS, International Journal of Remote Sensing, 45:5, 1522-1538, DOI: [10.1080/01431161.2024.2314001](https://doi.org/10.1080/01431161.2024.2314001)

To link to this article: <https://doi.org/10.1080/01431161.2024.2314001>



Published online: 15 Feb 2024.



Submit your article to this journal [↗](#)



Article views: 4









View related articles [↗](#)



View Crossmark data [↗](#)



A new algorithm for mapping large inland water bodies using CYGNSS

Maoxiang Chang ^{a,b}, Peng Li ^{a,b}, Yufeng Hu ^c, Yue Sun ^d, Houjie Wang ^{a,b}
and Zhenhong Li ^c

^aInstitute of Estuarine and Coastal Zone, College of Marine Geosciences, Key Laboratory of Submarine Geosciences and Prospecting Technology, Ocean University of China, Qingdao, China; ^bLaboratory of Marine Geology, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China; ^cCollege of Geological Engineering and Geomatics, Chang'an University, Xi'an, China; ^dFrontiers Science Center for Deep Ocean Multispheres and Earth System, Key Laboratory of Physical Oceanography, Institute for Advanced Ocean Studies, Ocean University of China, Qingdao, China

ABSTRACT

The persistence of inland water bodies affects climate, biodiversity, and human societies. Among the multiple remote sensing methods, Global Navigation Satellite System Reflectometry (GNSS-R) technology has shown great potential for inland water bodies mapping at very high temporal resolutions. For large inland water bodies with dimensions larger than 350 km, long fetches would roughen water surfaces, thereby causing low-power GNSS returns, posing challenges for mapping them using GNSS-R data, particularly in the case of ultra-large lakes with complicated shorelines, such as Lake Victoria. In this study, we propose an algorithm for mapping Lake Victoria at a $0.01^\circ \times 0.01^\circ$ spatial resolution using GNSS-R data from the Cyclone Global Navigation Satellite System (CYGNSS). By mainly leveraging the surface reflectivity signal, our algorithm extracts lake boundaries and fills the interior to map Lake Victoria. The probability of detection (POD) of the resultant water mask was approximately 90%. This simple and robust algorithm could enhance the capability of monitoring global fast-changing inland water bodies using GNSS-R data, especially in the pan-tropical areas.

ARTICLE HISTORY

Received 2 October 2023
Accepted 30 January 2024

KEYWORDS

GNSS-R; large inland water bodies; Lake Victoria; CYGNSS

1. Introduction

Inland water bodies are of great importance to human society, ecosystem stability, and the global climate (Holgerson and Raymond 2016). Mapping inland water bodies and their dynamics provides essential services for resource management, food production, and climate assessment (Al-Khaldi, Johnson, Gleason, Chew, et al. 2021; Gerlein-Safdi et al. 2021), which have significant implications for social-ecological system resilience (Piao et al. 2010). However, owing to the constantly changing nature of inland water bodies (Woolway and Christopher 2019), information about their abundance and distribution remains limited across the globe (Feng et al. 2019).

To address the above issues, satellite-remote-sensing methods have been widely applied for mapping inland water bodies over large areas (Huang et al. 2018). Current

remote sensing methods for global inland water bodies mapping can be generally grouped into two categories. For optical sensors, two main watermask datasets are available. The first dataset is the MOD44W V6 product (<https://lpdaac.usgs.gov/products/mod44wv006/>), derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data, provides a yearly global map of surface water at 250 m spatial resolution from 2000 to 2015. The second dataset, the Global Surface Water (GSW) watermask, is a Landsat-derived dataset that maps global inland surface water at 30 m resolution from 1984 to 2021 (Pekel et al. 2016). However, these two optical products cannot map water bodies under clouds or vegetation, which is a significant limitation in tropical areas. In rainy seasons, cloud cover can be 100% (Martins et al. 2018), and dense vegetation cover can also fully conceal the presence of certain water bodies.

Microwave sensors can penetrate clouds and vegetation. However, existing accessible or upcoming data have relatively long repeat intervals, for example, 12 days for the Sentinel-1 mission and 21 days for the Surface Water and Ocean Topography (SWOT) mission, which makes them unsuitable for monitoring fast-changing inland water bodies, especially in tropical areas.

Global Navigation Satellite System Reflectometry (GNSS-R) is an emerging remote sensing technique. Its signals are generated by the forward scattering of existing GNSS signals (L-band) from the Earth's surface, leading to all-weather, all-day, and vegetation-penetrating monitoring with a small file size. Theoretically, to execute a GNSS-R mission, only a low-cost receiving system needs to be launched. With the rapidly increasing density of GNSS signals (e.g. Galileo and BEIDOU) and given sufficient receivers, this bistatic radar configuration concept enables the production of a large number of randomly distributed ground observations with very short revisit intervals (Loria et al. 2020) and at moderate spatial resolutions (Al-Khalidi, Johnson, Gleason, Chew, et al. 2021). Originally designed for tropical cyclone monitoring, the Cyclone Global Navigation Satellite System (CYGNSS) mission was a successful proof of GNSS-R technology (Ruf et al. 2016). This mission has hour-level repeat time and its smallest theoretical spatial resolution is 3.5×0.5 km on land (Chew and Small 2020).

The hour-level revisit time of the CYGNSS data makes it possible to understand the surface hydrological situation on a large scale in a timely manner. Recently, the sensitivity of CYGNSS to inland water bodies has been explored in a more detailed manner (Loria et al. 2020; Zeiger et al. 2022) and large-area CYGNSS water body maps have been produced (Al-Khalidi, Johnson, Gleason, Chew, et al. 2021; Chew, Reager, and Small 2018; Gerlein-Safdi and Ruf 2019; Gerlein-Safdi et al. 2021; Zeiger et al. 2022). Carreno-Luengo et al. (2024) utilized surface reflectivity and power ratio detectors to map inland water bodies with dense vegetation coverage, and verified the coherent signal reflection of calm surface water and the signal attenuation by vegetation canopy. These studies highlight the great potential of dynamic pan-tropical inland water body mapping using CYGNSS data.

However, the existing studies encounter difficulties from data quality instability, insufficient data volume, incomplete understanding of water body scattering characteristics, and lack of proper algorithms, limiting the ability to map extensive inland waters with rough surfaces rapidly and comprehensively. Compared with the specular scattering signals of calm waters, diffuse scattering signals from rough surfaces show a much lower strength. In addition, wind-generated waves can roughen water surfaces on large

inland water bodies larger than 350 km (Al-Khaldi, Johnson, Gleason, Chew, et al. 2021), causing CYGNSS signals from these water bodies appearing more similar to those from land areas.

In this study, we propose a simple and robust algorithm to compensate for the current shortfalls of large-area GNSS-R waterbody mapping. By processing the CYGNSS signal-to-noise ratio (SNR) data, a water mask of Lake Victoria was produced and validated. The advantages and limitations of this algorithm were discussed by comparing it with the published fine-resolution global water data sets.

2. Datasets and methodology

2.1. Study area

Lake Victoria, with a complex shoreline stretching 3,500 km and a surface area of 68,800 km², is the second largest freshwater lake in the world and the largest within the data coverage of CYGNSS (Egessa et al. 2020). Located in the East African Plateau, the lake is shared by three African countries: Uganda, Kenya, and Tanzania (Figure 1). Owing to the proximity of Lake Victoria to the equator, its climate is tropical, with average temperatures ranging from 23°C to 27°C throughout the year. There are two rainy seasons: from March to May and from October to December (Olokotum et al. 2020).

