Analysis of CYGNSS coherent reflectivity over land for the characterization of pan-tropical inundation dynamics

Pierre Zeiger[a,*], Frédéric Frappart[b], José Darrozes[c], Catherine Prigent[d,e], Carlos Jiménez[d,e]

[a]*Laboratoire d'Etudes en Géophysique et Océanographie Spatiale (LEGOS), Université Paul Sabatier (UPS), Toulouse, France*
[b]*Interaction Sol Plante Atmosphère (ISPA), INRAE Bordeaux, Villenave-d'Ornon, France*
[c]*Géosciences Environnement Toulouse (GET), Université Paul Sabatier (UPS), Toulouse, France*
[d]*CNRS, LERMA, Observatoire de Paris, Sorbonne Université, Université PSL, Paris, France*
[e]*Estellus, Paris, France*

Abstract

The monitoring of flood and wetland dynamics at global scale is hampered by several limitations, including a reduced data availability in tropical areas due to the presence of clouds affecting visible and infrared imagery, or low spatial and/or temporal resolutions affecting passive and active microwave Earth Observation (EO) data. As a consequence, surface water extent estimates and their temporal variations remain challenging especially in equatorial river basins. Global Navigation Satellite System Reflectometry (GNSS-R) L-band signals recorded onboard Cyclone GNSS (CYGNSS) mission, composed of 8 Low Elevation Orbit (LEO) satellites, provide information on surface properties at high temporal resolution from 2017 up to now. CYGNSS bistatic observations were analyzed for detecting permanent water and seasonal floodplains over the full coverage of the mission, from 40°S to 40°N. We computed CYGNSS reflectivity associated to the coherent component of the received power, that was gridded at 0.1° spatial resolution with a 7-day time sampling afterwards. Several statistical metrics were derived from CYGNSS reflectivity, including the weighted mean and standard deviation, the median and the 90^{th} percentile (respectively Γ_{mean}, Γ_{std}, Γ_{median} and $\Gamma_{90\%}$) in each pixel. These parameters were clustered using the K-means algorithm with an implementation of the Dynamic Time Warping (DTW) similarity measure. They were compared to static inundation maps, and to dynamic estimations of surface water extent both at the global and regional scales, using the Global Inundation Extent from Multi-Satellites (GIEMS) and MODIS-based products. The difference between $\Gamma_{90\%}$ and Γ_{median} shows the best sensitivity to the presence of water. The river streams and lakes are correctly detected, and a strong seasonality is identified in CYGNSS reflectivity over the largest floodplains, with the exception of the Cuvette Centrale of Congo which is covered by dense vegetation. This seasonal reflectivity signal correlates well with inundation maps: Pearson's correlation coefficient between Γ_{median} and surface water extent from both GIEMS and MODIS is over 0.8 in the largest floodplains. The spatial patterns of reflectivity are consistent with static inundation maps: at the time of maximum flooding extent, a spatial correlation coefficient around 0.75 with Γ_{median} is obtained for several basins. We also evaluated the dependence of CYGNSS-derived clusters and reflectivity on the dominant land cover type and on the density of Above Groud Biomass (AGB) in the pixel. On the one hand, misclassifications of flooded pixels were observed over vegetated regions, probably due to uncertainties related to the attenuation by the vegetation in both CYGNSS and reference datasets. On the other hand, flooded pixels with a mean AGB

*Corresponding author

up to ∼300 Mg/ha were correctly detected with the clustering. High reflectivity values are also observed over rocky soils in arid regions and create false alarms. Finally, strong winds on large lakes cause surface roughness, and lower reflectivity values are observed in this case which weaken the detection of open water. While these constraints are to be taken in account and corrected in a future model, a pan-tropical mapping of surface water extent dynamics using CYGNSS can be envisaged.

Keywords: GNSS-Reflectometry, CYGNSS, coherent reflectivity, flood dynamics, surface water extent

1. Introduction

Wetland ecosystems and floodplains play a key role for the management of water and natural resources and for climate change, although they cover only 8% of the land surfaces (Davidson et al., 2018). They temper the water cycles through the regulation of river discharges and the mitigation of floods (Bullock & Acreman, 2003; Acreman & Holden, 2013). They are particularly sensitive to climate change, while contributing a lot to the world's greenhouse gases emissions through the global biogeochemical cycles. According to estimations in the literature, wetlands are the source of 20-25% of the world's methane emissions (Bartlett & Harriss, 1993; Whalen, 2005; Bergamaschi et al., 2007; Bloom et al., 2010; Ringeval et al., 2010; Melton et al., 2013; Nisbet et al., 2014; Saunois et al., 2020), and they store 16 to 33% of the soil carbon pool (Maltby & Immirzi, 1993; Page et al., 2011; Mitsch et al., 2013). Beyond the natural biogeochemical cycles, the anthropogenic activity strongly impacts wetlands, and exerts feedback loops along with climate change through the variations of the carbon stocks and methane emissions. Moreover, wetlands are known to be major reserves of biodiversity (Mitra et al., 2005; Junk et al., 2006; Webb et al., 2010). The temporal and spatial variations in inundation extent also affect the propagation of infectious diseases (Kouadio et al., 2012; Suk et al., 2020). All these effects affect the millions of people living worldwide in wetlands and relying on a healthy wetland ecosystem (Maltby & Acreman, 2011). Besides, unprecented floods and droughts increase the vulnerability in many regions of the world in spite of the improvement of risk management policies (Kreibich et al., 2022). For these aforementioned factors, a better monitoring from regional to global scales of the flood extent and dynamics is needed.

Inundation mapping usually depends on remote sensing due to the lack of *in-situ* data and the difficulties to perform measurements in remote regions. Long time-series of inundation extent at 30 m spatial resolution were obtained with the processing of Landsat multispectral images (Pekel et al., 2016), but at the expense of a low temporal resolution. The Moderate Resolution Imaging Spectroradiometer (MODIS) instrument was also used to produce a delineation of flooded area with a 8-day temporal resolution, but with a lower spatial resolution from 250 m to 1 km (Chen et al., 2013; Di Vittorio & Georgakakos, 2018). Nevertheless, while these products provide reliable information on open water bodies, they suffer from limitations due to the presence of clouds in equatorial areas, and are unable to detect water under dense canopies such as the inundated forests present in the Amazon and the Congo basins.

Microwave remote sensing is less affected by these limitations in both passive and active domains. The

presence of freshwater is responsible for: i) a decrease in the brightness temperature, in passive mode, affecting more sensibly the horizontal than the vertical polarization (Choudhury, 1991; Sippel et al., 1994; Prigent et al., 2001), ii) in active mode, low backscattering over open water and high backscattering in presence of water under vegetation owing to the double-bounce effect for side-looking radar such as Synthetic Aperture Radar (SAR) (Richards et al., 1987), and iii) also in active mode, high backscattering over open water and wetlands for nadir-looking radar altimeters (Frappart et al., 2021), except over large lakes where surface roughness (*e.g.* waves on windy days) can cause a strong scattering. L-band in particular is able to penetrate deeper the vegetation cover than the higher frequency microwave bands. It allows the detection of water under the vegetation even in equatorial basins, in both passive (Parrens et al., 2017) and active (Hess et al., 2003; Betbeder et al., 2014) domains. Nevertheless, major drawbacks limit the use of microwave observations for flood monitoring. First, passive microwave observations have a coarse spatial resolution, generally lower than 25 km, which limits their use for flood monitoring in spite of their quasi-daily temporal repeat. For example, the Soil Moisture and Ocean Salinity (SMOS, Kerr et al. (2001)) and the Soil Moisture Active Passive (SMAP, Entekhabi et al. (2010)) missions measure the brightness temperatures at L-band, with a nominal spatial resolution of several tens of kilometers and a revisit period of 2-3 days. Then, the use of Synthetic Aperture Radar (SAR) active microwave images was also limited by a low temporal sampling before the launch of Sentinel-1, the difficulty to identify the signature of water in complex environments, and the limited availability of L-band images at global scale. Finally, the use of radar altimetry for flood mapping is limited by its acquisition mode along the satellite tracks that does not offer a global coverage of land surfaces and by its low temporal resolutions ranging from 10 to 35 days. As a consequence, most of the studies were limited to regional mapping of the flood dynamics (Hamilton et al., 2004; Kuenzer et al., 2013; Parrens et al., 2017). The combination of multi-satellite information allows to overcome some of these limitations. The Global Inundation Extent Multi-Satellite (GIEMS) product combines information from passive microwave, radar scatterometers, and visible / near-infrared images to account for vegetation effect in the flood detection. It has been providing a continuous monitoring of the wetland dynamics globally, at monthly time scale and $\sim$0.25° spatial resolution since 1992 (Prigent et al., 2007, 2020).

Global Navigation Satellite System (GNSS) Reflectometry (GNSS-R) onboard satellite platforms can be a great help to improve the spatial and temporal resolutions of wetlands dynamics mapping. It uses the L-band (f = 1.575 GHz for L1) GNSS signals scattered by the Earth's surface, and collected by a receiver as a multistatic measurement technique (Martin-Neira, 1993; Zavorotny et al., 2014). In-situ and airborne GNSS-R measurements have shown sensitivity to various oceanic and land geophysical parameters (Ruffini et al., 2004; Cardellach et al., 2011; Egido et al., 2014). Over the last years, satellite missions carrying GNSS-R receivers have been launched and emerge as new tools for global applications.

The UK TechDemoSat-1 (TDS-1) mission was a proof of concept of spaceborne GNSS-R measurements. It successfully allowed it to retrieve either oceanic parameters such as sea level and wind speed (Foti et al., 2015; Clarizia et al., 2016), or land geophysical parameters such as soil moisture and vegetation (Camps et al., 2016;

Chew et al., 2016). It was followed by the launch of the NASA Cyclone GNSS (CYGNSS) mission, whose 8 Low Earth Orbit (LEO) micro-satellites are designed to monitor the formation of tropical cyclones through wind speed retrieval (Ruf et al., 2016). CYGNSS observations are also collected over land over the pan-tropical area (±38° latitude). The design of the mission suits the study of land parameters with a fine spatiotemporal resolution, as each of the 8 satellites records simultaneously 4 observations at a sampling rate of 1 Hz (2 Hz after July 2019). CYGNSS observations are sensitive to the properties of the reflecting surface, including surface roughness, soil moisture (SM), vegetation water content (VWC) and biomass density (Carreno-Luengo et al., 2019). One of the major applications over land of this dataset is soil moisture retrieval with a good accuracy (Root Mean Square Error ranging from 0.04 to 0.07 cm^3/cm^3) (Chew & Small, 2018; Kim & Lakshmi, 2018; Al-Khaldi et al., 2019; Clarizia et al., 2019; Eroglu et al., 2019; Senyurek et al., 2020). It has also been demonstrated that CYGNSS can be used to increase both the spatial and temporal sampling of existing SM products derived from passive microwave sensors (*e.g.*, SMAP, SMOS), that suffer from a lower spatiotemporal resolution (Yan et al., 2020).

CYGNSS reflectivity, based on the Delay Doppler Map (DDM) Peak power and metadata used for the calibration of the signal, was also used for identifying flood signatures in the spaceborne GNSS-R observations. In particular, the reflectivity maps show the changes in inundation extent following hurricanes, typhoons or extreme rain events (Chew et al., 2018; Morris et al., 2019; Wan et al., 2019; Ghasemigoudarzi et al., 2020; Rajabi et al., 2020). The methodologies mostly consist in an empirical thresholding of the CYGNSS reflectivity or Signal-to-Noise Ratio (SNR), except Ghasemigoudarzi et al. (2020) which used the Random Under-Sampling Boosted (RUSBoost) classification algorithm to identify flooded and non-flooded CYGNSS data. Gerlein-Safdi & Ruf (2019) used CYGNSS to produce annual water masks in the Congo basin, with a 0.1° spatial resolution, using a spatial comparison of one pixel's value with its neighbors and the random walker segmentation method. Then, Gerlein-Safdi et al. (2021) used the same approach temporally, for comparing the monthly mean of a pixel with the average value of its time series. They produced monthly water masks at 0.01° spatial resolution over the Pantanal and Sudd wetlands. Finally, Jensen et al. (2018) studied the relationship between CYGNSS reflectivity, floods and vegetation in tropical wetlands in the hydrographic basin of the Amazon. Based on this analysis, Rodriguez-Alvarez et al. (2019) classified the open water and flooded vegetation over the same study area. Note that all these studies were performed at the regional scale. At larger scale, CYGNSS reflectivity and coherence over flooded areas vary temporally and spatially due to vegetation cover, due to the changes in roughness affecting soil and water surfaces (Chew & Small, 2020), and to the bistatic geometry of the acquisition Loria et al. (2020).

Recent improvements have been made in the detection of surface water using several indicators of the coherence of CYGNSS observations. In particular, the raw Intermediate Frequency (IF) data occasionally collected were analyzed in several studies (Li et al., 2021, 2022a; Collett et al., 2022). They are unprocessed signal samples which contain phase information at a high sampling rate, before the incoherent averaging at 1-2 Hz used to derive Level-1 DDM products. Li et al. (2021) calculated the complex DDM and derived a coherent coefficient

(CoC) for BeiDou-3 raw IF. They found that a high CoC is highly linked with the presence of water bodies. Li et al. (2022a) proposed a multi-GNSS complex waveform product with a high sampling rate (1000 Hz), that enables the detection of surface water at ∼100 m spatial resolution along the reflection track. And Collett et al. (2022) showed that the coherence of CYGNSS raw IF samples highly relies on the standing surface water, and if no water body is found in the footprint, coherence is found for saturated soils ($SM \geq 0.45 cm^3/cm^3$) with low large-scale roughness. An entropy-based metric was also proposed by Russo et al. (2022) to characterize the coherence of CYGNSS raw IF. It was very sensitive to open water and floods in forested areas along Yucatan Lake and the Mississippi River (Chapman et al., 2022). An operational and continuous generation of raw IF data at different frequencies and polarizations is planned for the future European Space Agency (ESA) GNSS-R mission, HydroGNSS, of which the detection of surface water is one of the main science objectives (Unwin et al., 2021). But currently, the availability of raw IF data is low compared to CYGNSS Level-1 data, and they are more an object of investigation rather than an operational solution for the monitoring of flood dynamics. For this reason, Al-Khaldi et al. (2021b) gave an estimation of the coherence based on the shape of CYGNSS Level-1 DDM. Some pan-tropical water masks were further proposed in Al-Khaldi et al. (2021a) using the aggregation of CYGNSS observations with varying spatio-temporal resolution: 1 km/1 year, 3 km/3 months, and 6 km/2 weeks. At 1-km spatial resolution, the mask is annual at best but the permanent water bodies are well delineated, including small tributaries of the Amazon and Congo rivers. On the contrary, lower spatial resolution allows a much higher temporal sampling and the detection of flood events.

In this study, we focus on CYGNSS reflectivity to map flood dynamics with a 0.1° spatial resolution and a 7-day temporal resolution. The CYGNSS reflectivity time series are classified using a K-means clustering technique with Dynamic Time Warping (DTW) similarity measurement (Müller, 2007), to delineate the flooded areas. Results are compared to reference static and dynamic maps of inundations at global and regional scales, among other ancillary datasets. Section 2 presents the datasets, Section 3 the methodology of this study, while Section 4 and Section 5 show the results and a discussion about the main conclusions of this paper.

2. CYGNSS and reference datasets

2.1. CYGNSS

The CYGNSS mission is composed of 8 LEO micro-satellites covering the pan-tropical area (±38° latitude). Each satellite carries onboard a Delay Doppler Mapping Instrument (DDMI), which is composed of a GPS receiver, nadir-looking antennas for collecting GNSS signals, and a zenith-looking antenna for geolocation purpose. The DDMI records simultaneously 4 reflected signals integrated over a second (0.5 s after July, 2019), so the CYGNSS mission provides 32 (64, respectively) observations per second in the area of coverage.The temporal sampling is high, with a median and mean revisit time over a 25 km pixel of 3 hours and 7 hours, respectively (Ruf et al., 2016). With higher spatial resolution, the revisit time increases. Still, substantial improvements in terms of temporal resolution are expected using CYGNSS when compared to monostatic radars.

Due to the bistatic configuration of the measurements between an emitting Global Positioning System (GPS) satellite and a receiving CYGNSS observatory, the sampling of CYGNSS observations over the Earth's surface is pseudo-random. The L-band signals ($f = 1.575$ GHz for GPS L1) are emitted by GNSS satellites and received by each of the 8 CYGNSS micro-satellites, with an incidence angle ranging from 0° to ∼70° and following a Gaussian distribution centered on ∼30°. Over the ocean, the reflected GPS signals are dominated by the incoherent component due to an important surface roughness. The spatial resolution is there degraded and wind speed can be estimated with a 25 km spatial resolution and a sub-daily mean revisit time (Clarizia & Ruf, 2016). However, the reflected signals over land can be dominated by the coherent component, in case of the presence of water bodies or areas characterized by a high surface moisture content in the glistening zone, with low surface roughness at the scale of the GPS L1 wavelength (19.03 cm). The spatial resolution depends, in that case, on the size of the first Fresnel zone which is larger than ∼0.6 km × 0.6 km (the minimum area in the case of a 0° incidence angle) (Eroglu et al., 2019; Rodriguez-Alvarez et al., 2019). The finest theoretical spatial resolution for a CYGNSS observation over flooded regions is therefore ∼0.6 km × 6.6 km, with an elongation effect due to the integration of the reflected signals over 1 second (0.5 s, respectively) for observations before (after) July, 2019. Yet, CYGNSS observations are sensitive to the presence of a small fraction of water in the glistening zone, whose contribution dominates the total response of the reflecting surface. The maps of reflectivity over river basins, such as the Amazon, show a strong scattering over small tributaries, down to a few hundred meters.

The CYGNSS observables used in this study come from CYGNSS Level 1 science data record files (CYGNSS, 2020). They contain the 17x11 Delay Doppler Maps (DDMs) of raw bins and analog scattered power, along with all the parameters used for data processing, *i.e.* the geometry of the acquisition and data flags. We used the version 3.0 of L1 data over one year, from August 1^{st}, 2018 to July 31^{st}, 2019. The observations before July, 2019 are sampled at 1 Hz (2 Hz after), so we assume a ∼0.6 km × 6.6 km spatial resolution although the along-track resolution is twice higher for the last month of data. The daily files (one for each CYGNSS satellite) are distributed in the Physical Oceanography Distributed Active Archive Center (PODAAC: https://podaac-opendap.jpl.nasa.gov/opendap/hyrax/allData/cygnss/L1/v3.0/). The CYGNSS DDM of scattered analog power are used in this study, along with quality flags and metadata from the L1 files. The description of the preprocessing and analysis of CYGNSS dataset can be found in Section 3 (methodology).