2.2. Datasets

2.2.1. CYGNSS data

Launched in Dec 2016, the CYGNSS mission consists of eight microsatellites distributed in low Earth orbit, covering a pan-tropical area ($\pm 38^\circ$ latitude). The Delay Doppler Mapping Instrument (DDMI) onboard each microsatellite tracks 4 distinct GNSS reflections simultaneously. Currently, the constellation is capable of producing up to 32 observations at a 2 Hz reporting interval (0.5 s). The exact revisit cycle is difficult to calculate because of the pseudo-random sampling of the CYGNSS constellation (Bussy-Virat, Ruf, and Ridley 2019). The mean and median revisit interval were 2.8 h and 7.2 h, respectively. The spatial resolution was not related to the antenna size. Instead, the surface roughness matters.

Currently, mapping inland water surfaces using CYGNSS surface reflectivity data assumes that the reflected signal is dominated by coherent reflection. This assumption was observed and tested in previous studies (Gerlein-Safdi and Ruf 2019; Loria et al. 2023; Morris et al. 2019; Nghiem et al. 2017; Zhang et al. 2021). The influence of incoherent components can be marginal (Dong and Jin 2021). Coherent reflections arising from smooth surfaces and coherence refer to a scattering pattern in which the majority of returns inside one observation footprint have the same phase shifts and therefore add coherently (Al-Khaldi, Johnson, Gleason, Loria, et al. 2021; Loria et al. 2023). In such cases, each footprint has a theoretical spatial resolution of 3.5×0.5 km (Chew and Small 2020). However, small fractions of water could dominate the reflection in the glistening zone (Zeiger et al. 2022), expecting a resolution of 600 m to 1 km in inland water mapping (Al-Khaldi, Johnson, Gleason, Chew, et al. 2021).

In this study, we used version 3.1 of CYGNSS level 1 data, spanning from Aug. 2021 to Jul. 2022. The two measurements used were the Delay Doppler Maps (DDM) and SNR. The

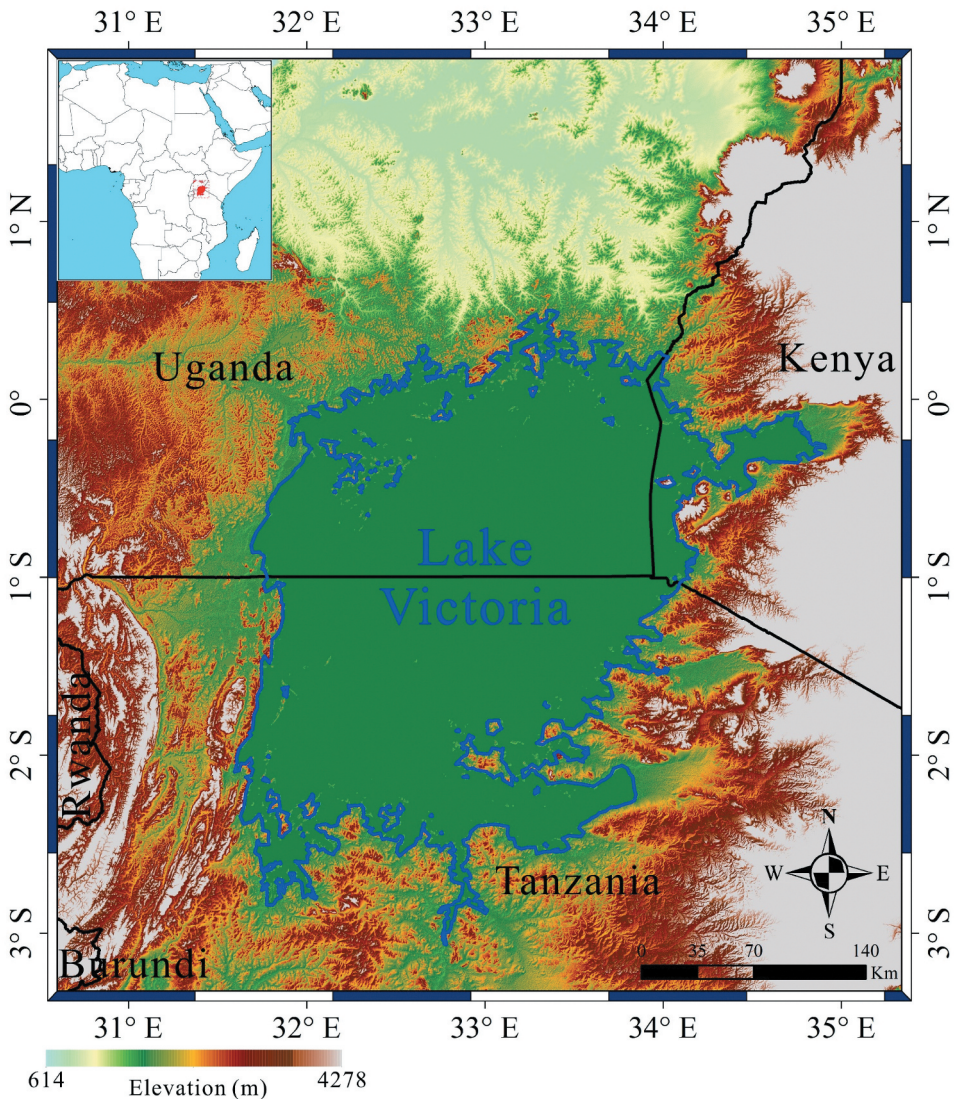


Figure 1. Relief map around Lake Victoria. The relief map is from SRTM (shuttle radar topography mission) digital elevation datasets (Farr et al. 2007). The blue line denotes the boundary of Lake Victoria, derived from global surface water (GSW) datasets (Pekel et al. 2016).

Level 1 DDM comprises 17 delay bins and 11 Doppler bins. The delay is the time difference between the arrival of the direct GPS signal and the signal scattered by the ground surface. The Doppler shift arises owing to the relative motions of the satellites and ground surface. The ratio of the maximum value to the noise average within a signal raw counts DDM is the SNR.

2.2.2. Other datasets

GSW data document global surface water location and seasonality with a resolution of 30 m (Pekel et al. 2016). Since 1984, more than three million Landsat images have been used

to measure long-term changes in global surface water. We used the latest yearly water classification map from 2021 archived in the Google Earth Engine (GEE) for this study. We categorized both the ‘Seasonal water’ and ‘Permanent water’ classes as ‘Water’, while ‘Not water’ were classified as ‘Land’.

GlobalLand30 (GL30) is a 30-metre spatial resolution global land cover data developed by the National Geomatics Centre of China (Jun, Ban, and Songnian 2014). The latest GL30 is the version 2020, extracted from Landsat, Chinese HJ-1, and Chinese GF-1 satellite images.

The European Space Agency (ESA) WorldCover 2020 product (ESA10) leverages images from Sentinel-1 and Sentinel-2 for near real-time development and validation. It classifies global land cover in 2020 at 10 m resolution (doi:10.5281/zenodo.5571936).

2.3. Methodology

2.3.1. DDM power-spread detector method and SNR correction

DDM power-spread detector (DPSD) method detects the status of power ‘spread’ within the DDM to identify inland water bodies (Al-Khaldi, Johnson, Gleason, Chew, et al. 2021). The smooth surfaces of calm inland water give rise to mirror-like reflections. The power is concentrated around the specular point, resulting in the DDM power-spread pattern shown in Figure 2(a). By contrast, rough surfaces diffuse signals, resulting in an apparent spread of power beyond the specular point in the DDM (Figure 2(b)). The DPSD method compares the power within a limited space neighbouring the specular bin to the power outside this region, which is developed as:

$$PR = \frac{C_{in}}{C_{out}} \quad (1)$$

$$C_{in} = \sum_{i=-1}^1 \sum_{j=-2}^2 DDM(\tau_M + i, f_M + j) \quad (2)$$

$$C_{out} = \sum_{i=1}^{N_r} \sum_{j=1}^{N_f} DDM(i, j) - C_{in} \quad (3)$$

where PR is the power ratio, C_{in} is the total power within the 3×5 delay-Doppler area around the specular bin, and C_{out} is the complete power within the DDM

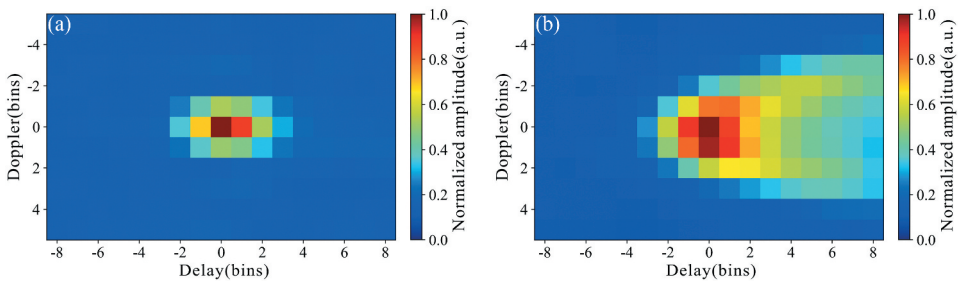


Figure 2. Illustration of different DDM power spread patterns. (a) DDM of smooth surfaces, and (b) DDM of rough surfaces.

excluding C_{in} . N_r and N_f are the numbers of delay bins and Doppler bins, respectively. τ_M and f_M are the delay Doppler positions of the specular bin (maximum power).

A higher PR indicates a more concentrated power distribution within the DDM, which indicates a higher possibility of inland surface water existence. After thermal noise removal, a threshold of $PR \geq 2$ was selected to identify calm inland waters (Al-Khaldi, Johnson, Gleason, Loria, et al. 2021).

As previously illustrated, when analysing the inland surface reflectivity from GNSS-R, coherent surface reflections are assumed. The peak scattered power of the coherent component (within a DDM) is

$$P_{DDM} = \frac{P_t G_t}{4\pi(R_r + R_t)^2} \frac{G_r \lambda^2}{4\pi} \Gamma \quad (4)$$

Where P_t is the transmitted power, G_t is the transmitting antenna gain, R_r is the distance between the receiver and specular point, R_t is the distance between the transmitter and specular point, G_r is the receiver antenna gain, λ is the wavelength of the GPS signal (0.1903 m), and Γ is the surface reflectivity. Because P_t and G_t are relatively stable over a brief period, we simplify the function as:

$$P_{DDM} \propto \frac{G_r \Gamma}{(R_r + R_t)^2} \quad (5)$$

The SNR is the peak power of the DDM normalized to a noise floor, which is defined as the mean cross-correlation before the leading edge of the reflection within the DDM:

$$SNR_{dB} = 10 \log_{10} \frac{P_{DDM}}{P_{noise}} \quad (6)$$

Where P_{noise} is the noise power within a DDM and SNR_{dB} is the SNR in dB. The relationship between the SNR and surface reflectivity can then be derived as follows:

$$SNR_{dB} \propto 10 \log_{10} \Gamma \quad (7)$$

The SNR is affected by various factors. As previously illustrated (Chew et al. 2017; Morris et al. 2019), we used the following expression to correct the SNR.

$$SNR_c = SNR_{dB} - G_r + 20 \log_{10}(R_r + R_t) \quad (8)$$

Where SNR_c is the corrected SNR, and G_r is the receiver antenna gain in dB. It is clear that SNR_c and surface reflectivity are not equal in quantity but are rather proportional to each other. The proportional factor remains subject to variability related to surface roughness, vegetation, instrument calibration, and others. The elimination and quantification of these uncertainties will be addressed in future research.

In this study, we do not concentrate on the absolute value of SNR_c , but on the relative changes. In general, compared with the DPSD method, the SNR_c is more sensitive to inland surface water (soil moisture). In addition, an increase in the SNR_c is expected under wetter ground conditions (Morris et al. 2019). These are the theoretical bases for inland water bodies mapping with the CYGNSS SNR_c data.

2.3.2. CYGNSS data preprocessing

First, the per-DDM quality flag, which is in a binary format, indicates various poor-quality conditions of the CYGNSS data. A flag combination is applied, that is, 's-band powered up', 'large spacecraft attitude error', 'black-body ddm', 'ddm is test pattern', and 'low-confidence gps eirp estimate'. Data flagged within this combination were removed. This combination has been tested previously and ensures correct data filtering with the preservation of most samples possible (Zeiger et al. 2022).

Second, a DDM with a low SNR is dominantly affected by noise. The SNR exclusion threshold was set to 2 dB as described in previous studies (Al-Khaldi, Johnson, Gleason, Loria, et al. 2021; Dong and Jin 2021).

Unevenly distributed CYGNSS data must be gridded to ensure a consistent reference. As the 1 year/1 km compromise of CYGNSS data has been tested and suggested previously (Al-Khaldi, Johnson, Gleason, Chew, et al. 2021), the one-year CYGNSS data were gridded at 0.01° spatial resolution (~ 1 km) in this study. For the PR map, the value of one grid is the 80th percentile of all samples falling within the same grid (Al-Khaldi, Johnson, Gleason, Chew, et al. 2021). For the SNRc map, to attenuate the influence of noise, we used the median of all samples within that grid.

2.3.3. Thresholding, noise removing, and void filling

Previous studies on surface water dynamics with CYGNSS SNR have mainly focused on the use of a suitable threshold (Chew, Reager, and Small 2018; Morris et al. 2019; Wan et al. 2019). A correct threshold can effectively separate targeted information from the background. Owing to factors such as small-scale roughness, large-scale topography, land cover, and permittivity (Al-Khaldi, Johnson, Gleason, Loria, et al. 2021), land surfaces produce significant diffuse scattering, resulting in a much weaker signal strength compared with calm water surfaces. In a large inland water body, such as Lake Victoria, long fetches and lake swells can also roughen the water surfaces, leading to a decrease in surface reflectivity (Chew and Small 2020). It is evident from Figure 3(c) that a significant number of pixels within Lake Victoria are confused with land SNRc values. Without further processing, direct threshold segmentation can introduce large mapping errors.

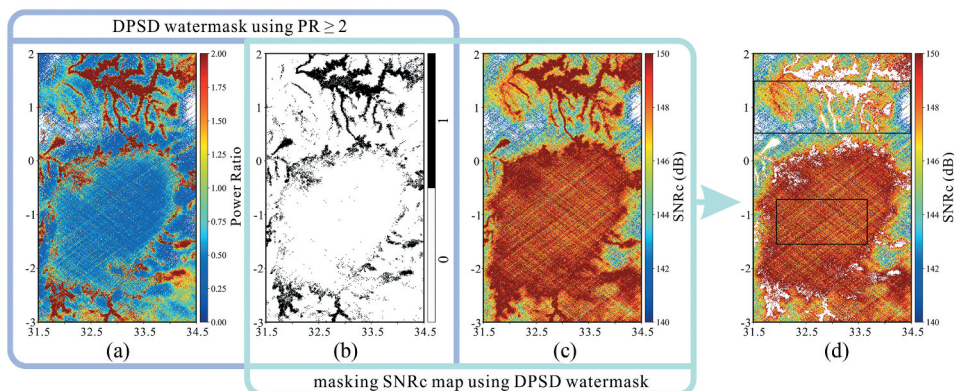


Figure 3. Processes of SNRc sampling in the study area. (a) 80th quantile of PR over the one-year period, (b) binary classification of (a) using threshold $PR \geq 2$, (c) median SNRc composite map over the one-year period of the study area, (d) SNRc map masked by (b). Rectangles in (d) denote two sampling regions that represent land and Lake Victoria.