2.2. Other sources of data

Other data sources were used to analyze and validate the results presented below. They consist in static and dynamic inundation maps available at regional and global scales, land cover thematic map and Above Ground Biomass (AGB).

2.2.1. Static inundation maps

Two different datasets were considered for a delineation of flooded areas: the Level 3 Global Lakes and Wetlands Database (GLWD-3) at 30" resolution (Lehner & Döll, 2004), and the wetlands maps at 15" (500 m

at the Equator) from Tootchi et al. (2019). Note that these two datasets do not provide any temporal variations and are thus static. After some analyzes, the latter was selected because it provides maps of both the Regularly Flooded Wetlands (RFWs, overlapping open water with several inundation datasets) and the Composite Water (CWs) which associates RFWs with Groundwater modelling (GDWs). It is of interest to compare RFWs to CWs maps, because CYGNSS may be sensitive to the groundwater-driven wetlands that represent a considerable source of floods. The CW-WTD product is used as it provides both RFWs areas, and a groundwater wetlands map derived from a direct water table depth (WTD) modeling (Fan et al., 2013).

The RFWs in Tootchi et al. (2019) were derived merging 3 inundation datasets, namely the open water and flooded areas extracted from the ESA-CCI Land Cover (see Section 2.2.3), the mean annual maximum water extent from GIEMS-D15 (Fluet-Chouinard et al., 2015), and the maximum surface water from the Joint Research Center (JRC) product (Pekel et al., 2016). It assumes that every single product has observation gaps and should be complemented by the use of other ones. It is particularly the case between the high-resolution JRC maps based on Landsat, and GIEMS-D15 which downscales the coarse 0.25° estimations from GIEMS at 15" resolution. The JRC product is unable to detect water under dense vegetation cover, while GIEMS-D15 is able to detect the floods in tropical or boreal forests. However, GIEMS-D15 is dominant when compared to other products. It represents 55% of the total RFWs areas, and only 10% of its extent is confirmed by either CCI land cover or JRC surface water. As a consequence, the uncertainties of GIEMS are directly propagated in the RFWs dataset. In particular, a recurrent overestimation of Surface Water Extent (SWE) in GIEMS has been found in the Ganges irrigated paddy fields, in the Sahel wetlands, and in coastal areas, due to a confusion between the signals from surface water and wet soils. Also, as the original spatial resolution of GIEMS is 0.25°, the sensitivity to small river streams in several regions, including the Amazon and Congo basins, is found to be low in both GIEMS-D15 and Tootchi's RFWs dataset.

2.2.2. Dynamic inundation maps

Few global products exist that provide temporal variations of inundation extent at the global scale. To compare with CYGNSS data, we used the Global Inundation Extent from Multi-Satellite version 2 (GIEMS-2) dataset (Prigent et al., 2020). It contains a global estimation of monthly flooded area in a 0.25° grid since 1992. In this study, we have used the new global maps for years 2018 and 2019, matching the times of our CYGNSS dataset.

A more precise comparison is performed regionally between CYGNSS observables and multispectral-based flood maps derived from the gridded atmospherically corrected surface reflectances from MODIS, acquired in seven spectral bands from visible/ near infrared (NIR) to shortwave infrared (SWIR). The surface reflectances contained in the MOD09A1 product (8-day binned level 3, version 6) have a spatial resolution of 500 m and a temporal resolution of 8 days. Flood extent maps were generated using the threshold method proposed in Frappart et al. (2018) and Normandin et al. (2018). It is a simplified version of the multi-threshold approach developed by Sakamoto et al. (2007). In this approach, a MODIS pixel is considered fully or partially covered with water if: i) the value of its Enhanced Vegetation Index (EVI) (Huete et al., 1997) is lower or equal to 0.05

and the value of its land surface water index (LSWI) (Xiao et al., 2005) is negative or equal to zero, or ii) its EVI value is lower than 0.3 and the difference between EVI and LSWI is lower than 0.05. This method was used to produce inundation maps in the Lower Mekong Basin (LMB) in Asia, the Inner Niger Delta (IND) in Africa, and La Plata basin in South America during the study period.

2.2.3. CCI Land Cover

The characteristics of CYGNSS reflectivity highly rely on the type of Land Cover (LC) present in the glistening zone (Carreno-Luengo et al., 2019; Chew & Small, 2020). To analyze the capability of CYGNSS to provide reliable information on the presence of water over land depending on the LC type, we used global LC maps from the European Space Agency's (ESA) Climate Change Initiative (CCI) at 300 m resolution (ESA (2017), available at https://www.esa-landcover-cci.org/). These maps are obtained from 1992 to 2015 using various imagery data, including the Medium Resolution Imaging Spectrometer (MERIS) and Project for On-Board Autonomy – Vegetation (PROBA-V), and applying the GlobCover unsupervised classifier (Defourny et al., 2007). More recent maps have been produced and validated by the Copernicus Climate Change Service (C3S), so we used the 2019 map which corresponds the best to our CYGNSS time series. Additionally, we aggregated the initial 38 LC types defined following the United Nations Land Cover Classification System (UN-LCCS) (Di Gregorio, 2016) into 10 general land cover classes, which is summarized in Table 1. Croplands, forests and herbaceous are separated into dry and flooded regions to study the changes in CYGNSS reflectivity associated to the presence of water.

Table 1: List of the 10 aggregated land cover classes, and the associated class numbers in the CCI LC classification system.

Class numbering	**Class name**	**CCI Land Cover classes associated**
LC1	Open water	210
LC2	Cropland non irrigated	10, 11, 12, 30
LC3	Cropland irrigated	20
LC4	Non flooded forest	50, 60, 61, 62, 70, 71, 72, 80, 81, 82, 90
LC5	Flooded forest	160, 170
LC6	Dry shrubs, herbaceous	40, 100, 110, 120, 121, 122, 130
LC7	Flooded shrubs, herbaceous	180
LC8	Bare soils	140, 150, 151, 152, 153, 200, 201, 202
LC9	Snow, ice	220
LC10	Urban	190

2.2.4. Above Ground Biomass

Dense vegetation layers can attenuate and even extinct the coherent component of GNSS-R signals (Loria et al., 2020), and the corresponding decrease of CYGNSS reflectivity was also modeled with respect to the Above Ground Biomass (AGB) of the reflecting surface (Carreno-Luengo et al., 2020). We therefore analysed the distribution of areas detected as flooded and non-flooded using CYGNSS reflectivity, depending on the mean Above Ground Biomass (AGB) in the pixel. We used the global GlobBiomass AGB maps (Santoro, 2018) which are given at 1 ha (approximately 3.2 arcsec) spatial resolution for the reference year 2010. It was derived from a set of observations including the L-band backscatter from ALOS-PALSAR and multiple C-band observations from Advanced Synthetic Aperture Radar (ASAR) onboard Envisat, with lidar-based metrics from Ice Cloud and Land

Elevation Satellite (ICESat), and Landsat reflectances at several bands (Santoro et al., 2021). The study independently estimated the Growing Stock Volume (GSV) for the backscatters of ALOS-PALSAR and ASAR, which are combined to obtain a more robust dataset and then invert AGB. The AGB maps reproduced well the known spatial patterns with high level of detail. Over the tropics, the comparison with a database of field measurements highlighted a systematic underestimation of AGB in the densest forests. This is due to the low dynamic range of backscatter observations over regions with AGB higher than 250 Mg/ha, and to the unavailability of ASAR GSV estimates which causes the AGB inversion to only rely on ALOS-PALSAR backscatter.

2.2.5. Resampling of the ancillary datasets

All the ancillary datasets have higher spatial resolutions than our 0.1° CYGNSS grid, ranging from ∼5100 m to 1 km at the equator. To compare with CYGNSS observables, these data were upscaled into a matching 0.1° grid. For CCI LC, the percentage of each LC class in every pixel was computed. The mean AGB was extracted from GlobBiomass 1 ha spatial resolution dataset. For static and dynamic inundation maps, we calculated the open water and flood extent as a percentage of the pixel surface. A notable exception is GIEMS-2, whose spatial resolution is 0.25°. To compare at the global scale GIEMS vs CYGNSS, we also computed CYGNSS-derived parameters at 0.25° spatial resolution.

2.2.6. Water levels in 2018-2019

We have used water levels derived from radar altimetry as a proxy to estimate the severity of floods in the different river basins further analyzed, during the time span of our study (August 2018 - July 2019). We downloaded the time series of water levels from the Hydroweb database (Crétaux et al., 2011; Hydroweb) at several specific locations which are described in the caption of Table 2. We extracted the yearly maximal water heights since 2016 (launch of Sentinel-3A). We then computed the anomaly of 2018-2019 peak water height compared to the average annual maximum, which are listed in Table 2. The relative anomaly represents the anomaly normalized by the average amplitude of seasonal variations, and is also shown. We notice high water levels in the Mekong River (2.96 m over the average, with a relative anomaly of 36.5%), and in South America: 0.76 - 0.77 m for the Orinoco and Madeira rivers, with a relative anomaly of 11.1 - 11.5%. This could be due to a weak El Niño in 2018-2019 that may also have affected the Yangtze basin (not shown here). The Parana has a relative anomaly of 42.2% mainly due to a strong decrease of water heights and inundation extent after the severe droughts of the summer 2019, posterior to the time span of our study. For the other basins, the Congo, Ganges and Brahmaputra rivers show low negative anomalies of 0.19 m to 0.39 m (-2.3% to -14.3%), while the Niger shows a positive anomaly of 0.54 m (+7.2%), likely linked to high floods in the Inner Niger Delta. We can also mention the very long-lasting 2019 spring floods along the Mississippi River, that have already been analyzed in several CYGNSS-related studies (Li et al., 2021; Chapman et al., 2022).

Table 2: Anomalies of the 2018-2019 peak water height compare to the average yearly maximum, at the locations listed hereafter. The **Orinoco** River, downstream the Llanos de Orinoco; the **Amazon** River close to its estuary, and the **Madeira** River (one of the biggest tributaries of the Amazon), downstream the confluence between the Beni, Madre de Dios and Mamore rivers; the **Parana** River, after its confluence with Rio Salado; the **Congo** River, after its confluence with the Ubangui; the **Niger** River, downstream the Inner Niger Delta (IND); the **Ganges** and **Brahmaputra** rivers just before their confluence; and finally, the **Mekong** River at the outlet of Tonle Sap.

River	Orinoco	Amazon	Madeira	Parana	Congo	Niger	Ganges	Brahm.	Mekong
Peak anomaly (m)	0.77	-0.03	0.76	1.24	-0.39	0.54	-0.19	-0.3	2.96
Rel. anomaly (%)	11.5	-0.4	11.1	42.2	-14.3	7.2	-2.3	-5.6	36.5

3. Methods

3.1. Preprocessing of the CYGNSS dataset

An overview of the processing chain of CYGNSS data developed in this study is presented in Figure 1. First, the peak of each CYGNSS DDM in the daily L1 files (see Section 2.1) is extracted and the entire 17x11 L1 DDM is removed, as a further processing would be very time-consuming. The CYGNSS overland flag is used to remove all reflections over the oceans, and some other quality flags are applied: *S-band powered up*, *Large spacecraft attitude error*, *Black-body DDM*, *DDM is test pattern*, *Low confidence GPS EIRP estimate*. This set of flags was used in previous studies to ensure a correct filtering with the removal of the least samples possible (Chew & Small, 2018; Eroglu et al., 2019). We also tried to apply more constraining masks using all the flags combined in the CYGNSS overall quality flag with the logical exception of the overland flag (Clarizia et al., 2019; Rodriguez-Alvarez et al., 2019). However, this latter approach removed a major part of CYGNSS observations over land (Eroglu et al., 2019), and decreased the performance of land geophysical parameter retrieval with a high spatiotemporal resolution.

The CYGNSS L1 algorithms performed an estimation of the specular point location based on a mean sea surface model, which is close to the geoid (Gleason et al., 2019). This can affect the DDMs over land as topography is not taken into account. Before the end of 2017, data collected over 600 m were of poor quality, which lead several authors to apply an elevation cutoff (Eroglu et al., 2019; Rodriguez-Alvarez et al., 2019; Yan et al., 2020). Further versions of the L1 calibration procedures for land reflections include the topography to estimate more accurately the specular point location (Gleason et al., 2020). For the time span of our study, CYGNSS DDM are of good quality over a large variety of terrains and elevation ranges, including almost all the wetlands and water bodies in CYGNSS coverage. Notable exceptions are the Tibetan plateau in the Himalayas, and the Altiplano in South America (Gleason et al., 2020). Still, the position of the peak power in CYGNSS DDMs is sometimes shifted due to several factors, including the topography around the specular point. In this study, we filtered the observations when the peak of the 17x11 DDM is located in the three first and last delay rows as previously reported (Yan et al., 2020). This corresponds to a quality control affecting a small part ($\sim$4%) of the full dataset.

3.2. CYGNSS reflectivity

The computation of the reflectivity is mostly preferred to the use of the Signal-to-Noise Ratio (SNR) when analyzing CYGNSS data, as it combines information on the DDM peak power along with other CYGNSS metadata useful for the calibration of the signal. The GNSS signals scattered by open water, floodplains and in general smooth land and ice surfaces are considered to be dominated by the coherent component (Carreno-Luengo et al.,

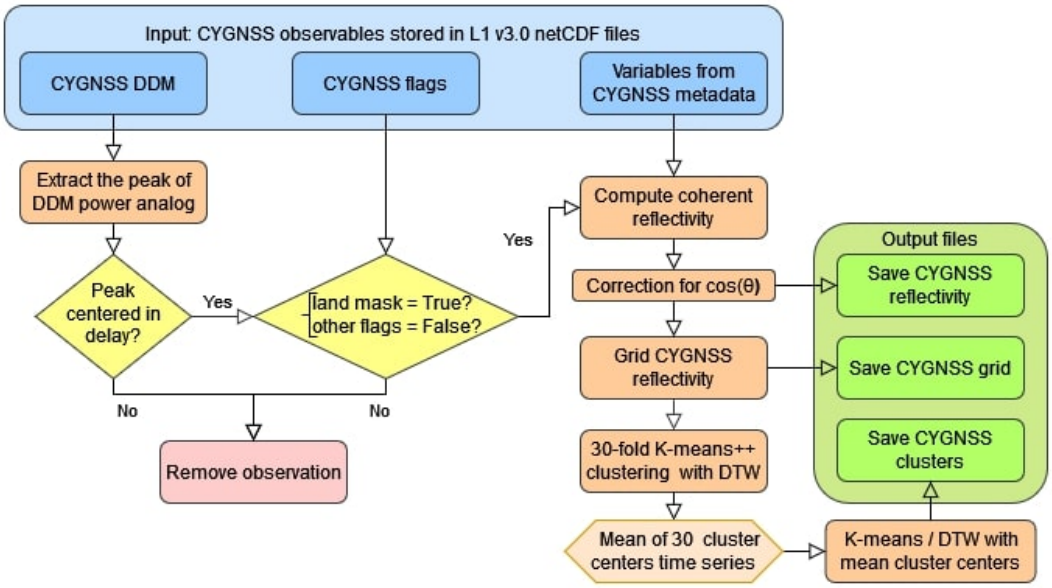

Figure 1: Overview of the processing chain for the analysis and the clustering of CYGNSS observations.

2019; Li et al., 2017; Rodriguez-Alvarez et al., 2019). The power ratio (PR) defined in Al-Khaldi et al. (2021a,b) also highlights high coherent returns with a low diffuse scattering over inland water bodies. In our study, CYGNSS reflectivity is hence determined using the coherent scattering equation (De Roo & Ulaby, 1994; Gleason et al., 2020):

$$\Gamma(\theta) = \left(\frac{4\pi}{\lambda}\right)^2 \frac{P_{DDM}(R_r + R_t)^2}{G_r G_t P_t} \tag{1}$$

where P_{DDM} is the peak of the DDM analog power, R_r and R_t are the receiver and the transmitter ranges (distance from the receiver and the transmitter to the specular point), G_r is the receiver antenna gain, $P_t G_t$ is the GPS Equivalent Isotopically Radiated Power (EIRP), and λ is the GPS L1 signal wavelength (i.e. $\lambda = \lambda_{L1} =$ 0.1903 m). P_{DDM} is computed using the 17 delay x 11 Doppler DDMs, while the other variables come from CYGNSS Level 1 metadata. In Equation (1), the reflectivity depends on the incidence angle (θ) of CYGNSS observations. In Figure 2, the curves of the mean reflectivity vs. θ are shown without correction (red curve), with a correction based on $cos(\theta)$ (in blue), and with a correction based on $cos^2(\theta)$ (in green). With no correction, the reflectivity decreases while increasing the incidence angle. With the $cos(\theta)$ correction, the reflectivity is $\sim$constant in all the incidence range, except for $\theta > 68^\circ$ (these values are finally filtered out). With the latter correction in $cos^2(\theta)$, the reflectivity increases exponentially with an increasing θ. These corrections based on $cos(\theta)$ or $cos^2(\theta)$ are commonly used in correcting the influence of the incidence angle on SAR backscatter (Ulaby et al., 1982; Kellndorfer et al., 1998). As a consequence, the reflectivity corrected from the incidence angle Γ is:

$$\Gamma = \Gamma(\theta = 0^\circ) = \frac{\Gamma(\theta)}{cos(\theta)} \tag{2}$$

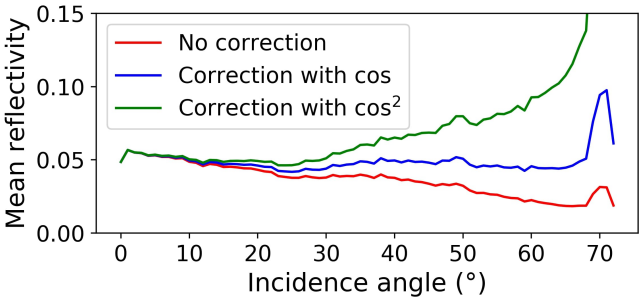

Figure 2: Mean reflectivity vs. the incidence angle for CYGNSS observations without correction for θ, with a correction based on $cos(\theta)$, and with a correction based on $cos^2(\theta)$

Once CYGNSS reflectivity was computed, these unevenly distributed values were gridded at 0.1° spatial resolution (∼11 km at the equator). Several tests were performed to evaluate the best compromise between spatial and temporal resolutions. For the mapping of variables exhibiting a high temporal variability, such as SM, a daily time step is preferable and the best spatial resolution associated is 0.25°. For studying flood dynamics, we prefered to ensure a higher spatial resolution to reduce the confusion between flooded and non-flooded areas. With a 0.1° grid, a 7-day time sampling can be reached. This observation is consistent with Al-Khaldi et al. (2021a), who mapped CYGNSS PR at 2 weeks/6 km, 3 months/3 km or 1 year/1 km. Moreover, our grid is based on a maximum of 32 observations per second (1 Hz sampling of each CYGNSS track). For data after July 2019 the sampling is 2 Hz, so the spatial resolution at weekly time scale could be improved. Finally, an alternative 0.25° grid was also computed to compare CYGNSS reflectivity with the flooded extent from GIEMS-2.

Due to the pseudo-random configuration of CYGNSS bistatic observations, some missing values were present in most of the pixels at specific time steps. A 30-day moving window was applied during the gridding step to fill these gaps. Figure 3 shows that this process increased the average number of samples per pixel, while the orbits of CYGNSS satellites ensure more observations at the extreme latitudes, and fewer ones in the equatorial regions. A 30-day Gaussian window with a standard deviation σ = 7 days was used to weight the observations. The weighted mean and standard deviation values of reflectivity per pixel were calculated, which are further expressed as Γ_{mean} and Γ_{std} in linear units. Other statistical parameters were used to describe the distribution of

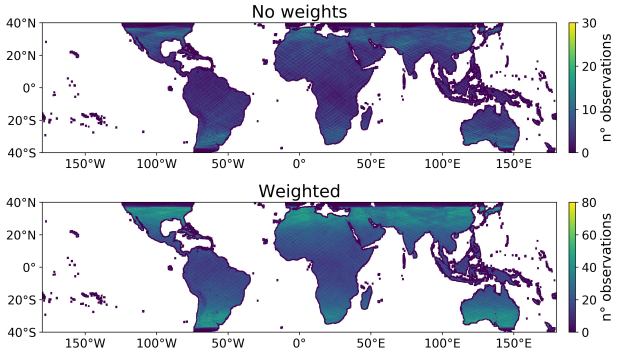

Figure 3: Number of observations per pixel in 7 days, from 2019-01-02 to 2019-01-08. **(a)** Count without the moving window, **(b)** count with a moving window and 1-month Gaussian weighting

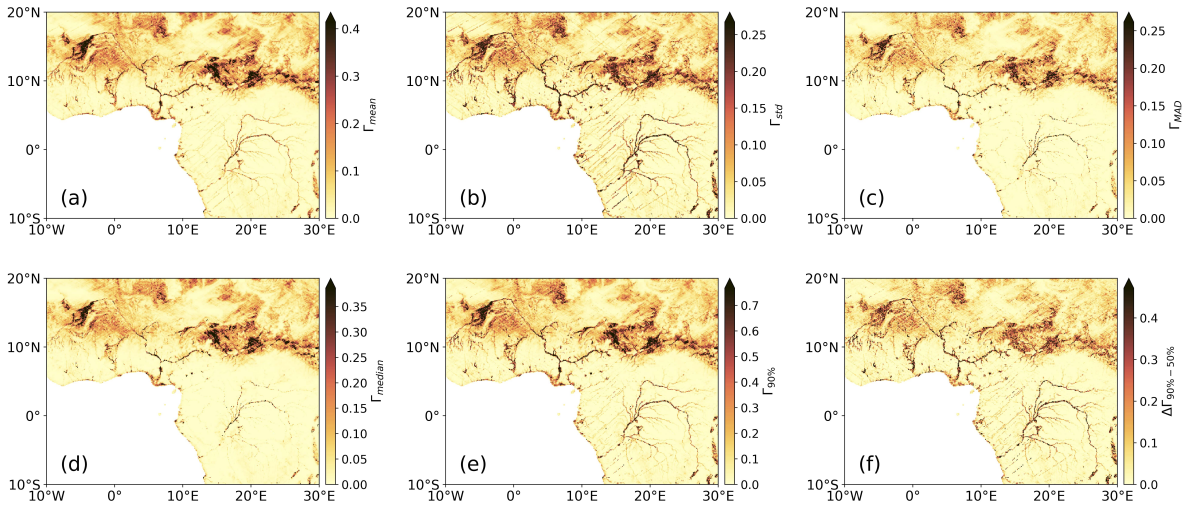

Figure 4: Parameters derived from CYGNSS reflectivity over western and central Africa, with the Congo, Niger and Chad hydrographic basins, at time step 2018-09-05. **(a)** Γ_{mean}, **(b)** Γ_{std}, **(c)** Γ_{MAD}, **(d)** Γ_{median}, **(e)** $\Gamma_{90\%}$, **(f)** $\Gamma_{90\%-50\%}$

CYGNSS reflectivity values in each pixel: the median (Γ_{median}), the 90^{th} percentile ($\Gamma_{90\%}$), the median absolute deviation (MAD, Γ_{MAD}), and the difference between $\Gamma_{90\%}$ and Γ_{median} ($\Gamma_{90\%-50\%} = \Gamma_{90\%} - \Gamma_{median}$). They were extracted at each time step of the 0.1° grid.

A subset of these variables is shown in Figure 4 over western and central Africa, including a major part of the Congo, Chad and Niger basins, for the first week of September when the Sahel region is affected by floods. The reflectivity derived from CYGNSS observations is usually high over flooded regions and rivers, which are smooth reflecting surfaces. All the observables show high values over large wetlands, such as the Inner Niger Delta (IND) and the floodplains around Lake Chad. Over smaller water bodies, the 0.1° pixels are composed of a fraction of water with non-inundated areas dominating around. As a consequence, Γ_{median} is low as it comes from an observation over soils non-covered with water, while $\Gamma_{90\%}$ corresponds to the highest reflectivity values associated to water bodies. For this reason, the detection of rivers is lower using Γ_{median} and Γ_{MAD} than with other variables, as it is observed along the streams of the Congo and Niger rivers and their tributaries. The interest of Γ_{median} and Γ_{MAD} is their robustness to noise, when compared to Γ_{mean} and Γ_{std}, respectively. Finally, the $\Gamma_{90\%-50\%}$ parameter is relevant, because it discriminates well the river streams (high $\Gamma_{90\%}$ and low Γ_{median}), the large water bodies and floodplains (high $\Gamma_{90\%}$ and high Γ_{median}), and the non-flooded areas (low $\Gamma_{90\%}$ and low Γ_{median}).

The maps of $\Gamma_{90\%}$ and $\Gamma_{90\%-50\%}$ in the full coverage of CYGNSS mission are shown in Figure 5. The Amazon, Parana, Congo and Niger Rivers and their tributaries, as well as other smaller streams in the pan-tropical area, are clearly delineated for both parameters. High reflectivity values are also obtained over the major floodplains in South America, Africa or Southeast Asia. Finally, a strong specular scattering is observed in the arid regions over bare soils, such as in the Sahara, the Arabian Peninsula, and Australia (see Section 5.4, discussion). Due to the homogeneity of the land cover in these areas, high values are obtained for both $\Gamma_{90\%}$ and Γ_{median}. As a consequence, the difference $\Gamma_{90\%-50\%}$ is low, which improves the separation between water bodies, floodplains

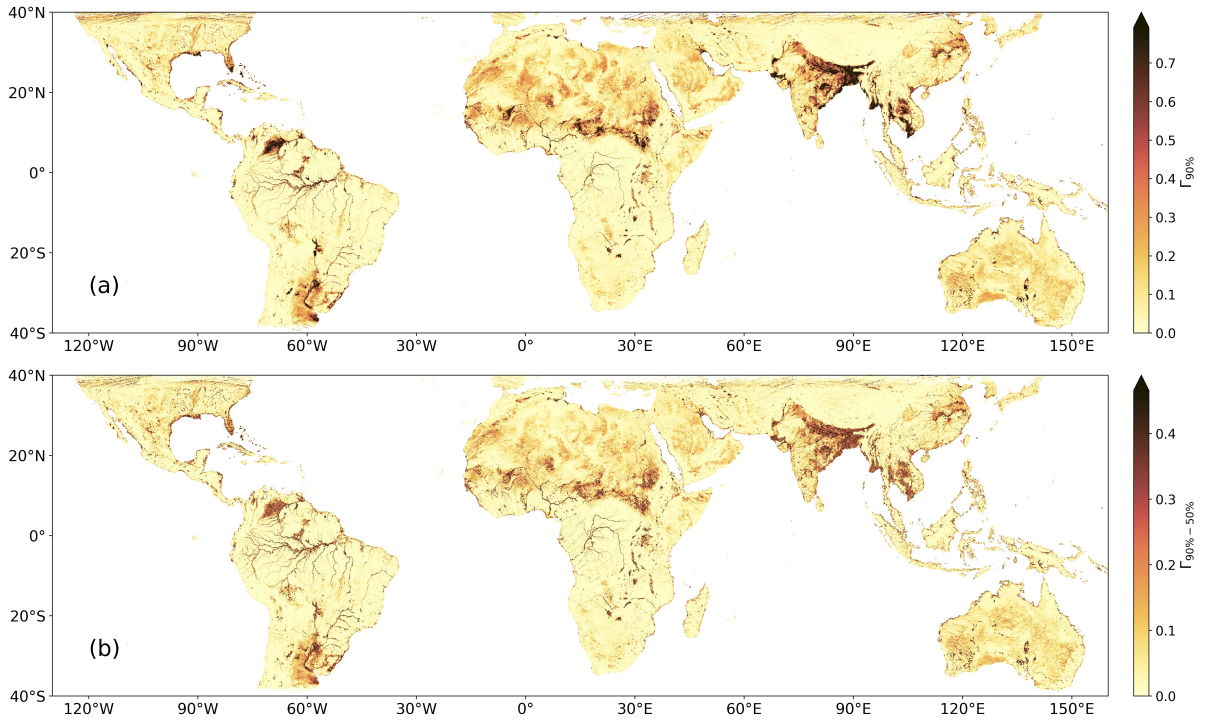

Figure 5: Values at time step 2018-09-05 of CYGNSS-derived parameters: **(a)** $\Gamma_{90\%}$, **(b)** $\Gamma_{90\%-50\%}$

and bare soils. For all the aforementioned factors, $\Gamma_{90\%-50\%}$ and $\Gamma_{90\%}$ were used along with Γ_{median} for further analysis in this paper.

3.3. Clustering

We clustered the CYGNSS-derived parameters Γ_{mean}, Γ_{std}, Γ_{median}, Γ_{MAD}, $\Gamma_{90\%}$ and $\Gamma_{90\%-50\%}$ using the unsupervised K-means technique (Macqueen, 1967). Starting from an initial set of centroids, it assigns all the points to the closest cluster center and then calculates the mean value of each cluster. This defines a new set of centroids, and the algorithm iterates until convergence. The K-means++ method is used to select the initial cluster centers, as it performs better and provides a good stability of the final solutions (Arthur & Vassilvitskii, 2007). The first centroïd is chosen randomly at this step. Then, the other centroids are iteratively determined between all the dataset points, with a probability equal to the inverse distance between a point and the closest centroïd, the weights being adjusted every iteration. This ensures a stable distribution of the cluster centers that is representative of the dataset.

In this study, we aim to detect inundated areas whatever their seasonality, with a time-series based approach. This is not possible using the K-means clustering with Euclidean distance, which makes the algorithm extremely sensitive to shifting and distortion in time. To overcome this limitation of the default K-means algorithm, we used the Dynamic Time Warping (DTW) similarity measure (Berndt & Clifford, 1994; Müller, 2007) which is implemented in the python package *tslearn* (Tavenard et al., 2020).

Given two time series $X = (x_1, x_2, \ldots, x_N), N \in \mathbb{N}$ and $Y = (y_1, y_2, \ldots, y_M), M \in \mathbb{N}$, the optimization problem associated with DTW is formulated as follow:

$$DTW(X,Y) = min_{\pi} \sqrt{\sum_{(i,j)\in\pi} d(x_i, y_j)} \quad (3)$$

where $\pi = [\pi_0, \pi_1, \ldots, \pi_K]$ with $\pi_k = (i_k, j_k)$ is called a warping path, and $d(\cdot,\cdot)$ is a distance metric (by default the Euclidean distance is used). The warping path must satisfy the following conditions:

(i) Boundary condition: $\pi_0 = (1,1)$ and $\pi_K = (N, M)$

(ii) Monotonicity condition: $i_k \leq i_{k+1}$ and $j_k \leq j_{k+1}, \forall k \in [1, K-1]$

(iii) Step size condition: $i_{k+1} - i_k \leq 1$ and $j_{k+1} - j_k \leq 1, \forall k \in [1, K-1]$

The minimization of all the potential warping paths according to the distance metric $d(\cdot,\cdot)$ results in the optimal warping path, which is here expressed as $DTW(X,Y)$. The optimal warping path is computed between the K-means cluster centers and every pixel at each iteration of the algorithm, and permits a shift and a distortion in time between the time series. This makes the DTW similarity measure well suited for our study. When applied to the detection of inundations, it allows to identify similarities in flood patterns from distinct regions of the world, with differences in phase and intensity. One limitation in our case is the boundary condition (i), because we only use one year of CYGNSS data to detect floods with a yearly seasonal cycle. Thus, we implemented in our methodology a padding of the CYGNSS dataset over 3 consecutive years to limit the boundary effects. The evaluation of a 3 to 5-year dataset of CYGNSS observations would be optimal. However, the choice of a padding was driven by constraints on our computation capacities. It assumes a constant seasonality, which seems reasonable. The methodology could still be extended to the full 5-year dataset of CYGNSS in the future.

The choice of an optimal number of centroids in the K-means algorithm can be discussed. Due to the important computational resources needed by the DTW similarity measure, we were not able to calculate any metric as the Calinski-Harabasz score (Calinski & Harabasz, 1974) and the Silhouette score (Rousseeuw, 1987) with this version of the algorithm. We finally adopted an empirical approach for the determination of the optimal number of clusters, and we tested our methodology with a number of classes ranging from 2 to 8. The results are evaluated graphically with the maps and time series of labelled pixels, and numerically with confusion matrices between CYGNSS clusters and flood reference classes based on Tootchi et al. (2019). We also evaluated the capability of each of the 6 CYGNSS-based parameters presented in Figure 4 to detect floods and water bodies, either individually or associated to each-other. The most interesting results are obtained using $\Gamma_{90\%-50\%}$, which associates the information of Γ_{median} and $\Gamma_{90\%}$ into a unique variable. We therefore only present the results based on this parameter, although all the configurations were evaluated.

3.4. Sensitivity to the choice of initial centroids

Although the use of K-means++ algorithm produced stabler results than a pure random selection of the initial centroids, some differences were observed between successive versions of the clustering. We computed

30 times the K-means++ / DTW algorithm with 4 clusters and the $\Gamma_{90\%-50\%}$ parameter (the best configuration, see Section 4), to perform a sensitivity analysis. We extracted the inertia (sum of distances of samples to their closest cluster centers) at each iteration. We also calculated the percentage of correspondence between the labelled pixels from every pair of distinct clustering results. If the correspondence reaches 100%, the two versions are exactly the same. Table 3 presents some statistical parameters for both the inertia and the correspondence, which are the mean, median, min, max and std values. As can be seen, the inertia covers a very low dynamic range from 0.1005 to 0.1055 with a standard deviation of 10^{-3}, indicating that the algorithm converges toward similar solutions, whatever the choice of the initial centroids by the K-means++ method. The dynamic range of the correspondence is higher ($\sim 80 - 99\%$), with a median value of 94.57% and a standard deviation of 4.07%, respectively. The results are usually quite stable, in the range of 92%-99%, with the exception of few ones. Graphically, the different solutions found in 30 iterations also look very consistent.

Table 3: Sensitivity of the clustering result to the choice of initial centroids by the K-means++ algorithm. We calculated the inertia of each clustering, and the correspondence between every pair of distinct clustering results. We present here the mean, median, minimum, maximum and standard deviation values of both parameters.

Result	mean	median	min	max	std
Inertia	0.1027	0.1026	0.1005	0.1055	0.0010
Correspondence (%)	93.76	94.64	80.26	98.78	4.01

We finally achieved a second round of the K-means / DTW clustering, which was based on the results of the sensitivity analysis. There, the 4 initial centroids were obtained by averaging the cluster centers of the 30 outputs obtained with K-means++ initialization. This step permits to obtain a robust classification that does not vary through the random choice of the initial centroïds, but is still based on unsupervised techniques. All the clustering results presented in the following section are computed that way.

4. Results

4.1. Clustering of CYGNSS reflectivity

We clustered the CYGNSS $\Gamma_{90\%-50\%}$ parameter with a number of classes *k* ranging from 2 to 8, in order to empirically determine the optimal number of clusters. For all the configurations, we plotted the maps of labelled pixels and we calculated a confusion matrix between the CYGNSS clusters and a reference set of water and inundation classes derived from RFWs maps (Tootchi et al., 2019). The definition of the reference classes is detailed in Table 4. Open water is grouped into a single class with a threshold at 10% corresponding to the empirical breaking point of permanent water (class A) vs. dry land (class E). Other classes, *i.e.* high floods (class B, maximum surface water extent SWE $\geq$ 80%), medium floods (class C, 40% $\leq$ SW < 80%) and low floods (class D, 5% $\leq$ SW < 40%), correspond to different levels of maximum inundation extent per pixel.

For a low number of clusters (*i.e.*, $k = 2$ or $k = 3$), the clustering problem was highly simplified as the separation of flooded vs. dry areas, according to the average values of $\Gamma_{90\%-50\%}$ related to the intensity of inundation. The results with $k = 4$ are more interesting, as the additional cluster represents the pixels in

Table 4: Definition of the reference classes of open water and floods with the RFWs dataset (Tootchi et al., 2019).