We first considered sampling Lake Victoria and land areas on SNRc maps to better understand their characteristics. To facilitate the process of land and Lake Victoria sampling, we masked other calm inland water bodies using $PR \geq 2$ (Figure 3(a,b)).

The use of PR in this water-masking step, instead of other water body products, has certain advantages. Both the DPSD and SNRc methods are derived from CYGNSS data, which means that they inherit the same characteristics of L-band forward scattering signals. New error sources will not be introduced in the masking process, and calm water bodies under the vegetation cover will also be masked. In addition, because of the DPSD method's sensitivity to surface roughness, watermasking using this method could mask out calm inland water bodies and retain the majority of Lake Victoria unmasked at the same time, making it convenient for subsequent large-scale sampling.

Here, the masked SNRc map can be easily sampled without paying significant attention to avoid inland calm water bodies. Two areas were sampled from the masked SNRc map (Figure 3(d)). One stands for Lake Victoria (0.74°S, 1.56°S, 31.9°E, 33.63°E), and the other represents the land area (1.4°N, 0.5°N, 31.5°E, 34.5°E).

Figure 4 shows a statistical histogram of the SNRc values in the two sampled regions. Although we masked out calm water surfaces, SNRc sampling values of land and Lake Victoria still overlapped with each other to an apparent extent. This was mainly caused by the mix of relatively low SNRc values of rough surfaces within Lake Victoria and relatively high SNRc values of wet soil on land. To address this problem, a threshold should be chosen carefully.

Because of the opportunistic nature and calibration uncertainties associated with CYGNSS data (Gleason et al. 2019), SNRc noises unrelated to surface properties could

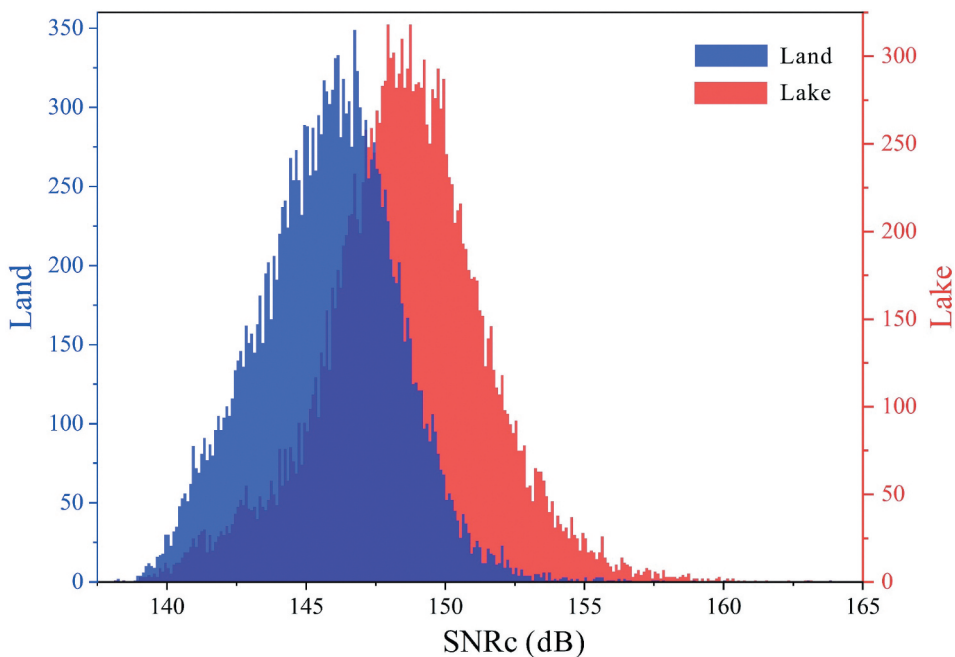


Figure 4. Statistical histogram of SNRc values in two sampled regions. Blue bars indicate SNRc values of land area, while red bars indicate SNRc values of Lake Victoria.

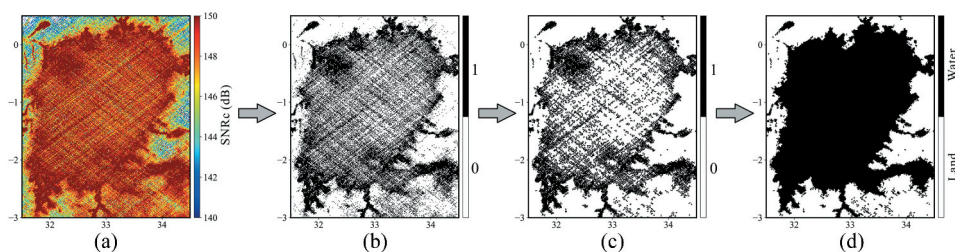


Figure 5. Flowchart for extracting Lake Victoria based on CYGNSS SNRc map. (a) median SNRc composite map over the one-year period, (b) binary classification of SNRc map using the threshold SNRc = 150 dB, (c) results of removing noises (clusters ≤ 3) in the after-thresholding SNRc map, and (d) final mapping result of Lake Victoria after void filling.

arise. These noises tend to be small clusters of high SNRc and are track-based (Gerlein-Safdi and Ruf 2019). To remove this noise, clusters smaller than a certain cluster size should be eliminated in the after-thresholding SNRc map.

After noise removal, substantial pixels in the inner Lake Victoria were masked, whereas the rims were barely affected. This is because the water around the rims of large lakes is sheltered by harbours or natural geography and usually has calm surfaces. Compared to rough water surfaces, the SNRc of calm water surfaces is steadier and more reliable. Therefore, this noise-removing step excludes unwanted noise and simultaneously maintains the outline of Lake Victoria. Finally, we filled the holes in the after-noise-removing binary results to derive the CYGNSS Lake Victoria map.

To determine the optimal threshold and cluster size, 126 possible scenarios are tested. These scenarios comprised 11 different thresholds (145 dB–155 dB, 0.5 dB step) and six different cluster sizes (0 to 5). In these scenarios, the SNRc map was first binarily segmented using a test threshold. Subsequently, clusters smaller than the test cluster size were removed, and the holes in the resultant binary map were filled to obtain the final CYGNSS Lake Victoria map. Taking threshold = 150 dB and cluster size ≤ 3 as an example, the workflow mentioned above is shown in Figure 5.

To verify the accuracy, we compared the 126 CYGNSS watermarks with GSW and the water body layer of the GL30 data. The GSW and GL30 data were gridded to $0.01^\circ \times 0.01^\circ$ using the nearest-neighbour algorithm. The verification process assumes that the two datasets reflect the true ground status. Based on this assumption, two sets of probability of detection (POD) were calculated (Figure 6). The POD (Wilks 2011) is defined as follows, which represents the probability of the CYGNSS-based water surface maps identified by the Landsat or Sentinel water mask.

$$\text{POD} = \frac{\text{hits}}{\text{hits} + \text{misses}} \quad (9)$$

3. Results

The algorithm overcomes the interference of rough lake water surfaces to CYGNSS signals and successfully maps the water bodies of Lake Victoria. Because the highest POD is achieved at a threshold of 150 dB and a cluster size of 3 (Figure 6), this combination was

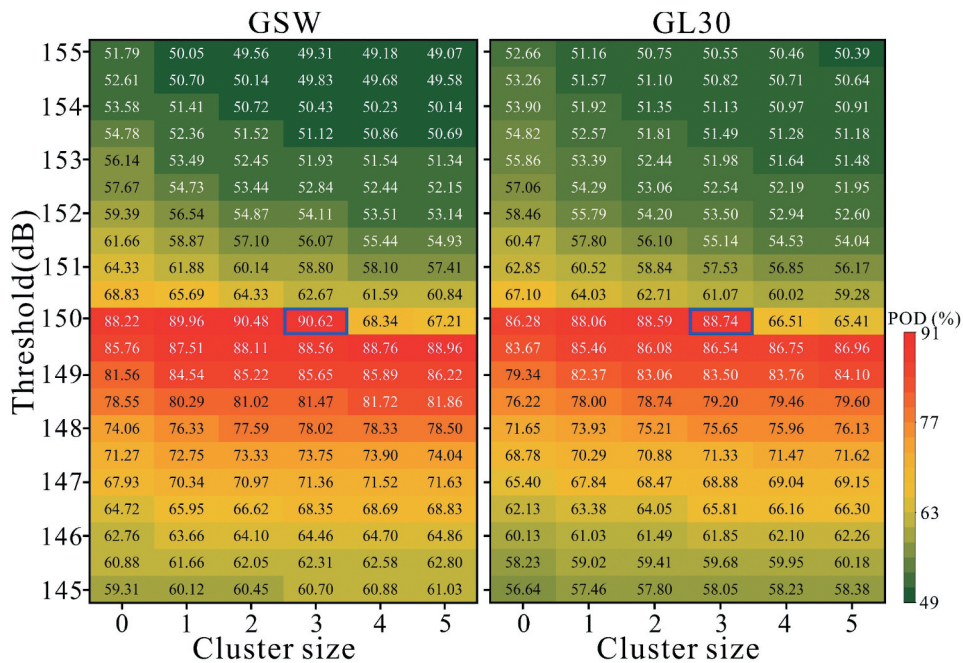


Figure 6. POD of CYGNSS watermask using different segmentation thresholds and noise-removing cluster sizes. The left and right subplots are CYGNSS watermask POD based on GSW data and GL30 data, respectively. The blue rectangles denote the highest POD in the subplots.

applied in our workflow of CYGNSS Lake Victoria mapping. The results are shown in Figure 7. Compared with other optical or microwave sensor results (Figure 7(b-d)), the CYGNSS lake map (Figure 7(a)) reasonably delineated the shape of Lake Victoria. However, visual examination indicates that the lake map produced by our algorithm using CYGNSS data tends to overestimate the extent of water surface coverage. Detailed information on these discrepancies is discussed in Section 3.

Assuming GSW as the ground truth, the POD around Lake Victoria (0.5°N, 3°S, 31.5°E, 34.5°E) was 90.6%. The ‘misses’ of CYGNSS results are mainly false positives (pixels identified as water in CYGNSS but not in GSW). The false positive rate (FPR) reaches 8.5%, while the false negative rate (FNR) for pixels identified as water in GSW but not in CYGNSS accounts for only 0.9%. However, when the CYGNSS map was compared to the GL30 water body layer, discrepancies were observed. Using GL30 as the standard, the POD varies to 88.7%, and the FPR increases to 10.8%. The FNR still constitutes a small fraction (0.5%). In the case of the ESA10 product, the POD reaches 88.8%, with the FPR accounting for 10.8% and the FNR for 0.4%.

4. Discussion

4.1. Uncertainties and limitations

Figure 8(b) shows the false alarm pattern of the CYGNSS watermask compared with the GL30 water body layer. The false positive pixels are mainly located on the rims of Lake

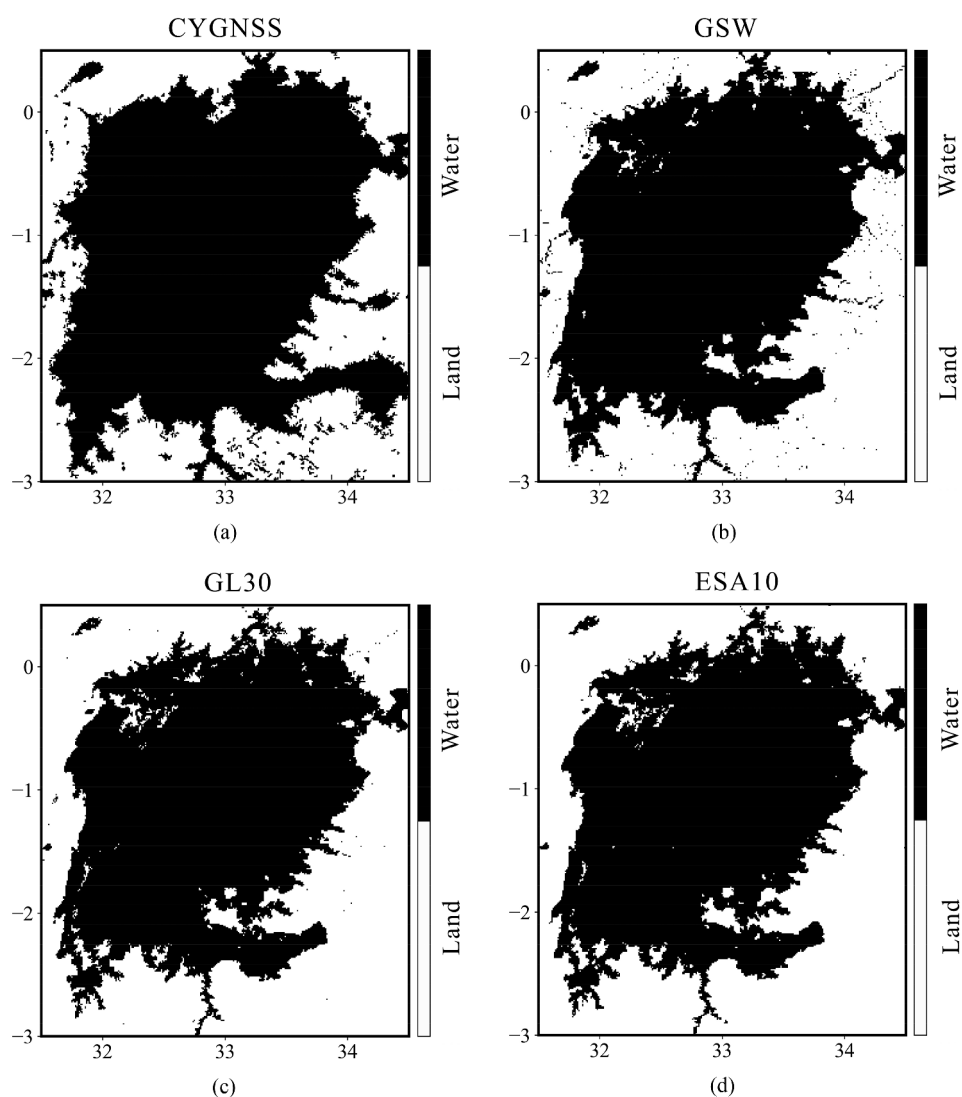


Figure 7. Different water body maps of Lake Victoria. (a) CYGNSS, (b) GSW 2021 yearly water classification map (Pekel et al. 2016), (c) water body layer of GL30 map in 2020 (Jun, Ban, and Songnian 2014), and (d) permanent water body layer of ESA WorldCover 10 m 2020 product. All the maps were resampled into the resolution of $0.01^\circ \times 0.01^\circ$ using the nearest neighbour algorithm.

Victoria, expanding their range. Islands inside Lake Victoria were mostly misidentified as water surfaces in the CYGNSS water body map. This was predominantly caused by the intrinsic characteristics of our algorithm. In this algorithm, we eliminated a considerable portion of the pixels using a strict threshold, leading to the emergence of lake rims. The pixels inside the rims were then filled and classified as water surfaces. As these lake islands are inside the lake rims, they are unavoidably included as water surfaces during the filling process.