Reference class	Name / interpretation	Open water %	Seasonal water %
A	Permanent water	$\geq 10\%$	-
B	High floods	$< 10\%$	$\geq 80\%$
C	Medium floods	$< 10\%$	$\geq 40\%$ and $< 80\%$
D	Low floods	$< 10\%$	$\geq 5\%$ and $< 40\%$
E	Non flooded areas	$< 10\%$	$< 5\%$

floodplains with a strong seasonal cycle during the year. Finally, the results with $k \geq 5$ show an increasing confusion between one or several clusters associated to both seasonal flood patterns and permanent water. The choice of $k = 4$ as the optimal configuration was based on both statistical outputs (confusion matrices), and graphical interpretation of the clustering results.

We present the confusion matrix and the results of the clustering with $\Gamma_{90\%-50\%}$ and $k = 4$ in Table 5 and Figure 6, respectively. The output clusters are referred as C1, C2, C3 and C4 from the lowest to the highest average $\Gamma_{90\%-50\%}$ values. The map of labelled pixels is shown in Figure 6.a, while Figure 6.b-j represent the average time series of $\Gamma_{90\%-50\%}$ and $\Gamma_{90\%}$ for the pixels classified in C3 and C4, over 9 large river basins. The cluster C4 (in blue in Figure 6) exhibits high values for both $\Gamma_{90\%-50\%}$ and $\Gamma_{90\%}$ over the entire time series, associated to a seasonality in phase with variations of inundation extent. The pixels in C4 are located on lakes, river streams such as the Amazon, Parana, Niger and Congo Rivers, and also some large floodplains with permanent water or high SM throughout the year. This cluster represents only 4.4% of the total labelled pixels, but includes 58.8% of the reference class A (open water), 32.7% of class B (high floods) and 19.0% of class C (medium floods), according to the results in Table 5 (percentage values in red).

Table 5: Confusion matrix between reference classes A to E defined in Table 4 and CYGNSS clusters C1-4 shown in Figure 6. For each cell in the table, the proportion of pixels in this cell vs. the total number of pixels from the reference class (column Total) is shown in red, and the proportion vs. the total number of pixels from the CYGNSS cluster (line All) is shown in blue. As an example, the box A / C1 contains 9.6% of the total pixels in reference class A, and 0.2% of all the pixels in CYGNSS cluster C1.

	C1	C2	C3	C4	Total
A	804 (9.6%, 0.2%)	650 (7.8%, 0.4%)	1993 (23.8%, 3.3%)	4911 (58.8%, 15.5%)	**8358** (1.2%)
B	2150 (9.5%, 0.4%)	3861 (17.1%, 2.6%)	9150 (40.6%, 15.2%)	7353 (32.7%, 23.3%)	**22514** (3.1%)
C	8037 (21.9%, 1.7%)	10620 (28.9%, 7.1%)	11121 (30.3%, 18.4%)	6980 (19.0%, 22.1%)	**36758** (5.1%)
D	41226 (43.3%, 8.5%)	28782 (30.2%, 19.2%)	16963 (17.8%, 28.1%)	8211 (8.6%, 26.0%)	**95182** (13.2%)
E	430066 (76.7%, 89.2%)	105628 (18.8%, 70.6%)	21092 (3.8%, 35.0%)	4160 (0.7%, 13.2%)	**560946** (77.5%)
All	**482283** (66.6%)	**149541** (20.7%)	**60319** (8.3%)	**31615** (4.4%)	**723758**

Then, the cluster C3 in dark green shows variations from low to high values of $\Gamma_{90\%-50\%}$ and $\Gamma_{90\%}$, with a seasonal trend particularly visible in $\Gamma_{90\%}$. The pixels in C3 are located on the major floodplains in the pan-tropical area: the Llanos de Orinoco, Llanos de Mojos, Rio Branco, Pantanal and Parana floodplains in South America, the Inner Niger Delta (IND), Lake Chad and along the Nile in Africa, the Ganges-Brahmaputra, Indus, Irrawaddy, Yangtze and Mekong basins in Asia. The seasonality in C3 exhibits maximum values of $\Gamma_{90\%-50\%}$ and $\Gamma_{90\%}$ during summer in the Orinoco (Figure 6.b), Niger (Figure 6.e), Lake Chad (Figure 6.f) and Ganges-Brahmaputra (Figure 6.h) basins, all located in the northern hemisphere. On the contrary, in the Amazon (Figure 6.c, across the equator) and Parana (Figure 6.d, in the southern hemisphere) basins, the maximum values

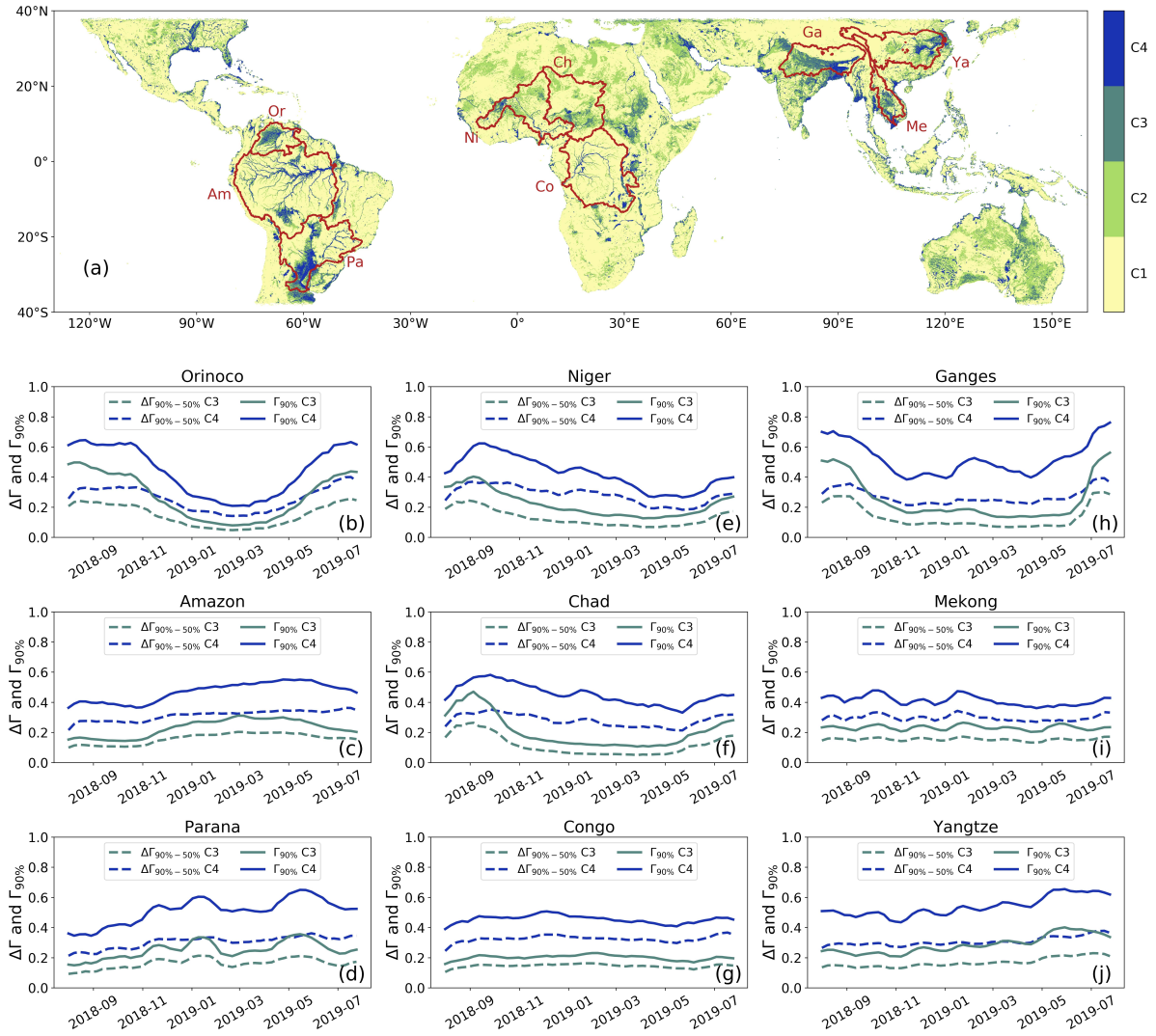

Figure 6: Clustering of $\Delta\Gamma = \Gamma_{90\%-50\%}$ with 4 clusters. **(a)** Map of the labelled CYGNSS pixels using the K-means / DTW algorithm with $k = 4$. **(b-j)** Average time series of $\Gamma_{90\%-50\%}$ and $\Gamma_{90\%}$ for clusters C3 and C4 in 9 large river basins: **(b)** Orinoco, **(c)** Amazon, **(d)** Parana / La Plata, **(e)** Niger, **(f)** Lake Chad, **(g)** Congo, **(h)** Ganges-Brahmaputra, **(i)** Mekong, and **(j)** Yangtze. The boundaries of all the 9 basins are superposed with the map of labelled clusters in **(a)**, with the two first letters of the basin name allowing its identification.

in C3 for $\Gamma_{90\%-50\%}$ and $\Gamma_{90\%}$ are obtained during spring. This validates the use of the DTW similarity measurement to cluster $\Gamma_{90\%-50\%}$ time series, for the extraction of flooded areas with distinct temporal cycles. For statistical results, C3 represents 8.3% of the total labelled pixels, and includes 23.8% of the reference class A, 40.6% of class B, 30.3% of class C, and 17.8% of class D (low floods) according to Table 5 (values in red). If we merge the two clusters C3 and C4, their combination represent only 12.7% of the world pixels, but 82.6% and 73.3% of the reference classes A and B, respectively. Thus, the identification of open water and flood signatures in CYGNSS reflectivity is highly reliable.

The cluster C2 has medium values of $\Gamma_{90\%-50\%}$ (not shown) corresponding to either water bodies and pixels with a fraction of water but not totally flooded (28.9% and 30.2% of classes C and D, respectively), or non-flooded regions from the reference class E. In Figure 6.a, it is possible to locate C2 mainly over bare soils in arid regions such as the Sahara, the Arabian Peninsula and Australia. Finally, the cluster C1 is dominant (66.6% of the labelled

pixels), and is interpreted as non-flooded areas although it includes about 9.5% of the pixels from both reference classes A and B.

In fact, some confusions are identified in the clustering despite a strong sensitivity to the presence of water. First, the pixels from reference class E which are theoretically non-inundated, represent 13.2% and 35.0% respectively (values in blue in Table 5) of the clusters C4 and C3, exhibiting the highest $\Gamma_{90\%-50\%}$ values. Then, the cluster C2 is a mix of pixels located over flooded and non-flooded regions. This is mainly due to a strong specular scattering over rocky bare soils (see the Section 5.4, discussion for more details). And lastly, while the open water and high floods (classes A and B) are well detected, it is not the case for the medium and low floods (classes C and D). They are associated to a mixing of wetlands and dry areas, and are quite equally divided into the 4 CYGNSS clusters. All these misclassifications can be related to: i) the influence of other factors such as the vegetation, SM, elevation, topography and small-scale roughness in the GNSS-R signals over land, ii) the heterogeneity of land cover and flood occurrence inside a CYGNSS pixel at 0.1° spatial resolution, and iii) errors in the definition of the reference classes, either due to the choice of our thresholds or to the high uncertainties associated to the RFWs product from Tootchi et al. (2019) itself (see Section 2.2.1 for more details).

4.2. Comparison with ancillary datasets

We have used CCI Land Cover maps and Tootchi's RFWs dataset described in Section 2 to investigate how the land cover changes and the occurrence of water influence CYGNSS-derived clusters. In Figure 7, these parameters are analyzed over three complementary study areas: the northern part of South America, between 20°S and 10°N, encompassing the Amazon, Orinoco, Tocantins and São Francisco river basins, as well as the Pantanal floodplains, the Titicaca and Poopó lakes (Figure 7.a1-5); Western and Central Africa around the gulf of Guinea, between 5°S and 20°N, including the Niger River Basin, Lake Chad and the Cuvette Centrale of Congo (Figure 7.b1-5); the Indian subcontinent including the Ganges-Brahmaputra and the Irrawaddy basins (Figure 7.c1-5). These regions include all types of land cover present in the pan-tropical area, *e.g.* equatorial forests, floodplains under canopy layers or with herbaceous covers, croplands irrigated or not, savannas, bare soils, high mountain ranges.

There is a good correspondence between the DTW clusters C3 and C4 in Figure 7.a1-c1 corresponding to areas with high values of $\Gamma_{90\%}$ and $\Gamma_{90\%-50\%}$, and the reference map of water extent derived from Tootchi et al. (2019) in Figure 7.a2-c2. The delineation of permanent water bodies and in particular the streams of the Amazon, Congo and Niger Rivers as well as their tributaries is clear, mainly in cluster C4. The most extensive floodplains are also detected in C3 and C4. In South America, we identify the Llanos in the Northeast Orinoco Basin (around 7°N and 70°W), Llanos de Mojos in the Southwest Amazon Basin (15°S and 65°W), the Rio Branco floodplain in the northern Amazon Basin (0° and 63°W), and the Pantanal floodplains (18°S and 58°W). In Africa, the IND (15°N and 5°W) and the floodplains around Lake Chad (10°N and 15°E) are well detected, but the Cuvette Centrale of Congo (0° and 17°E) is not. In the Indian subcontinent, a strong seasonal signal is obtained and mainly associated to C3 over the irrigated croplands (land cover class LC3 in cyan) along the Ganges plain. The Ganges-

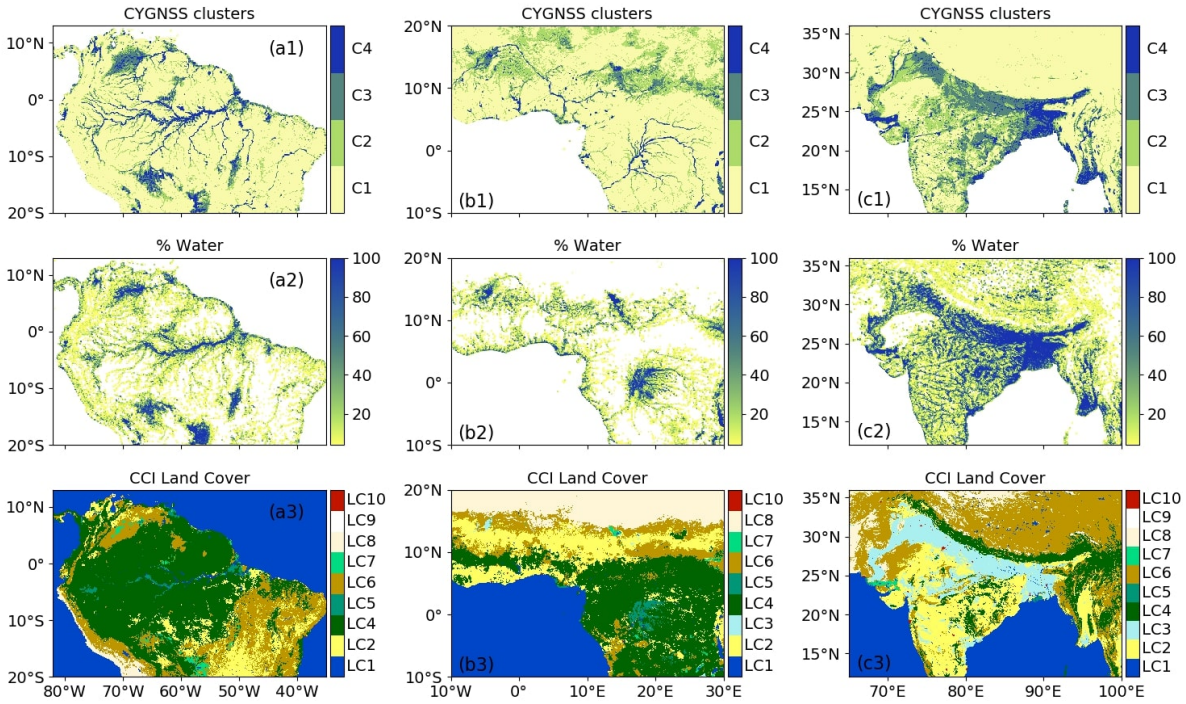

Figure 7: Comparison of CYGNSS-derived parameters with ancillary datasets over the northern part of South America **(a1-3)**, Western and Central Africa with the Gulf of Guinea **(b1-3)**, and the Indian subcontinent **(c1-3)**. **(a1, b1, c1)** CYGNSS clusters, **(a2, b2, c2)** Percentage of water from Tootchi's RFWs maps (Tootchi et al., 2019), **(a3, b3, c3)** CCI Land Cover with the 10 aggregated classes defined in Table 1

Brahmaputra delta, the Irrawaddy River and their tributaries are also well delineated. The interesting point is that CYGNSS reflectivity maps show a good ability for monitoring heavy seasonal floods, almost regardless of the LC types. Yet, the densest canopies cause a strong attenuation or a diffusion of the L-band signal. Over the Cuvette Centrale of Congo (bottom right in Figure 7.b1-3), we notice that CYGNSS likely underestimates the presence of floods nearby the streams of the Congo and Ubangi Rivers, if we refer to the RFWs dataset. Also, several lakes and wet areas in the Tibetan plateau in the upper right corner of Figure 7.c1-3 are not even detected using CYGNSS. The estimation of the specular point location is of poor confidence over the Tibetan plateau, which is about 5000 m high (Gleason et al., 2020), thus CYGNSS DDMs are not centered on the specular point and likely contain only thermal noise and diffuse scattering from surrounding areas.

The same datasets are analyzed in Figure 8 but on smaller regions, with a focus on some significant water bodies. Figure 8.a1-5 is centered on Lake Titicaca (3812 m high), Figure 8.b1-5 on Lake Victoria (1133 m high), Figure 8.c1-5 on the Lower Mekong Basin including Tonle Sap, the largest lake of south east Asia, and Figure 8.d1-5 on Lake Chad (both at low elevation). Several indications must be highlighted in Figure 8 . First, all water bodies are detected using CYGNSS, and mostly classified in C3 or C4 by the DTW clustering. These classes include a part of Lake Titicaca, Lake Poopó and the Salar de Uyuni located on the Altiplano at high elevation. Although the estimation of specular point location in this region has low confidence, a strong returned power is often observed with coherent reflection conditions. Then, it appears that the Tonle Sap and Lake Chad are classified in C4, while the surrounding floodplains are mostly classified in C3. This shows an example of the separation of permanent water vs. seasonal floods using CYGNSS. The Delta of Mekong mainly belongs to C4, likely due

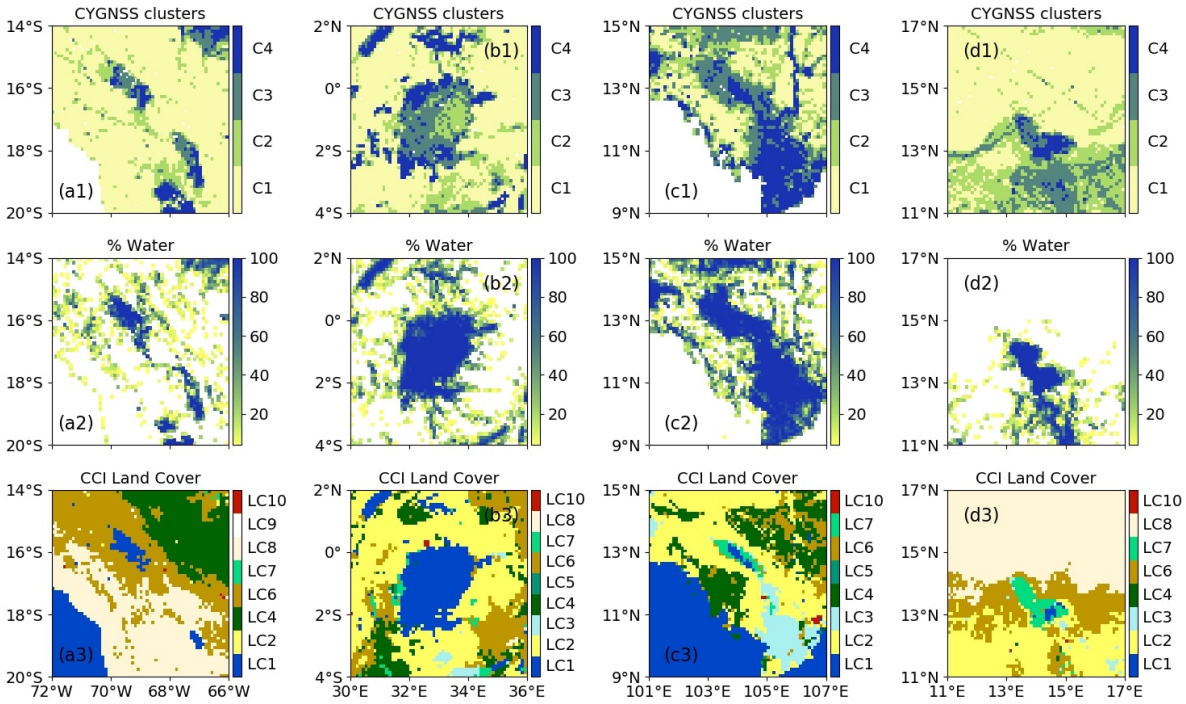

Figure 8: Comparison of CYGNSS-derived parameters with ancillary datasets over small subsets centered on water bodies. **(a1-3)** Portion of the Andes centered on Lake Titicaca, **(b1-3)** Lake Victoria, **(c1-3)** Delta of Mekong with the Tonle Sap, and **(d1-3)** Lake Chad and a portion of Sahel. The same datasets are used than in Figure 7

to continuous flooding or wet conditions in the paddy fields (Kuenzer et al., 2013). Thirdly, in Lake Victoria and Lake Titicaca, a spatial heterogeneity is observed with higher reflectivity (associated to C3 and C4) observed closer to the banks. This is attributed to the effect of the winds on large lakes and was previously reported in the literature, especially over Lake Victoria (Al-Khaldi et al., 2021b). The roughness of water surfaces due to the wind waves can nullify the coherent scattering assumption, and lead to a decrease of surface reflectivity (Chew & Small, 2020).

The CYGNSS reflectivity also depends on the dominant type of land cover inside a pixel. Figure 9 presents the distribution of CYGNSS clusters into the aggregated LC classes defined in Table 1, either as the total number of pixels (Figure 9.a) or the percentage (Figure 9.b) in each class. Only pixels with a dominant type of LC covering at least 80% of the pixel are considered to avoid an influence of LC mixing. First, we notice a strong dominance at the global scale of non-flooded LC types (bare soils, dry shrubs and herbaceous, non-flooded forests, croplands non irrigated), with fewer pixels affected by permanent or seasonal water. These dry classes are composed quasi-exclusively of the CYGNSS clusters C1 and C2, with the lowest reflectivity values, and include also a part of C3, with high seasonal variations of $\Gamma_{90\%}$ and other CYGNSS-derived parameters. Then, we make several assessments over flooded areas. The flooded forests are mostly composed of C1 and C2, while the flooded shrubs and herbaceous are mainly composed of C3 and C4, with higher $\Gamma_{90\%}$ and $\Gamma_{90\%-50\%}$ values and seasonality (Figure 6). In densely forested areas, the L-band reflected signals recorded by CYGNSS can be either strongly attenuated on their way to or from the reflecting surface, or scattered by the top of the canopy. In both cases, the ability of CYGNSS to detect inundations below the canopy is affected. The irrigated croplands, mainly located

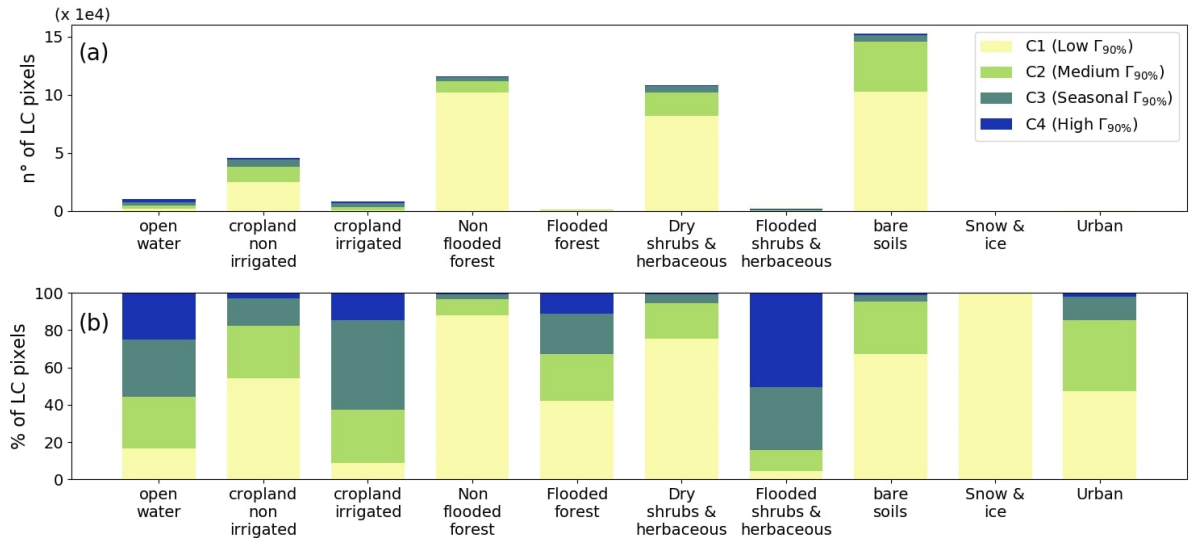

Figure 9: Repartition of the CYGNSS clusters into the main Land Cover types (see Table 1 for the aggregation of CCI LC classes). **(a)** Number of pixels from each cluster, **(b)** Percentage of pixels from each cluster.

in the paddy fields in Southeast Asia, are dominated by the pixels with a strong seasonal signal in C3. On the other hand, the croplands non-irrigated are dominated by the dry pixels in C1. Both also contribute in the cluster C2, highlighting a mixing between these two classes and some misclassifications in Table 5. Non-flooded croplands and herbaceous covers affected by large seasonal rainfall present high SM values, responsible for a high CYGNSS reflectivity. Finally, open water bodies are divided between low and high reflectivity clusters. This should be linked to: i) the attenuation or scattering of the signals in equatorial forests and the non-detection of small streams, and ii) lower returned signal power on large windy lakes as shown in Figure 8.

4.3. Comparison against flood products at regional and global scales

In this subsection, we analyze the correlations between CYGNSS-derived parameters and flood reference products. Because the spatial resolution of GIEMS is 0.25°, CYGNSS has also been gridded and clustered into an equivalent 0.25° grid for further analysis. We only present the results for Γ_{median} which show the best correlations, as $\Gamma_{90\%}$ is sometimes saturated with a fraction of water inside the pixel. The relationship between CYGNSS reflectivity and the fraction of water in the footprint is not linear (Chew & Small, 2020). Thus, we evaluated both Pearson's linear and Spearman's rank correlations. The latter evaluates whether the two variables are linked with a monotonic function, and should perform better for non-linear relationships. In fact, the results are slightly better using Spearman's R, in particular for the regional comparisons against MODIS dataset (Section 4.3.2). Spearman's temporal correlations between Γ_{median} and the dynamic inundation maps at global and regional scales are therefore presented in Section 4.3.1 and Section 4.3.2, and in Figure 10-11. Finally, the spatial correlations against RFWs static inundation maps are presented in Section 4.3.3.

4.3.1. Global comparison of CYGNSS and GIEMS

Figure 10 presents the Spearman's correlation coefficients in every pixel between CYGNSS Γ_{median} and GIEMS Surface Water Extent (SWE) at 0.25° (Figure 10.a). A strong correlation (R > 0.8) is obtained in most of the major floodplains located in CYGNSS's spatial coverage, where the seasonality drives both signals due to the high temporal variations of SWE. It is the case in Southeast Asia and India, characterized by the annual monsoon events and irrigated paddy fields in the Ganges-Brahmaputra, Irrawaddy, Mekong and Yangtze basins. In South America also, the seasonal rainfalls cause large floods in Llanos de Orinoco, in the Amazon basin (including Llanos de Mojos and the Branco River), or in La Plata basin (including the Pantanal wetlands). On the contrary, poor results are obtained along some of the Amazon and Congo tributaries, in the upstream parts of these basins. This can be attributed to either the vegetation effect on Γ_{median}, or to a lower seasonality in both signals due to a continuous, high SWE, leading mechanically to a lower correlation. Low or even sometimes negative correlations are also observed in some coastal areas. The surface water estimations from GIEMS can be contaminated by the ocean contribution, although it is expected to be filtered out. Finally, negative correlations are also observed over some land areas (Figure 10). We had a closer look at two specific targets in East Asia. These regions mostly comprise non irrigated croplands and herbaceous land cover, and they showed both a seasonal SWE cycle with low amplitude in GIEMS, and a low signal in CYGNSS with episodic variations. GIEMS likely overestimates SWE in these areas during the wet season, while CYGNSS peaks may be a response to SM or short duration floods following intense precipitation.

Figure 10.b presents the average of CYGNSS vs. GIEMS Spearman correlations over the same river basins than in Figure 6, for C3, C4 and all CYGNSS clusters. Cluster C3 in orange and C4 in red show an average correlation over 0.6 for most of the basins. The results in the Orinoco, Ganges, Niger, Lake Chad, Amazon and Parana basins are globally consistent because the water cycle is driven by seasonal floods and irrigation. Two main exceptions are the Congo and Mekong basins where the seasonal variations of reflectivity are very weak (below 0.1). The low correlations observed can be due to either a constant or random signal in both GIEMS and CYGNSS observables. In the Congo basin, the vegetation attenuates CYGNSS signals and provokes a lower detection of inundations below the canopy, and alternating rainfalls in the northern and southern hemispheres produce bimodal variations of inundation extent. In the Mekong basin, a continuous flooding or irrigation in paddy fields, including several parts of the Delta of Mekong (Kuenzer et al., 2013), certainly reduces the seasonality of the signals.

4.3.2. Regional comparison

The comparison with GIEMS SWE gives an indication of the correlation between CYGNSS reflectivity and flood dynamics at the global scale, but at the regional scale GIEMS is less adequate. First, it is gridded at 0.25° of spatial resolution, thus the information can be degraded when compared to CYGNSS 0.1° grid. Also, we gridded CYGNSS into at 0.25° for comparing against GIEMS, which is not the optimal spatial resolution of this product. Moreover, CYGNSS at 0.1° provides information about the flood dynamics up to a 7-day temporal resolution, and

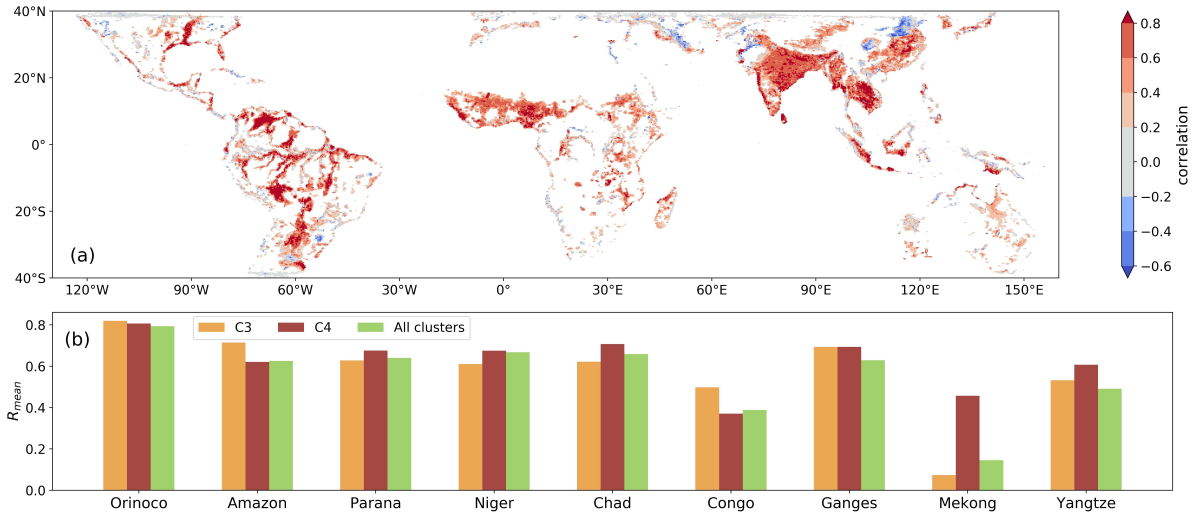

Figure 10: Spearman correlations between CYGNSS Γ_{median} at 0.25° spatial resolution and GIEMS-2 from August 2018 to July 2019. **(a)** Map of correlations over the flooded areas in GIEMS (maximal SWE > 20%), **(b)** Barplot of average correlations for CYGNSS clusters C3, C4 and all clusters in 9 river basins.

up to a daily time-scale at 0.25°, while GIEMS has a monthly time sampling. It also has uncertainties, especially in densely forested regions, where it is unclear whether flood signatures can be fully identified in the brightness temperatures at K-band from SSM/I. Finally, the large estimations of seasonal SWE in GIEMS, particularly in the Sahel region and the Indian subcontinent, may be overestimated due to the confusion between standing water and saturated soils during rainfall periods. For all these reasons, we also performed a regional comparison between CYGNSS Γ_{median} and the SWE derived from MODIS-based regional dynamic flood maps at 500 m spatial and 8-day temporal resolutions (see Section 2.2).

Figure 11 presents the results of this comparison over 3 regions: La Plata basin including the Parana, Uruguay and Paraguay rivers (Figure 11.a1-2), the Inner Niger Delta (IND) (Figure 11.b1-2), and the Lower Mekong Basin (LMB) including Tonle Sap and the Delta of Mekong (Figure 11.c1-2). Table 2 show that these regions have experienced above-average water level peaks during the study period, which are usually correlated with larger inundated areas. For every region, both maps of maximum flooded extent (as the percentage of water inside the pixel) and Spearman's correlation coefficients between the time series of SWE and Γ_{median} are shown. Only the correlations on pixels with a maximum flooded extent above 20% are presented. Over La Plata basin, we obtain medium (R > 0.4) to high (R > 0.8) values in the major flooded areas along the streams of the Parana and Paraguay rivers. In particular, the correlations are high in the Pantanal wetlands (∼58°W, 20°S), in the Paraguay River and its confluence with the Parana River, and in wetlands close the Salado River in the region of Santa Fé (∼61°W, 29°S). The values are lower (R ∼ 0.5) in the major part of the Parana flooded savannas. This could be due to continuous high reflectivity in this region affected by permanent floods or very high SM content. Low correlation values are obtained along the streams of the Uruguay and the upper Parana rivers, as well as in mountainous regions along the Andes.