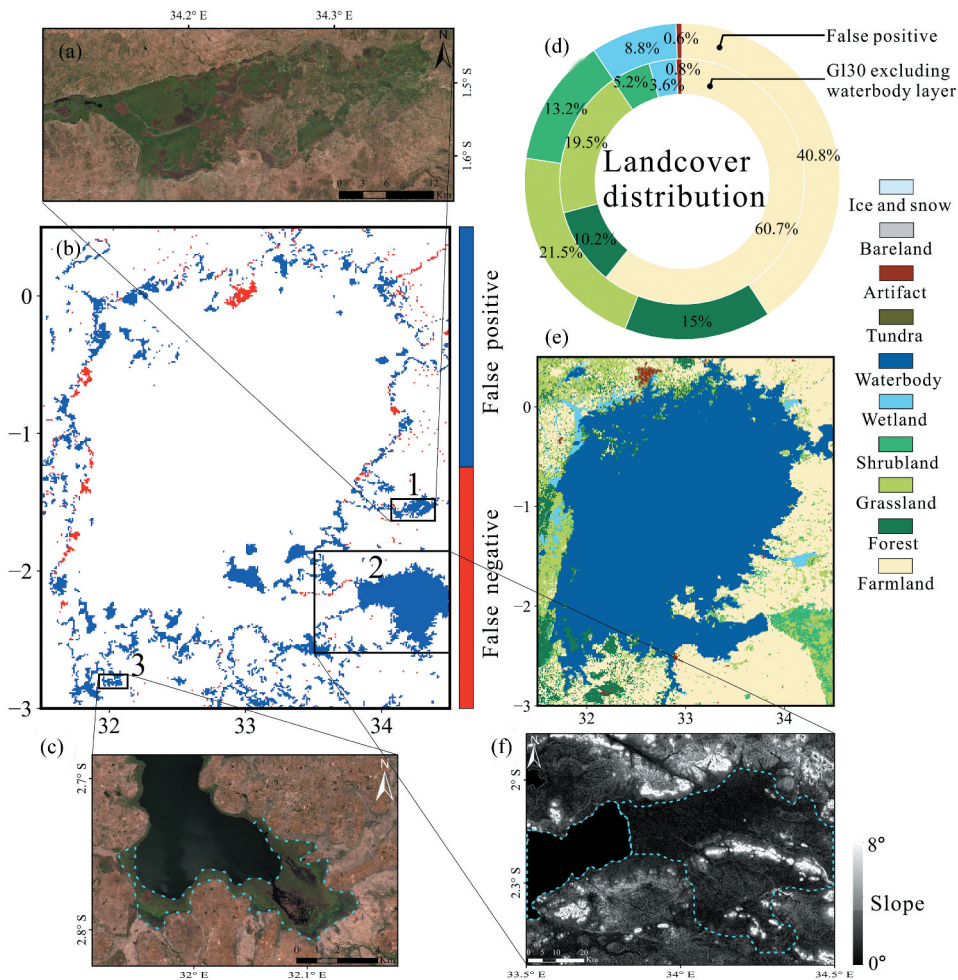


Figure 8. Details of CYGNSS water body map false alarms based on GL30 2020 data. (a) Sentinel-2 imagery of area 1, acquired on Aug 29, 2021. (b) Spatial distribution of false alarms. (c) Sentinel-2 imagery of area 3, acquired on Aug 2, 2021. (d) Landcover distribution related to GL30 2020. The inner circle indicates the landcover type constitution of GL30 2020 except for the “waterbody” class around Lake Victoria, while the outer layer is the distribution of landcovers where the false alarm pixels locate. (e) GL30 2020 data gridded into $0.01^\circ \times 0.01^\circ$. (f) Slope relief map of area 2 derived from SRTM digital elevation datasets. (d) and (e) share the same legend. Dotted lines in (c) and (f) denote the extent of wetlands and a small-slope region, respectively.

It was found that very small clustered, isolated false alarms were mainly false negatives (Figure 8(b)), and these false negatives tended to be distributed away from Lake Victoria. These pixels are small inland water bodies that reflect strong SNRc signals. However, the noise removal step in our algorithm has difficulty distinguishing real noise clusters from small water body clusters, resulting in the absence of very small water bodies in the CYGNSS water body map. These false negative pixels can then be produced.

Compared with the land cover composition of GL30, the share of the natural vegetated land cover types (forest, shrubland, and grassland) in false positive pixels increased

(Figure 8(d)). It is suggested that a certain portion of saturated soil or water under vegetation cover was also detected as water surfaces in this study. An example of this is shown in (Figure 8(c)). The water surface or saturated soils blocked by aquatic vegetation around the lake rims could not be detected by the optical sensors utilized in GL30, but they were identified by CYGNSS signals. This produces false positive pixels in this region. Similar conditions were also observed in wetland areas (Figure 8(a)).

It is notable that false alarms were concentrated around area 2 (Figure 8(b)). Vegetation, including forests, grasses, and shrubs, is widespread in this area (Figure 8(e)). In the equatorial climate zone, these vegetated areas have the potential to retain sufficient soil moisture to reflect strong GNSS-R signals. However, the shape of the false alarm ensemble in the area 2 did not strictly follow the shape of the composition of forest, grass, and shrub (Figure 8(e)). Meanwhile, we found that there exists a small-slope (smooth) region in the area 2 (except for a ridge in the middle). We roughly delineate the shape in Figure 8(f). Intriguingly, the outline of this smooth region resembles that of the false alarm ensemble. We suppose that the false alarm ensemble in the area 2 results from the combined effect of low surface roughness and high soil moisture. The quantification of the contribution of the two uncertainties requires further examination.

4.2. Insights from the data deliverable

First, the $0.01^\circ \times 0.01^\circ$ grid of the CYGNSS data is a trade-off between spatial and temporal coverage. Unlike other optical remote sensing sensors, the spatial resolution of CYGNSS is a function of the surface roughness (Chew and Small 2020). Signals scattered from significantly rough surfaces, characterized by a root mean square surface height deviation larger than 5–7 cm (Balakhder, Al-Khaldi, and Johnson 2019), can originate from a much wider area compared to smooth surfaces. Furthermore, the track-based, pseudo-random method of the spatial coverage of CYGNSS introduces additional complexity. Therefore, the intrinsic differences between the GL30 and CYGNSS data could not be eliminated.

Second, the CYGNSS observations are sensitive to calm water. The contribution of a small fraction of water to the footprint may dominate the scattering signals from the surface (Zeiger et al. 2022). Therefore, mixed pixels may also be identified as water, which can lead to a range exaggeration.

Third, the L-band CYGNSS signal (19 cm wavelength) is capable of penetrating dense vegetation. SNR_c is directly related to the surface reflectivity (Morris et al. 2019), and the surface reflectivity is closely related to the surface dielectric constant. Saturated soils (bare or under vegetation canopies) may not be easily distinguished from the water surface using the SNR_c.

This study demonstrates the ability of GNSS-R signals to map large inland water bodies with rough surfaces, filling a gap in this field. Our algorithm improved the ability of fast global water body mapping using GNSS-R signals. However, it should be noted that the 150 dB threshold we set may be case dependent. Different thresholds would result in different estimates of water body area. The threshold can vary depending on the environment. Tests and experiments are necessary if this algorithm is used to produce operational pan-tropical or global water body products. With more CYGNSS-like constellations being launched in the future, a finer, timelier, and more complete hydro-earth will be presented to the scientific community.

4.3. Implications for future work

Currently, water mapping based on GNSS-R technology is still in its early stages and requires improvement to catch up with other remote sensing products at 10 m or 30 m resolutions. Figure 7 showed that false positives dominate the differences between water body maps from GNSS-R and other fine-resolution remote sensing missions. However, when compared with satellite missions like Landsat or Sentinel constellation, the advantages and potentials of GNSS-R are still great.