The correlations over the LMB are heterogeneous, with high values over the floodplains and irrigated crop-

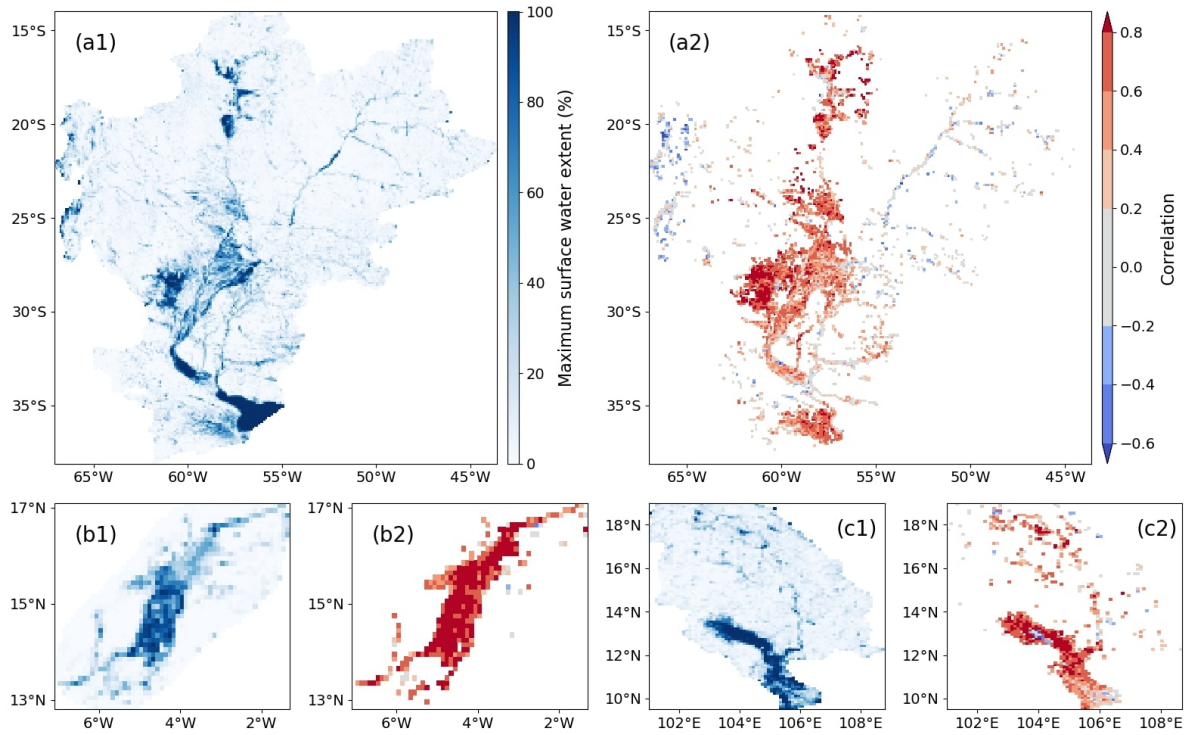

Figure 11: Correlation between CYGNSS Γ_{median} and SWE derived from MODIS over 3 flooded regions. **(a1-2)** show the maximum water extent in the year and the pixel-by-pixel correlations between water extent and Γ_{median} over La Plata basin, **(b1-2)** are the equivalent plots over the Inner Niger Delta (IND), and **(c1-2)** are the equivalent plots over the southern Mekong basin (including the Delta of Mekong and Tonle Sap).

lands near the Tonle Sap and along the stream of the Mekong River. On the contrary, low and even negative correlations are obtained over the Tonle Sap itself, in some parts of the Delta and in isolated pixels. Figure 12 present the time series of SWE and Γ_{median} in 4 neighboring pixels located near the Tonle Sap. The two left panel pixels (Figure 12.a and Figure 12.c) show a very high seasonality for both parameters, with a water extent varying from 0 to 100% during the year. It likely corresponds to seasonal floodplains, and the correlations calculated between the time series are around 0.9. On the contrary, the pixels plotted in Figure 12.b and Figure 12.d show limited variations in water extent throughout the year, with values systematically over 90% and 50%, respectively. They are mainly covered by permanent water from Tonle Sap. There, the variations of Γ_{median} are not correlated with variations in the inundation extent. Moreover, in all the pixels, we observe high fluctuations of Γ_{median} with an amplitude reaching ~ 0.4 and a period of one to several months. These fluctuations mostly occur when a large fraction of water is present inside the pixel. It could be linked to variations of water roughness over Tonle Sap, the greatest lake in southeast Asia. Moreover, similar time series of Γ_{median} with high amplitude fluctuations were found to be located close to the banks of Lake Victoria, where the direction and speed of winds can cause an alternation between coherent and incoherent scattering regimes. Unfortunately, we were not able to find wind speed data close enough to Tonle Sap to compare with Γ_{median} time series. The understanding of this phenomena could help further work on retrieving SWE using CYGNSS reflectivity or coherency. Other pixels in the Delta of Mekong show low correlation values, but large parts of this area are flooded or wet throughout the year (Kuenzer et al., 2013).

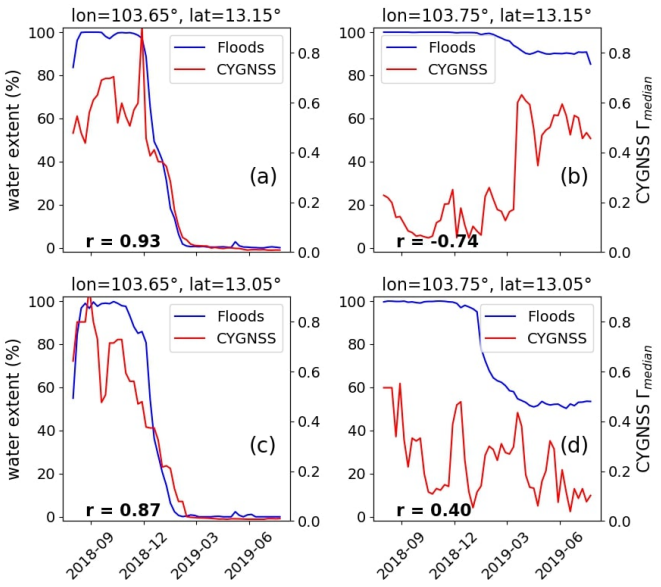

Figure 12: Time series of CYGNSS Γ_{median} (in red) and SWE from MODIS (in blue) with their correlation, for four pixels near the Tonle Sap in the Mekong basin.

Finally, the results over the IND show a very good consistency between the time series of SWE and Γ_{median}. The correlations are greater than 0.8 in most of the pixels. It indicates a strong correspondence between the dynamics of floods and CYGNSS reflectivity in this region. It is certainly due to two reasons: i) the low vegetation cover cause lower uncertainties in both CYGNSS observations and the reference inundation maps, and ii) a strong seasonality is observed in the precipitation and flooding events, so Γ_{median} is low during the dry season unlike other floodplains in the LMB and the Parana.

4.3.3. Spatial correlations at the regional scale

Spatial correlation coefficients were also calculated between Γ_{median} and the SWE derived from static RFWs maps (Tootchi et al., 2019), at the time of the maximum flooded extent during the year. They are estimated for the 9 river basins presented in Figure 6 and Figure 10, and values are reported in Table 6. The spatial correlations in the Orinoco, Amazon, Parana, Ganges-Brahmaputra and Yangtze basins are high, ranging between 0.74 and 0.77, showing a good correspondence between the reference maps and CYGNSS reflectivity at the flood maximum extent. On the contrary, spatial correlations are lower in the other basins and especially in the Sahel region (0.52 for Niger, 0.51 for Lake Chad). This highlights the contribution of multiple factors in the scattering of GNSS signals over these areas, and maybe uncertainties in the RFWs dataset. While the temporal correlations indicate whether the variations of SWE are linked with an increase in CYGNSS reflectivity, the spatial correlations are particularly sensitive to a saturation of both the reference and CYGNSS signals during the wet season. High SM can provoke high CYGNSS reflectivity, and also an overestimation of the fraction of water estimated in GIEMS, whose uncertainty is further propagated in the RFWs dataset (see Section 2.2.1for more details).

Table 6: Spatial correlations between CYGNSS Γ_{median} and the percentage of water derived from RFWs static inundation maps (Tootchi et al., 2019), in 9 river basins inside the area of coverage. The values of Γ_{median} are extracted at the time of the maximum flooded extent.

River basin	Orinoco	Amazon	Parana	Niger	Chad	Congo	Ganges	Mekong	Yangtze
Spatial R	0.76	0.77	0.75	0.52	0.51	0.58	0.75	0.66	0.74

5. Discussion

The objective of this study was to assess the potential of CYGNSS reflectivity for a pan-tropical mapping of flood dynamics. Based on the results presented in Section 4, we discuss several points to pave the way to a CYGNSS-based inundation product. To begin with, we take a broader view on the interests and limitations of our methodology, as well as the uncertainties in the reference datasets considered. We then discuss the attenuation of GNSS-R signals by the vegetation, especially over tropical forests, and the feasibility of flood detection under dense canopies using CYGNSS reflectivity. We also point out the misleading role of high forward scattering in the specular direction in arid areas, due to changes in morphology, *i.e.* roughness for CYGNSS, but also lithology (sand vs. rock, see Section 5.4).

5.1. Results of the K-means Clustering

We have used the K-means clustering algorithm with a Dynamic Time Warping similarity measure to perform the clustering of CYGNSS reflectivity time series. Other common unsupervised classification algorithms were tested such as Agglomerative Clustering or Birch, but the best results were obtained with K-means. The implementation of a DTW similarity measure instead of a simple Euclidean distance makes it more robust to shifts in phase and distortion in time between time series of reflectivity. Promising results were obtained and analyzed in this study, but the empirical approach led us to discuss our choices in the implementation of the methodology.

First, the optimal number of clusters needed to be determined. We were not able to calculate any criteria like the Calinski-Harabasz and the Silhouette scores with the implementation of the DTW similarity measure. We also tested our methodology with a common Euclidean distance in the K-means approach, which gave us an optimal number of 2 clusters with both the Calinski-Harabasz and the Silhouette scores. This is a highly simplified version of the phenomena as dry land vs. flooded areas, and does not give information on the seasonality of floodplains. Thus, we empirically determined the optimal number of clusters, which was found to be k = 4 (Figure 6 and Table 5). The two clusters with highest reflectivity are mostly associated to the occurrence of water, one with permanent or long-lasting floods, the second with high seasonal variations. A third cluster is composed of a mix of wetlands and bare soils with medium reflectivity, and can be especially sensitive to an increase of CYGNSS reflectivity over non-flooded areas due to high SM content. The last cluster is mainly associated to dry soils, never flooded throughout the year. We also tried the clustering with several combinations of parameters from Figure 4: 1) Γ_{mean} only, 2) Γ_{median} only, 3) $\Gamma_{90\%}$ only, 4) $\Gamma_{90\%-50\%}$ only, 5) Γ_{median} and $\Gamma_{90\%}$, 6) Γ_{mean} and Γ_{std}, 7) Γ_{mean}, Γ_{std}, $\Gamma_{90\%}$ and Γ_{median}. All the confusion matrices and the plots (similar to Table 5 and

Figure 6, respectively) were evaluated for each version of the clustering, leading to the determination of the best scenario.

To ensure the stability of our clustering, we performed a sensitivity analysis on 30 iterations of the K-means++ / DTW clustering. It was shown that the 30 results are quite stables, especially in terms of inertia (see Table 3). However, the random (although weighted) choice of the initial 4 centroids by the K-means++ algorithm leads to differences in the final results. In 25 results out of 30, very small differences are observed in the output labels, especially for clusters C3 and C4 associated to water bodies (see Figure 6). The last 5 versions show lower agreement with each-other. The correspondence (*i.e.* percentage of pixels similarly labelled between two versions of the clustering) has a minimum of 80.26%, and a median of 94.64% (Table 3). As a consequence, we adapted our methodology to guarantee the reproducibility of the clustering. The mean time series of cluster centers C1 to C4 were computed from the 30 random DTW outputs, and selected as user-defined centroids for a new K-means / DTW clustering. This final version was analyzed in Section 4.

Finally, the DTW was implemented to manage shifts in time between the flooding events in different regions of the world. However, it has a limitation due to the boundary condition. The DTW algorithm starts with the first index and finishes with the last index of each time series that are compared. As a consequence, the values of $\Gamma_{90\%-50\%}$ and other CYGNSS parameters at the boundaries of the time series play a determinant role for the clustering. We use only one year of CYGNSS observations because of the high computational capacities required to process and analyze this dataset, and due to the experimental nature of our methodology whose interest needed to be proven before a global application to the 5-year CYGNSS dataset. To avoid the boundary effect, we implemented a padding of the 1-year time series of our parameters to create artificially a 3-year dataset. The results prove that this technique does not affect the ability of the K-means / DTW clustering to extract a seasonal flood signal from both the wetlands in the northern (*e.g.* Orinoco, Rio Branco, Ganges, Yangtze, Mekong) and southern (*e.g.* Llanos de Mojos, Pantanal, Parana) hemispheres.

5.2. Reference inundation maps

The reference datasets used in this study are also sources of uncertainties. The static inundation maps from Tootchi et al. (2019) were used as a delineation of permanent water bodies and regularly flooded areas. We first tried to use the composite wetlands (CWs) maps as it contains both the regularly flooded wetlands (RFWs) and the groundwater-driven wetlands (GDWs). It seemed logical to us that GDWs also play a role in the scattering of GNSS signals, as they are an important source of SM in saturated soils. However, we noticed a strong saturation of the CWs maps in several river basins. As an example, the entire Amazon basin and a great part of Sahel were listed as partially or totally affected by floods, which is largely exaggerated. Also, spatial correlation coefficients calculated with CYGNSS Γ_{median} show higher values with RFWs maps both at the global and regional scales. For the Amazon basin, the spatial correlation is 0.77 between Γ_{mean} and the SWE from RFWs maps (Table 6), and is 0.48 with CWs maps. As a consequence, we decided to use the RFWs maps for comparison with CYGNSS datasets. Additionally, the thresholding of RFWs variables into the 5 reference classes defined in (Table 4) is

debatable. The open water class (A) corresponds to pixels covered with, at least, 10% of open water. When plotting Γ_{mean} as a function of open water extent, the Γ_{mean} values are saturated above this threshold. On the contrary, the function of Γ_{mean} vs. inundation extent is linear, with no saturation of the signal when increasing the fraction of flooded areas in the pixel. The separation of reference classes B, C and D is then arbitrary, with thresholds corresponding to high, medium or low inundation extent in the pixel.

The uncertainties of GIEMS are mainly associated to its low spatial resolution (0.25°), leading to a low sensitivity to small water bodies. It is also influenced by dense vegetation covers, and high seasonal SM content confounded with surface water in several regions (Sahel, Indian subcontinent). These are sources of uncertainty for our analysis because even Tootchi's RFWs dataset is based on GIEMS-D15 (Fluet-Chouinard et al., 2015), derived from GIEMS itself. Still, the comparison of CYGNSS vs. GIEMS and MODIS estimated SWE looks consistent at the regional scale, although the spatial resolutions are different (0.25° for the first, 0.1° for the latter).

5.3. Reflectivity over vegetated areas

The detection of floods under equatorial forests remains difficult even with the use of microwave signals. It is a limitation of GIEMS (Prigent et al., 2020) for example. The use of CYGNSS data is promising because: i) L-band GNSS signals penetrate deeper the canopy than shorter wavelengths/higher frequencies in the microwave domain, ii) it provides information at a higher spatiotemporal resolution than the passive microwave sensors. Several studies took advantage of these characteristics. Rodriguez-Alvarez et al. (2019) classified CYGNSS corrected Signal-to-Noise Ratio (SNR) as flooded vegetation (FV), open water and dry land. This study was performed in a small subset of the Amazon basin and showed a detection of FV around 70%. Then, Carreno-Luengo et al. (2020) studied the relationship between CYGNSS observables, including the reflectivity Γ, and AGB over subsets of the Congo and Amazon basins. A polynomial fit was used to characterize the vegetation attenuation at several incidence angles, and to further derive CYGNSS-based maps of AGB over equatorial forests. These maps were consistent with reference datasets up to ~ 350 Mg/ha. Also, Li et al. (2021) derived the coherency of Beidou-3 raw IF tracks over flooded areas along the Mississippi River, in densely vegetated areas. High coherency, which is linked to the presence of water below the canopy, was obtained for AGB up to 200-300 Mg/ha. And finally, the PR defined in Al-Khaldi et al. (2021a,b) as a coherency proxy has shown low sensitivity to vegetation, allowing the detection of small river streams obscured by trees.

In Figure 6.a and Figure 7, the streams of the Congo and Amazon rivers and their tributaries are well delineated. We can still question whether CYGNSS is able to monitor nearby floodplains under vegetation cover. Figure 13 presents the distribution of AGB in CYGNSS clusters C1 to C4, for both flooded and non-flooded pixels. All pixels with an occurrence of water greater than 0% in the 0.1° regridded RFWs dataset are considered as flooded, which likely overestimates the inundated areas. Only pixels in South America are considered in this figure, to obtain a balance between flooded and dry pixels, and between high and low AGB values. The dynamic range of AGB is approximately 0-300 Mg/ha and shows two peaks for all clusters. The first one, between 0 and 50 Mg/ha, corresponds both to dry pixels with few vegetation, and to large floodplains with herbaceous-type

land cover, as in the Llanos de Orinoco, Llanos de Mojos, and the Pantanal wetlands (see Figure 7 for dominant land cover types in South America). The second one, between 200 and 300 Mg/ha, corresponds to dense forests mainly located in the Amazon basin. These forests are either dry (cluster C1), wet or flooded (clusters C2 to C4). In particular, flooded pixels in C3 and C4 mostly represent the large floodplains and open water (see Figure 6, Figure 7 and Table 5), with low AGB. However, the fourth decile (*i.e.* the top 25% of AGB values) in flooded pixels from C3 and C4 ranges from ∼150 Mg/ha to ∼300 Mg/ha. The results in Figure 13 show that the K-means / DTW clustering based on CYGNSS reflectivity is able to detect either flooded areas below the canopy or narrow river streams surrounded by vegetation, in pixels with dense forests and average AGB as high as 250-300 Mg/ha. Some sources of uncertainties can be related to: i) signals from wet, saturated soils without standing water, ii) errors in the definition of flooded areas using RFWs dataset, and iii) the attenuation of GNSS-R signals by the vegetation, especially at high incidence angle, leading to a potential extinction of the coherent component on which we base our analysis (Al-Khaldi et al., 2021a; Loria et al., 2020). The results presented in Figure 9 show a lower detection of the inundations under forest land cover types, when compared to an herbaceous cover.

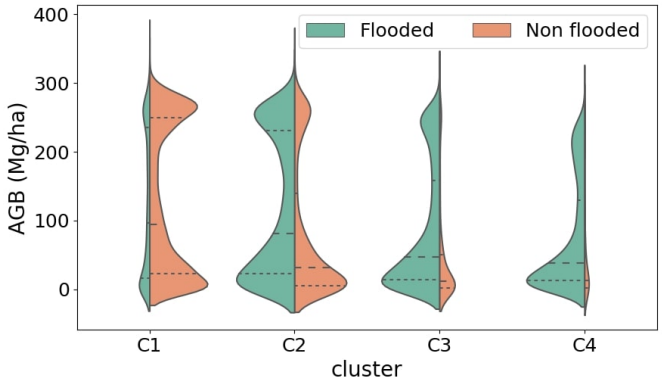

Figure 13: Distribution of the AGB for flooded and non-flooded pixels over South America in the 4 CYGNSS clusters. The dashed lines represent the quartiles of the distributions. Flooded pixels are defined with a fraction of water greater than 0% in the 0.1° regridded RFWs dataset.

5.4. Reflectivity over bare soils

In Figure 5 and Figure 6.a, we notice the heterogeneity of $\Gamma_{90\%}$, $\Gamma_{90\%-50\%}$ and CYGNSS clusters over arid regions, where bare soils is the dominant land cover type. In particular, reflections over deserts in the Sahara and the Arabica Peninsula have a high dynamic range, with $\Gamma_{90\%}$ values ranging from 0 to ∼0.4. As the time series of $\Gamma_{90\%}$ and $\Gamma_{90\%-50\%}$ are almost constant in the year due to a very low moisture, these areas can be interpreted as open water bodies, in particular when the reflected signals over the latter are attenuated as for wind conditions over lakes (see Figure 8). An important forward scattering of GNSS-R signals has already been observed in the literature, and mainly linked to areas of low small-scale roughness. In particular, the water mask given by Al-Khaldi et al. (2021a) also had false alarms over desert regions characterized by (almost) flat surfaces, based on their PR coherency metric. They handled the problem with a filtering of the locations showing a recurrent coherence, and a 0% occurrence of water in the mask from Pekel et al. (2016). This type of filtering

could be easily applied to our reflectivity dataset, and is a requirement for a further inversion of a fractional water extent product.

The state of the art beyond the scope of CYGNSS-related studies shows that anomalies in the radar signals over arid regions are well known for several active and passive sensors, and can be related to the geological and lithologic properties. Prigent et al. (2015) analyzed the multiangle backscattering at Ku-band from Tropical Rainfall Measurement Mission / Precipitation Radar (TRMM / PR) and QuikSCAT in the inter-tropical zone covered by CYGNSS. They showed that sand dunes have lower backscattering coefficients when compared to rock deserts, because of an important volume scattering in dry sand and a relatively low large-scale roughness. Similar effects were also observed at C-band using ASCAT (Fatras et al., 2015) . This phenomenon increases with an increasing incidence angle. Similarly, the microwave emissivities over arid regions show higher values for sand dunes due to the contribution of both the surface and deeper soils to microwave emissions (Jiménez et al., 2010). GNSS-R forward scattering is likely subjected to the same phenomena, as the spatial agreement between $\Gamma_{90\%}$ and $\Gamma_{90\%-50\%}$ maps and TRMM / PR-based maps of backscattering coefficients in Prigent et al. (2015) is good.

5.5. Guidelines for a CYGNSS-based dynamic product of SWE

This study shows that both permanent water and seasonal floodplains can be monitored using CYGNSS reflectivity. In particular, the 0.1° spatial resolution and 7-day temporal resolution grid merges by time and location the bistatic observations, and offers the possibility to compute time series of several statistical parameters (see Section 3). In particular, $\Gamma_{90\%}$ is very sensitive to the presence of a fraction of water, and $\Gamma_{90\%-50\%}$ can be used to extract pixels with recurrence or occurrence of water. Also, the comparison of Γ_{mean} with the fraction of water in the pixel, either from static or dynamic estimations, shows an interesting linear trend. This linearity could be further exploited to produce a CYGNSS-based SWE product, covering the pan-tropical area.

For this, the confounding effects of several geophysical parameters need to be removed. False alarms over deserts can be avoided with a simple filtering as in Al-Khaldi et al. (2021a). Then, a correction of reflectivity for the vegetation attenuation must be implemented, which is usually performed using the incidence angle and ancillary L-band Vegetation Optical Depth (L-VOD) data provided by radiometers such as SMOS (Wigneron et al., 2021) or SMAP (Konings et al., 2017; Li et al., 2022b). Also, GIEMS-2 (Prigent et al., 2020), assumes the linearity of SWE retrieval under given vegetation conditions, and the dataset is therefore binned depending on the values of vegetation parameters. Based on this idea, CYGNSS dataset could be binned with either NDVI, AGB or VOD to improve the linearity of Γ_{mean} vs. SWE relations. The effect of small-scale and large scale roughness on the performances of future SWE product could be important to study. Water roughness over large lakes is a limitation due to a dominant incoherent scattering regime, so it must also be taken in account. As we focus on dynamic estimations of SWE in floodplains, recurrent open water bodies could be filtered out as they are well delineated using other remote sensing sensors. Finally, the confounding effect of SM can lead to an overestimation of flooded areas in regions affected by large seasonal rainfall. Further investigations could use

the coherence proxy defined in Al-Khaldi et al. (2021a,b), in helping to filter out the pixels with strong reflectivity that is associated to very wet soils with no surface water.

6. Conclusion

We have analyzed CYGNSS land surface reflectivity to evaluate its potential for a global mapping of flood dynamics. A 0.1° spatial resolution and weekly time sampling of CYGNSS reflectivity was found to be the most suitable compromise between high spatial and temporal resolutions. We have used a K-means clustering technique with DTW similarity measure to separate: i) the low, constant reflectivity signals from dry land, ii) the high signals from open water and long-lasting inundations, and iii) the seasonal signals associated to large floodplains. Static and dynamic inundation maps along with other ancillary datasets were used to analyze the clustering results. The largest water bodies, floodplains and irrigated croplands areas are detected. Various sources of misclassification are identified. The flooded areas are mostly detected in absence of vegetation or under herbaceous cover, but less accurately under forests. The concordance between CYGNSS parameters and the static inundation maps is not good in the Cuvette Centrale of Congo in particular. The detection of water bodies is also weakened under windy conditions over large lakes, as it was already reported by several studies. In the deserts, a low reflectivity is observed on sand dunes where the penetration of microwave signals is high, and a high reflectivity is observed on flat, rocky regions where a specular scattering was found to create false alarms. However, CYGNSS clusters and parameters make possible to identify the main floodplains and open water areas, including samples in areas with AGB as high as $\sim$300 Mg/ha along the streams of Amazon and Congo rivers. The spatial correlations between Γ_{median} and static inundation maps were calculated at the time of the maximum SWE. They show high values (R ranging from 0.74 to 0.77) in 5 large river basins and lower values in Congo, Lake Chad and Niger, highlighting the contribution of multiple known factors (vegetation, soil moisture, small and large scales roughness, type of soil, SWE) in the GNSS-R forward-scattered signals. Temporal correlations were also calculated between Γ_{median} and SWE from either GIEMS or regional MODIS-based inundation maps. Regional comparisons over the IND, Parana and Mekong basins perform well and are consistent with global comparison using GIEMS. High values (R > 0.8) are obtained on the principal floodplains in the CYGNSS coverage, with once again the exception of the Cuvette Centrale of Congo. Lower correlations are obtained in the Amazon and Congo tributaries, either due to an attenuation of GNSS-R signals by vegetation canopies, or to the uncertainties of reference datasets in these areas.

Finally, our results show that a global mapping of inundation dynamics using CYGNSS reflectivity is possible, as some open water and flood patterns were clearly identified in the K-means / DTW clustering results. We tried to list the several sources of misclassification to identify contributing factors in CYGNSS reflectivity either at the global or regional scale. We look forward to producing a dynamic mapping of floods using CYGNSS, with ancillary and reference datasets helping to characterize the relation between CYGNSS reflectivity and SWE.

Acknowledgements

P. Zeiger is funded by a PhD grant from Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI). This work was supported by Centre National de la Recherche Scientifique (CNRS) and Centre National d'Etudes Spatiales (CNES) through INSU PNTS grant URGENS and TOSCA grants SWHYM and SCOMAG, respectively.

We thank all the reviewers for their helpful and detailed suggestions, to help us improve the revised version of this manuscript.

References

Acreman, M., & Holden, J. (2013). How Wetlands Affect Floods. *Wetlands*, *33*, 773–786. doi:10.1007/s13157-013-0473-2.

Al-Khaldi, M. M., Johnson, J. T., Gleason, S., Chew, C. C., Gerlein-Safdi, C., Shah, R., & Zuffada, C. (2021a). Inland water body mapping using cygnss coherence detection. *IEEE Transactions on Geoscience and Remote Sensing*, *59*, 7385–7394. doi:10.1109/TGRS.2020.3047075.

Al-Khaldi, M. M., Johnson, J. T., Gleason, S., Loria, E., O'Brien, A. J., & Yi, Y. (2021b). An algorithm for detecting coherence in cyclone global navigation satellite system mission level-1 delay-doppler maps. *IEEE Transactions on Geoscience and Remote Sensing*, *59*, 4454–4463. doi:10.1109/TGRS.2020.3009784.

Al-Khaldi, M. M., Johnson, J. T., O'Brien, A. J., Balenzano, A., & Mattia, F. (2019). Time-Series Retrieval of Soil Moisture Using CYGNSS. *IEEE Trans. Geosci. Remote Sensing*, *57*, 10.

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The Advantages of Careful Seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 1027–1035). Philadelphia, USA.

Bartlett, K. B., & Harriss, R. C. (1993). Review and assessment of methane emissions from wetlands. *Chemosphere*, *26*, 261–320. doi:10.1016/0045-6535(93)90427-7.

Bergamaschi, P., Frankenberg, C., Meirink, J. F., Krol, M., Dentener, F., Wagner, T., Platt, U., Kaplan, J. O., Körner, S., Heimann, M., Dlugokencky, E. J., & Goede, A. (2007). Satellite chartography of atmospheric methane from SCIAMACHY on board ENVISAT: 2. Evaluation based on inverse model simulations. *J. Geophys. Res.*, *112*, D02304. doi:10.1029/2006JD007268.

Berndt, D. J., & Clifford, J. (1994). Using Dynamic Time Warping to Find Patterns in Time Series. In *AAAI-94 Workshop on Knowledge Discovery in Databases* (pp. 359–370). volume 10.

Betbeder, J., Gond, V., Frappart, F., Baghdadi, N. N., Briant, G., & Bartholome, E. (2014). Mapping of Central Africa Forested Wetlands Using Remote Sensing. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, *7*, 531–542. doi:10.1109/JSTARS.2013.2269733.

907 Bloom, A. A., Palmer, P. I., Fraser, A., Reay, D. S., & Frankenberg, C. (2010). Large-Scale Controls of Methanogen-
908 esis Inferred from Methane and Gravity Spaceborne Data. *Science*, *327*, 322–325. doi:10.1126/science.
909 1175176.

910 Bullock, A., & Acreman, M. (2003). The role of wetlands in the hydrological cycle. *Hydrol. Earth Syst. Sci.*, *7*,
911 358–389. doi:10.5194/hess-7-358-2003.

912 Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Comm. in Stats. - Theory & Methods*,
913 *3*, 1–27. doi:10.1080/03610927408827101.

914 Camps, A., Park, H., Pablos, M., Foti, G., Gommenginger, C. P., Liu, P. W., & Judge, J. (2016). Sensitivity of
915 GNSS-R Spaceborne Observations to Soil Moisture and Vegetation. *IEEE J. Sel. Top. Appl. Earth Observations*
916 *Remote Sensing*, *9*, 4730–4742. doi:10.1109/JSTARS.2016.2588467.

917 Cardellach, E., Fabra, F., Nogués-Correig, O., Oliveras, S., Ribó, S., & Rius, A. (2011). GNSS-R ground-based and
918 airborne campaigns for ocean, land, ice, and snow techniques: Application to the GOLD-RTR data sets. *Radio*
919 *Science*, *46*, 1–16. doi:10.1029/2011RS004683.

920 Carreno-Luengo, H., Luzi, G., & Crosetto, M. (2019). Sensitivity of CyGNSS Bistatic Reflectivity and SMAP Mi-
921 crowave Radiometry Brightness Temperature to Geophysical Parameters over Land Surfaces. *IEEE J. Sel. Top.*
922 *Appl. Earth Observations Remote Sensing*, *12*, 107–122. doi:10.1109/JSTARS.2018.2856588. Pub-
923 lisher: IEEE.

924 Carreno-Luengo, H., Luzi, G., & Crosetto, M. (2020). Above-ground biomass retrieval over tropical forests: A
925 novel gnss-r approach with cygnss. *Remote Sensing*, *12*. doi:10.3390/rs12091368.

926 Chapman, B. D., Russo, I. M., Galdi, C., Morris, M., di Bisceglie, M., Zuffada, C., Downs, B., Lavalle, M., Loria, E.,
927 & O'Brien, A. J. (2022). Comparison of sar and cygnss surface water extent metrics. *IEEE Journal of Selected*
928 *Topics in Applied Earth Observations and Remote Sensing*, *15*, 3235–3245. doi:10.1109/JSTARS.2022.
929 3162764.

930 Chen, Y., Huang, C., Ticehurst, C., Merrin, L., & Thew, P. (2013). An Evaluation of MODIS Daily and 8-day
931 Composite Products for Floodplain and Wetland Inundation Mapping. *Wetlands*, *33*, 823–835. doi:10.1007/
932 s13157-013-0439-4.

933 Chew, C., Reager, J. T., & Small, E. (2018). CYGNSS data map flood inundation during the 2017 Atlantic hurricane
934 season. *Sci Rep*, *8*, 9336. doi:10.1038/s41598-018-27673-x.

935 Chew, C., Shah, R., Zuffada, C., Hajj, G., Masters, D., & Mannucci, A. J. (2016). Demonstrating soil moisture
936 remote sensing with observations from the UK TechDemoSat-1 satellite mission. *Geophysical Research Letters*,
937 *43*, 3317–3324. doi:10.1002/2016GL068189.

Chew, C., & Small, E. (2020). Estimating inundation extent using CYGNSS data: A conceptual modeling study. *Remote Sensing of Environment*, *246*, 111869. doi:10.1016/j.rse.2020.111869.

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. doi:10.1029/2018GL077905.

Choudhury, B. J. (1991). Passive microwave remote sensing contribution to hydrological variables. *Surveys in Geophysics*, *12*, 63–84.

Clarizia, M. P., Pierdicca, N., Costantini, F., & Floury, N. (2019). Analysis of CYGNSS Data for Soil Moisture Retrieval. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, *12*, 9.

Clarizia, M. P., & Ruf, C. S. (2016). Wind speed retrieval algorithm for the cyclone global navigation satellite system (cygnss) mission. *IEEE Transactions on Geoscience and Remote Sensing*, *54*, 4419–4432. doi:10.1109/TGRS.2016.2541343.

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

Collett, I., Wang, Y., Shah, R., & Morton, Y. J. (2022). Phase coherence of gps signal land reflections and its dependence on surface characteristics. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1–5. doi:10.1109/LGRS.2021.3094407.

Crétaux, J.-F., Arsen, A., Calmant, S., Kouraev, A., Vuglinski, V., Bergé-Nguyen, M., Gennero, M.-C., Nino, F., Rio, R. A. D., Cazenave, A., & Maisongrande, P. (2011). SOLS: A lake database to monitor in the Near Real Time water level and storage variations from remote sensing data. *Advances in Space Research*, *47*, 1497–1507. doi:https://doi.org/10.1016/j.asr.2011.01.004.

CYGNSS (2020). CYGNSS level 1 science data record version 3.0. ver. 3.0. PO.DAAC, CA, USA. URL: https://doi.org/10.5067/CYGNS-L1X30 accessed: 2021-10-01.

Davidson, N. C., Fluet-Chouinard, E., & Finlayson, C. M. (2018). Global extent and distribution of wetlands: trends and issues. *Mar. Freshwater Res.*, *69*, 620. doi:10.1071/MF17019.

De Roo, R. D., & Ulaby, F. T. (1994). Bistatic specular scattering from rough dielectric surfaces. *IEEE Transactions on Antennas and Propagation*, *42*, 220–231.

Defourny, P., Vancutsem, C., Bicheron, P., Brockmann, C., Nino, F., Schouten, L., & Leroy, M. (2007). GlobCover: A 300M Global Land Cover Product for 2005 Using ENVISAT MERIS Time Series. In *Proceedings of the ISPRS Commission VII Symposium Remote Sensing: From Pixels to Processes* (p. 4). Enschede, The Netherlands.

Di Gregorio, A. (2016). *Land Cover Classification System - Classification concepts Software version 3*. Technical Report Food and Agriculture Organisation Roma, Italy.

Di Vittorio, C. A., & Georgakakos, A. P. (2018). Land cover classification and wetland inundation mapping using MODIS. *Remote Sensing of Environment*, *204*, 1–17. doi:10.1016/j.rse.2017.11.001.

Egido, A., Paloscia, S., Motte, E., Guerriero, L., Pierdicca, N., Caparrini, M., Santi, E., Fontanelli, G., & Floury, N. (2014). Airborne GNSS-R polarimetric measurements for soil moisture and above-ground biomass estimation. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, *7*, 1522–1532. doi:10.1109/JSTARS.2014.2322854. Publisher: IEEE.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., & al. (2010). The Soil Moisture Active Passive (SMAP) Mission. *Proc. IEEE*, *98*, 704–716. doi:10.1109/JPROC.2010.2043918.

Eroglu, O., Kurum, M., Boyd, D., & Gurbuz, A. C. (2019). High spatio-temporal resolution cygnss soil moisture estimates using artificial neural networks. *Remote Sensing*, *11*. doi:10.3390/rs11192272.

ESA (2017). Land cover CCI product user guide version 2. tech. rep. (2017). URL: maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf.

Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global Patterns of Groundwater Table Depth. *Science*, *339*, 940–943. doi:10.1126/science.1229881.

Fatras, C., Frappart, F., Mougin, E., Frison, P.-L., Faye, G., Borderies, P., & Jarlan, L. (2015). Spaceborne altimetry and scatterometry backscattering signatures at C- and Ku-bands over West Africa. *Remote Sensing of Environment*, *159*, 117–133. doi:10.1016/j.rse.2014.12.005.

Fluet-Chouinard, E., Lehner, B., Rebelo, L.-M., Papa, F., & Hamilton, S. K. (2015). Development of a global inundation map at high spatial resolution from topographic downscaling of coarse-scale remote sensing data. *Remote Sensing of Environment*, *158*, 348–361. doi:https://doi.org/10.1016/j.rse.2014.10.015.

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. *Geophys. Res. Lett.*, *42*, 5435–5441. doi:10.1002/2015GL064204.

Frappart, F., Biancamaria, S., Normandin, C., Blarel, F., Bourrel, L., Aumont, M., Azemar, P., Vu, P.-L., Le Toan, T., Lubac, B., & Darrozes, J. (2018). Influence of recent climatic events on the surface water storage of the Tonle Sap Lake. *Science of The Total Environment*, *636*, 1520–1533. doi:10.1016/j.scitotenv.2018.04.326.