Fine-spatial-resolution remote sensing images are often influenced by the similarity of water and shadow in spectral index, and the index of small water body changes due to land disturbance (Carreno-Luengo et al. 2024). Besides, the availability of optical sensors is greatly hindered by night and clouds (Chang, Li, Li, et al. 2022; Chang, Li, Sun, et al. 2022). Persistent cloud cover often impedes the timely use of optical sensors in flood and hazard monitoring. For microwave sensors, such as the C-band 10 m Sentinel-1 mission, they can generate water masks in all weather conditions and throughout the day (Liu et al. 2022; Tu et al. 2021; Zhu et al. 2021). However, for optical or radar sensors, a revisit time of a few to a dozen days is relatively long for monitoring short-term events such as flooding.

In contrast, water mapping with GNSS-R technology can easily handle issues related to clouds, night-time, and coarse temporal resolution. Furthermore, the L-band GNSS-R signals have shown great applicability in mapping water bodies under heavy biomass up to 400 ton/ha in complicated heterogeneous environments, e.g. rainforests in the Amazon basin and Pantanal wetlands (Carreno-Luengo et al. 2024). In addition, the GNSS-R technique uses transmitted signals originally intended for positioning. This 'free' signal source significantly reduces the cost. With the continuous development of GNSS systems, the spatiotemporal coverage and resolution of GNSS-R water bodies mapping can be improved to a large extent. The emergence of new researches and understandings will help compensate for the existing shortcomings.

5. Conclusion

Taking advantage of the CYGNSS data, we proposed an algorithm that extracts the boundaries of Lake Victoria and fills the inside, achieving the mapping of large inland water bodies with rough surfaces. The resultant water body map of Lake Victoria showed good performance when compared to the GSW, GL30 and ESA10 data. We found that false alarms mainly arise from the unique scattering characteristics of the CYGNSS signals.

Although our algorithm requires improvements in terms of accuracy, particularly in distinguishing inner lake islands, we have made an important step towards inland rough surface water body mapping using CYGNSS data, facilitating complete and quick-response large-scale water body monitoring through GNSS-R techniques.

Acknowledgement

This study was supported by the Natural Science Foundation of China (No. 42041005-4). The CYGNSS data are freely available at <https://cygnss.engin.umich.edu/data-products/>. We thank the

Google Earth Engine (GEE) platform, which provides publicly accessible data and computing power. The authors are grateful to the anonymous reviewers and the editor for their constructive and excellent reviews, which greatly improved the quality of the article.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Natural Science Foundation of China [42041005-4].

ORCID

Maoxiang Chang  <http://orcid.org/0009-0001-9699-0234>

Peng Li  <http://orcid.org/0000-0002-0707-4999>

Yufeng Hu  <http://orcid.org/0000-0001-9097-9010>

Yue Sun  <http://orcid.org/0009-0004-7678-5745>

Houjie Wang  <http://orcid.org/0000-0001-6409-0592>

Zhenhong Li  <http://orcid.org/0000-0002-8054-7449>

Credit author statement

Maoxiang Chang, and Peng Li: Writing, Investigation; **Maoxiang Chang, Peng Li, Zhenhong Li:** Conceptualization, Supervision; **Peng Li, Zhenhong Li, and Maoxiang Chang:** Resources; **Peng Li, Zhenhong Li, and Maoxiang Chang:** Software; **Maoxiang Chang, Yue Sun, and Peng Li:** Formal analysis, Data Curation; **Maoxiang Chang, Peng Li, and Zhenhong Li:** Methodology; **Maoxiang Chang, Peng Li, Yufeng Hu, Yue Sun, Zhenhong Li, and Houjie Wang:** Review, Validation.

Data availability statement

The example code for CYGNSS data processing and large inland water bodies mapping algorithms are openly available through Zenodo (<https://zenodo.org/record/7832537>). Other materials related to this study are available from the corresponding author upon reasonable request.

References

- Al-Khaldi, M. M., J. T. Johnson, S. Gleason, C. C. Chew, C. Gerlein-Safdi, R. Shah, and C. Zuffada. 2021. "Inland Water Body Mapping Using CYGNSS Coherence Detection." *IEEE Transactions on Geoscience & Remote Sensing* 59 (9): 7385–7394. <https://doi.org/10.1109/TGRS.2020.3047075>.
- Al-Khaldi, M. M., J. T. Johnson, S. Gleason, E. Loria, A. J. O'Brien, and Y. Yi. 2021. "An Algorithm for Detecting Coherence in Cyclone Global Navigation Satellite System Mission Level-1 Delay-Doppler Maps." *IEEE Transactions on Geoscience & Remote Sensing* 59 (5): 4454–4463. <https://doi.org/10.1109/TGRS.2020.3009784>.
- Balakhder, A. M., M. M. Al-Khaldi, and J. T. Johnson. 2019. "On the Coherency of Ocean and Land Surface Specular Scattering for GNSS-R and Signals of Opportunity Systems." *IEEE Transactions on Geoscience & Remote Sensing* 57 (12): 10426–10436. <https://doi.org/10.1109/TGRS.2019.2935257>.

- Bussy-Virat, C. D., C. S. Ruf, and A. J. Ridley. 2019. "Relationship Between Temporal and Spatial Resolution for a Constellation of GNSS-R Satellites." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 12 (1): 16–25. <https://doi.org/10.1109/JSTARS.2018.2833426>.
- Carreno-Luengo, H., C. S. Ruf, S. Gleason, and A. Russel. 2024. "Detection of Inland Water Bodies Under Dense Biomass by CYGNSS." *Remote Sensing of Environment* 301:113896. <https://doi.org/10.1016/j.rse.2023.113896>.
- Chang, M., P. Li, Z. Li, and H. Wang. 2022. "Mapping Tidal Flats of the Bohai and Yellow Seas Using Time Series Sentinel-2 Images and Google Earth Engine." *Remote Sensing* 14 (8): 1789. <https://doi.org/10.3390/rs14081789>.
- Chang, M., P. Li, Y. Sun, H. Wang, and L. Zhenhong. 2022. "Mapping Dynamic Turbidity Maximum Zone of the Yellow River Estuary from 38 Years of Landsat Imagery." *Remote Sensing* 14 (15): 3782. <https://doi.org/10.3390/rs14153782>.
- Chew, C., S. Lowe, N. Parazoo, S. Esterhuizen, S. Oveisgharan, E. Podest, C. Zuffada, and A. Freedman. 2017. "SMAP Radar Receiver Measures Land Surface Freeze/Thaw State Through Capture of Forward-Scattered L-Band Signals." *Remote Sensing of Environment* 198: 333–344. <https://doi.org/10.1016/j.rse.2017.06.020>.
- Chew, C., J. T. Reager, and E. Small. 2018. "CYGNSS Data Map Flood Inundation During the 2017 Atlantic Hurricane Season." *Scientific Reports* 8 (1): 9336. <https://doi.org/10.1038/s41598-018-27673-x>.
- Chew, C., and E. Small. 2020. "Estimating inundation extent using CYGNSS data: A conceptual modeling study." *Remote Sensing of Environment* 246: 111869. <https://doi.org/10.1016/j.rse.2020.111869>.
- Dong, Z., and S. Jin. 2021. "Evaluation of the Land GNSS-Reflected DDM Coherence on Soil Moisture Estimation from CYGNSS Data." *Remote Sensing* 13 (4): 570. <https://doi.org/10.3390/rs13040570>.
- Egessa, R., A. Nankabirwa, H. Ocaya, and W. Gandhi Pabire. 2020. "Microplastic Pollution in Surface Water of Lake Victoria." *Science of the Total Environment* 741: 140201. <https://doi.org/10.1016/j.scitotenv.2020.140201>.
- Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, et al. 2007. "The Shuttle Radar Topography Mission." *Reviews of Geophysics* 45 (2). <https://doi.org/10.1029/2005RG000183>.
- Feng, S., S. Liu, Z. Huang, L. Jing, M. Zhao, X. Peng, W. Yan, et al. 2019. "Inland Water Bodies in China: Features Discovered in the Long-Term Satellite Data." *Proceedings of the National Academy of Sciences* 116 (51): 25491–25496. <https://doi.org/10.1073/pnas.1910872116>.
- Gerlein-Safdi, C., A. A. Bloom, G. Plant, A. K. Eric, and S. R. Christopher. 2021. "Improving Representation of Tropical Wetland Methane Emissions with CYGNSS Inundation Maps." *Global Biogeochemical Cycles* 35 (12): e2020GB006890. <https://doi.org/10.1029/2020GB006890>.
- Gerlein-Safdi, C., and C. S. Ruf. 2019. "A CYGNSS-Based Algorithm for the Detection of Inland Waterbodies." *Geophysical Research Letters* 46 (21): 12065–12072. <https://doi.org/10.1029/2019GL085134>.
- Gleason, S., C. S. Ruf, A. J. O'Brien, and D. S. McKague. 2019. "The CYGNSS Level 1 Calibration Algorithm and Error Analysis Based on On-Orbit Measurements." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 12 (1): 37–49. <https://doi.org/10.1109/JSTARS.2018.2832981>.
- Holgerson, M. A., and P. A. Raymond. 2016. "Large Contribution to Inland Water CO₂ and CH₄ Emissions from Very Small Ponds." *Nature Geoscience* 9 (3): 222–226. <https://doi.org/10.1038/ngeo2654>.
- Huang, C., Y. Chen, S. Zhang, and W. Jianping. 2018. "Detecting, Extracting, and Monitoring Surface Water from Space Using Optical Sensors: A Review." *Reviews of Geophysics* 56 (2): 333–360. <https://doi.org/10.1029/2018RG000598>.
- Jun, C., Y. Ban, and L. Songnian. 2014. "Open Access to Earth Land-Cover Map." *Nature* 514 (7523): 434. <https://doi.org/10.1038/514434c>.
- Liu, J., L. Peng, T. Canran, H. Wang, Z. Zhou, Z. Feng, F. Shen, and L. Zhenhong. 2022. "Spatiotemporal Change Detection of Coastal Wetlands Using Multi-Band SAR Coherence and Synergetic Classification." *Remote Sensing* 14 (11): 2610. <https://doi.org/10.3390/rs14112610>.

- Loria, E., I. M. Russo, Y. Wang, G. Giangregorio, C. Galdi, M. D. Bisceglie, B. Wilson-Downs, et al. 2023. "Comparison of GNSS-R Coherent Reflection Detection Algorithms Using Simulated and Measured CYGNSS Data." *IEEE Transactions on Geoscience & Remote Sensing* 61:1–16. <https://doi.org/10.1109/TGRS.2023.3277411>.
- Loria, E., A. O'Brien, V. Zavorotny, B. Downs, and C. Zuffada. 2020. "Analysis of Scattering Characteristics from Inland Bodies of Water Observed by CYGNSS." *Remote Sensing of Environment* 245:111825. <https://doi.org/10.1016/j.rse.2020.111825>.
- Martins, V. S., E. M. L. M. Novo, A. Lyapustin, L. E. O. C. Aragão, S. R. Freitas, and C. C. F. Barbosa. 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 & Remote Sensing* 145:309–327. <https://doi.org/10.1016/j.isprsjprs.2018.05.013>.
- Morris, M., C. Chew, J. T. Reager, R. Shah, and C. Zuffada. 2019. "A Novel Approach to Monitoring Wetland Dynamics Using CYGNSS: Everglades Case Study." *Remote Sensing of Environment* 233: 111417. <https://doi.org/10.1016/j.rse.2019.111417>.
- Nghiem, S. V., C. Zuffada, R. Shah, C. Chew, S. T. Lowe, A. J. Mannucci, E. Cardellach, G. R. Brakenridge, G. Geller, and A. Rosenqvist. 2017. "Wetland Monitoring with Global Navigation Satellite System Reflectometry." *Earth & Space Science* 4 (1): 16–39. <https://doi.org/10.1002/2016EA000194>.
- Olokotum, M., V. Mitroi, M. Troussellier, R. Semyalo, C. Bernard, B. Montuelle, W. Okello, C. Quiblier, and J.-F. Humbert. 2020. "A Review of the Socioecological Causes and Consequences of Cyanobacterial Blooms in Lake Victoria." *Harmful Algae* 96:101829. <https://doi.org/10.1016/j.hal.2020.101829>.
- Pekel, J.-F., A. Cottam, N. Gorelick, and A. S. Belward. 2016. "High-Resolution Mapping of Global Surface Water and Its Long-Term Changes." *Nature* 540 (7633): 418–422. <https://doi.org/10.1038/nature20584>.
- Piao, S., P. Ciais, Y. Huang, Z. Shen, S. Peng, L. Junsheng, L. Zhou, et al. 2010. "The Impacts of Climate Change on Water Resources and Agriculture in China." *Nature* 467 (7311): 43–51. <https://doi.org/10.1038/nature09364>.
- Ruf, C. S., R. Atlas, P. S. Chang, M. Paola Clarizia, J. L. Garrison, S. Gleason, S. J. Katzberg, 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>.
- Tu, C., L. Peng, L. Zhenhong, H. Wang, S. Yin, L. Dahui, Q. Zhu, M. Chang, J. Liu, and G. Wang. 2021. "Synergetic Classification of Coastal Wetlands Over the Yellow River Delta with GF-3 Full-Polarization SAR and Zhuhai-1 OHS Hyperspectral Remote Sensing." *Remote Sensing* 13 (21): 4444. <https://doi.org/10.3390/rs13214444>.
- Wan, W., B. Liu, Z. Zeng, X. Chen, W. Guiping, X. Liwen, X. Chen, and Y. Hong. 2019. "Using CYGNSS Data to Monitor China's Flood Inundation During Typhoon and Extreme Precipitation Events in 2017." *Remote Sensing* 11 (7): 854. <https://doi.org/10.3390/rs11070854>.
- Wilks, D. S. 2011. *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Academic Press. USA: Elsevier.
- Woolway, R. I., and J. M. Christopher. 2019. "Worldwide Alteration of Lake Mixing Regimes in Response to Climate Change." *Nature Geoscience* 12 (4): 271–276. <https://doi.org/10.1038/s41561-019-0322-x>.
- Zeiger, P., F. Frappart, J. Darrozes, C. Prigent, and C. Jiménez. 2022. "Analysis of CYGNSS Coherent Reflectivity Over Land for the Characterization of Pan-Tropical Inundation Dynamics." *Remote Sensing of Environment* 282: 113278. <https://doi.org/10.1016/j.rse.2022.113278>.
- Zhang, S., M. Zhongmin, L. Zhenhong, P. Zhang, Q. Liu, Y. Nan, J. Zhang, H. Shengwei, Y. Feng, and H. Zhao. 2021. "Using CYGNSS Data to Map Flood Inundation During the 2021 Extreme Precipitation in Henan Province, China." *Remote Sensing* 13 (24): 5181. <https://doi.org/10.3390/rs13245181>.
- Zhu, Q., L. Peng, L. Zhenhong, P. Sixun, W. Xiao, B. Naishuang, and H. Wang. 2021. "Spatiotemporal Changes of Coastline Over the Yellow River Delta in the Previous 40 Years with Optical and SAR Remote Sensing." *Remote Sensing* 13 (10): 1940. <https://doi.org/10.3390/rs13101940>.