Frappart, F., Zeiger, P., Betbeder, J., Gond, V., Bellot, R., Baghdadi, N., Blarel, F., Darrozes, J., Bourrel, L., & Seyler, F. (2021). Automatic Detection of Inland Water Bodies along Altimetry Tracks for Estimating Surface Water Storage Variations in the Congo Basin. *Remote Sensing*, *13*, 22.

Gerlein-Safdi, C., Bloom, A. A., Plant, G., Kort, E. A., & Ruf, C. S. (2021). Improving representation of tropical wetland methane emissions with cygnss inundation maps. *Global Biogeochemical Cycles*, *35*, e2020GB006890. doi:https://doi.org/10.1029/2020GB006890.

Gerlein-Safdi, C., & Ruf, C. S. (2019). A CYGNSS-Based Algorithm for the Detection of Inland Waterbodies. *Geophysical Research Letters*, *46*, 12065–12072. doi:10.1029/2019GL085134.

Ghasemigoudarzi, P., Huang, W., De Silva, O., Yan, Q., & Power, D. T. (2020). Flash Flood Detection From CYGNSS Data Using the RUSBoost Algorithm. *IEEE Access*, *8*, 171864–171881. doi:10.1109/ACCESS.2020.3025302.

Gleason, S., O'Brien, A., Russel, A., Al-Khaldi, M. M., & Johnson, J. T. (2020). Geolocation, calibration and surface resolution of cygnss gnss-r land observations. *Remote Sensing*, *12*. doi:10.3390/rs12081317.

Gleason, S., Ruf, C. S., O'Brien, A. J., & McKague, D. S. (2019). The CYGNSS Level 1 Calibration Algorithm and Error Analysis Based on On-Orbit Measurements. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, *12*, 37–49. doi:10.1109/JSTARS.2018.2832981.

Hamilton, S. K., Sippel, S. J., & Melack, J. M. (2004). Seasonal inundation patterns in two large savanna floodplains of South America: the Llanos de Moxos(Bolivia) and the Llanos del Orinoco(Venezuela and Colombia). *Hydrol. Process.*, *18*, 2103–2116. doi:10.1002/hyp.5559.

Hess, L. L., Melack, J. M., Novo, E. M., Barbosa, C. C., & Gastil, M. (2003). Dual-season mapping of wetland inundation and vegetation for the central Amazon basin. *Remote Sensing of Environment*, *87*, 404–428. doi:10.1016/j.rse.2003.04.001.

Huete, A., Liu, H., Batchily, K., & van Leeuwen, W. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, *59*, 440–451. doi:10.1016/S0034-4257(96)00112-5.

Hydroweb (). Water height time series available online:. URL: http://hydroweb.theia-land.fr (accessed on 07 July 2022).

Jensen, K., McDonald, K., Podest, E., Rodriguez-Alvarez, N., Horna, V., & Steiner, N. (2018). Assessing L-Band GNSS-reflectometry and imaging radar for detecting sub-canopy inundation dynamics in a tropicalwetlands complex. *Remote Sensing*, *10*. doi:10.3390/rs10091431.

Jiménez, C., Catherinot, J., Prigent, C., & Roger, J. (2010). Relations between geological characteristics and satellite-derived infrared and microwave emissivities over deserts in northern Africa and the Arabian Peninsula. *J. Geophys. Res.*, *115*, D20311. doi:10.1029/2010JD013959.

Junk, W. J., Brown, M., Campbell, I. C., Finlayson, M., Gopal, B., Ramberg, L., & Warner, B. G. (2006). The comparative biodiversity of seven globally important wetlands: a synthesis. *Aquat. Sci.*, *68*, 400–414. doi:10.1007/s00027-006-0856-z.

Kellndorfer, J., Pierce, L., Dobson, M., & Ulaby, F. (1998). Toward consistent regional-to-global-scale vegetation characterization using orbital sar systems. *IEEE Transactions on Geoscience and Remote Sensing*, *36*, 1396–1411. doi:10.1109/36.718844.

Kerr, Y., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J., Font, J., & Berger, M. (2001). Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Trans. Geosci. Remote Sensing*, *39*, 1729–1735. doi:10.1109/36.942551.

Kim, H., & Lakshmi, V. (2018). Use of Cyclone Global Navigation Satellite System (CyGNSS) Observations for Estimation of Soil Moisture. *Geophysical Research Letters*, *45*, 11.

Konings, A. G., Piles, M., Das, N., & Entekhabi, D. (2017). L-band vegetation optical depth and effective scattering albedo estimation from SMAP. *Remote Sensing of Environment*, *198*, 460–470. doi:10.1016/j.rse.2017.06.037.

Kouadio, I. K., Aljunid, S., Kamigaki, T., Hammad, K., & Oshitani, H. (2012). Infectious diseases following natural disasters: prevention and control measures. *Expert Review of Anti-infective Therapy*, *10*, 95–104. doi:10.1586/eri.11.155.

Kreibich, H., Van Loon, A. F., Schröter, K., Ward, P. J., Mazzoleni, M., Sairam, N., Abeshu, G. W., Agafonova, S., AghaKouchak, A., Aksoy, H. et al. (2022). The challenge of unprecedented floods and droughts in risk management. *Nature*, (pp. 1–7). doi:10.1038/s41586-022-04917-5.

Kuenzer, C., Guo, H., Huth, J., Leinenkugel, P., Li, X., & Dech, S. (2013). Flood Mapping and Flood Dynamics of the Mekong Delta: ENVISAT-ASAR-WSM Based Time Series Analyses. *Remote Sensing*, *5*, 687–715. doi:10.3390/rs5020687.

Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, *296*, 1–22. doi:10.1016/j.jhydrol.2004.03.028.

Li, W., Cardellach, E., Fabra, F., Rius, A., Ribó, S., & Martín-Neira, M. (2017). First spaceborne phase altimetry over sea ice using TechDemoSat-1 GNSS-R signals. *Geophys. Res. Lett.*, *44*, 8369–8376. doi:10.1002/2017GL074513.

Li, W., CARDELLACH, E., RIBÓ, S., RIUS, A., & ZHOU, B. (2021). First spaceborne demonstration of beidou-3 signals for gnss reflectometry from cygnss constellation. *Chinese Journal of Aeronautics*, *34*, 1–10. doi:https://doi.org/10.1016/j.cja.2020.11.016.

Li, W., Cardellach, E., Ribó, S., Oliveras, S., & Rius, A. (2022a). Exploration of multi-mission spaceborne gnss-r raw if data sets: Processing, data products and potential applications. *Remote Sensing*, *14*. doi:10.3390/rs14061344.

Li, X., Wigneron, J.-P., Fan, L., Frappart, F., Yueh, S. H., Colliander, A., Ebtehaj, A., Gao, L., Fernandez-Moran, R., Liu, X., & al. (2022b). A new SMAP soil moisture and vegetation optical depth product (SMAP-IB): Algorithm, assessment and inter-comparison. *Remote Sensing of Environment*, *271*, 112921. doi:10.1016/j.rse.2022.112921.

Loria, E., O'Brien, A., Zavorotny, V., Downs, B., & Zuffada, C. (2020). Analysis of scattering characteristics from inland bodies of water observed by cygnss. *Remote Sensing of Environment*, *245*, 111825. doi:https://doi.org/10.1016/j.rse.2020.111825.

Macqueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Fifth Berkeley Symposium*, (p. 17).

Maltby, E., & Acreman, M. C. (2011). Ecosystem services of wetlands: pathfinder for a new paradigm. *Hydrological Sciences Journal*, *56*, 1341–1359. doi:10.1080/02626667.2011.631014.

Maltby, E., & Immirzi, P. (1993). Carbon dynamics in peatlands and other wetland soils regional and global perspectives. *Chemosphere*, *27*, 999–1023. doi:10.1016/0045-6535(93)90065-D.

Martin-Neira, M. (1993). *A passive reflectometry and interferometry system (PARIS): application to ocean altimetry* volume 17. ISSN: 03792285 Publication Title: ESA Journal.

Melton, J. R., Wania, R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Beerling, D. J., Chen, G., & al. (2013). Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP). *Biogeosciences*, *10*, 753–788. doi:10.5194/bg-10-753-2013.

Mitra, S., Wassmann, R., & Vlek, P. L. G. (2005). An appraisal of global wetland area and its organic carbon stock. *Current Science*, *88*, 12.

Mitsch, W. J., Bernal, B., Nahlik, A. M., Mander, , Zhang, L., Anderson, C. J., Jørgensen, S. E., & Brix, H. (2013). Wetlands, carbon, and climate change. *Landscape Ecol*, *28*, 583–597. doi:10.1007/s10980-012-9758-8.

Morris, M., Chew, C., Reager, J. T., Shah, R., & Zuffada, C. (2019). A novel approach to monitoring wetland dynamics using CYGNSS: Everglades case study. *Remote Sensing of Environment*, *233*, 111417. doi:10.1016/j.rse.2019.111417.

Müller, M. (2007). Dynamic Time Warping. In M. Müller (Ed.), *Information Retrieval for Music and Motion* (pp. 69–84). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-74048-3_4.

Nisbet, E. G., Dlugokencky, E. J., & Bousquet, P. (2014). Methane on the Rise—Again. *Science*, *343*, 493–495. doi:10.1126/science.1247828.

Normandin, C., Frappart, F., Lubac, B., Bélanger, S., Marieu, V., Blarel, F., Robinet, A., & Guiastrennec-Faugas, L. (2018). Quantification of surface water volume changes in the Mackenzie Delta using satellite multi-mission data. *Hydrol. Earth Syst. Sci.*, *22*, 1543–1561. doi:10.5194/hess-22-1543-2018.

Page, S. E., Rieley, J. O., & Banks, C. J. (2011). Global and regional importance of the tropical peatland carbon pool. *Global Change Biology*, *17*, 798–818. doi:10.1111/j.1365-2486.2010.02279.x.

Parrens, M., Al Bitar, A., Frappart, F., Papa, F., Calmant, S., Crétaux, J.-F., Wigneron, J.-P., & Kerr, Y. (2017). Mapping Dynamic Water Fraction under the Tropical Rain Forests of the Amazonian Basin from SMOS Brightness Temperatures. *Water*, *9*, 350. doi:10.3390/w9050350.

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*, 418–422. doi:10.1038/nature20584.

Prigent, C., Aires, F., Jimenez, C., Papa, F., & Roger, J. (2015). Multiangle Backscattering Observations of Continental Surfaces in Ku-Band (13 GHz) From Satellites: Understanding the Signals, Particularly in Arid Regions. *IEEE Trans. Geosci. Remote Sensing*, *53*, 1364–1373. doi:10.1109/TGRS.2014.2338913.

Prigent, C., Jimenez, C., & Bousquet, P. (2020). Satellite-Derived Global Surface Water Extent and Dynamics Over the Last 25 Years (GIEMS-2). *J. Geophys. Res. Atmos.*, *125*. doi:10.1029/2019JD030711.

Prigent, C., Matthews, E., Aires, F., & Rossow, W. B. (2001). Remote sensing of global wetland dynamics with multiple satellite data sets. *Geophys. Res. Lett.*, *28*, 4631–4634. doi:10.1029/2001GL013263.

Prigent, C., Papa, F., Aires, F., Rossow, W. B., & Matthews, E. (2007). Global inundation dynamics inferred from multiple satellite observations, 1993–2000. *J. Geophys. Res.*, *112*, D12107. doi:10.1029/2006JD007847.

Rajabi, M., Nahavandchi, H., & Hoseini, M. (2020). Evaluation of CYGNSS Observations for Flood Detection and Mapping during Sistan and Baluchestan Torrential Rain in 2020. *Water*, *12*, 2047. doi:10.3390/w12072047.

Richards, J. A., Woodgate, P. W., & Skidmore, A. K. (1987). An explanation of enhanced radar backscattering from flooded forests. *International Journal of Remote Sensing*, *8*, 1093–1100. doi:10.1080/01431168708954756.

Ringeval, B., de Noblet-Ducoudré, N., Ciais, P., Bousquet, P., Prigent, C., Papa, F., & Rossow, W. B. (2010). An attempt to quantify the impact of changes in wetland extent on methane emissions on the seasonal and interannual time scales. *Global Biogeochem. Cycles*, *24*. doi:10.1029/2008GB003354.

Rodriguez-Alvarez, N., Podest, E., Jensen, K., & McDonald, K. C. (2019). Classifying Inundation in a Tropical Wetlands Complex with GNSS-R. *Remote Sensing*, *11*, 1053. doi:10.3390/rs11091053.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. doi:10.1016/0377-0427(87)90125-7.

Ruf, C. S., Atlas, R., Chang, P. S., Clarizia, M. P., Garrison, J. L., Gleason, S., Katzberg, S. J., Jelenak, Z., Johnson, J. T., Majumdar, S. J., O'Brien, A., Posselt, D. J., Ridley, A. J., Rose, R. J., & Zavorotny, V. U. (2016). New ocean winds satellite mission to probe hurricanes and tropical convection. *Bulletin of the American Meteorological Society*, *97*, 385–395. doi:10.1175/BAMS-D-14-00218.1.

Ruffini, G., Soulat, F., Caparrini, M., Germain, O., & Martín-Neira, M. (2004). The Eddy Experiment: Accurate GNSS-R ocean altimetry from low altitude aircraft. *Geophys. Res. Lett.*, *31*. doi:10.1029/2004GL019994.

Russo, I. M., Bisceglie, M. d., Galdi, C., Lavalle, M., & Zuffada, C. (2022). Entropy-based coherence metric for land applications of gnss-r. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–13. doi:10.1109/TGRS.2021.3125858.

Sakamoto, T., Van Nguyen, N., Kotera, A., Ohno, H., Ishitsuka, N., & Yokozawa, M. (2007). Detecting temporal changes in the extent of annual flooding within the Cambodia and the Vietnamese Mekong Delta from MODIS time-series imagery. *Remote Sensing of Environment*, *109*, 295–313. doi:10.1016/j.rse.2007.01.011.

Santoro, M. (2018). GlobBiomass - global datasets of forest biomass. doi:10.1594/PANGAEA.894711.

Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D. M. A., Avitabile, V., Araza, A., de Bruin, S., Herold, M., Quegan, S., Rodríguez-Veiga, P., Balzter, H., Carreiras, J., Schepaschenko, D., Korets, M., Shimada, M., Itoh, T., Moreno Martínez, A., Cavlovic, J., Cazzolla Gatti, R., da Conceição Bispo, P., Dewnath, N., Labrière, N., Liang, J., Lindsell, J., Mitchard, E. T. A., Morel, A., Pacheco Pascagaza, A. M., Ryan, C. M., Slik, F., Vaglio Laurin, G., Verbeeck, H., Wijaya, A., & Willcock, S. (2021). The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. *Earth System Science Data*, *13*, 3927–3950. doi:10.5194/essd-13-3927-2021.

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., Raymond, P. A., Dlugokencky, E. J., Houweling, S., Patra, P. K., & al. (2020). The global methane budget 2000–2017. *Earth System Science Data*, *12*, 1561–1623. doi:10.5194/essd-12-1561-2020.

Senyurek, V., Lei, F., Boyd, D., Kurum, M., Gurbuz, A. C., & Moorhead, R. (2020). Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS. *Remote Sensing*, *12*, 1–24. doi:10.3390/rs12071168.

Sippel, S. J., Hamilton, S. K., Melack, J. M., & Choudhury, B. J. (1994). Determination of inundation area in the Amazon River floodplain using the SMMR 37 GHz polarization difference. *Remote Sensing of Environment*, *48*, 70–76. doi:10.1016/0034-4257(94)90115-5.

Suk, J. E., Vaughan, E. C., Cook, R. G., & Semenza, J. C. (2020). Natural disasters and infectious disease in Europe: a literature review to identify cascading risk pathways. *European Journal of Public Health*, *30*, 928–935. doi:10.1093/eurpub/ckz111.

Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Rußwurm, M., Kolar, K., & Woods, E. (2020). Tslearn, A Machine Learning Toolkit for Time Series Data. *Journal of Machine Learning Research*, *21*, 1–6.

Tootchi, A., Jost, A., & Ducharne, A. (2019). Multi-source global wetland maps combining surface water imagery and groundwater constraints. *Earth Syst. Sci. Data*, *11*, 189–220.

Ulaby, F., Moore, R., & Fung, A. (1982). Microwave remote sensing: Active and passive. volume 2-radar remote sensing and surface scattering and emission theory, .

Unwin, M. J., Pierdicca, N., Cardellach, E., Rautiainen, K., Foti, G., Blunt, P., Guerriero, L., Santi, E., & Tossaint, M. (2021). An introduction to the hydrognss gnss reflectometry remote sensing mission. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 6987–6999. doi:10.1109/JSTARS.2021.3089550.

Wan, W., Liu, B., Zeng, Z., Chen, X., Wu, G., Xu, L., Chen, X., & Hong, Y. (2019). Using CYGNSS Data to Monitor China's Flood Inundation during Typhoon and Extreme Precipitation Events in 2017. *Remote Sensing*, *11*, 854. doi:10.3390/rs11070854.

Webb, E. B., Smith, L. M., Vrtiska, M. P., & Lagrange, T. G. (2010). Effects of Local and Landscape Variables on Wetland Bird Habitat Use During Migration Through the Rainwater Basin. *Journal of Wildlife Management*, *74*, 109–119. doi:10.2193/2008-577.

Whalen, S. (2005). Biogeochemistry of Methane Exchange between Natural Wetlands and the Atmosphere. *Environmental Engineering Science*, *22*, 73–94. doi:10.1089/ees.2005.22.73.

Wigneron, J.-P., Li, X., Frappart, F., Fan, L., Al-Yaari, A., De Lannoy, G., Liu, X., Wang, M., Le Masson, E., & Moisy, C. (2021). SMOS-IC data record of soil moisture and L-VOD: Historical development, applications and perspectives. *Remote Sensing of Environment*, (p. 20).

Xiao, X., Boles, S., Liu, J., Zhuang, D., Frolking, S., Li, C., Salas, W., & Moore, B. (2005). Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote Sensing of Environment*, *95*, 480–492. doi:10.1016/j.rse.2004.12.009.

Yan, Q., Huang, W., Jin, S., & Jia, Y. (2020). Pan-tropical soil moisture mapping based on a three-layer model from CYGNSS GNSS-R data. *Remote Sensing of Environment*, *247*, 111944. doi:10.1016/j.rse.2020.111944.

Zavorotny, V. U., Gleason, S., Cardellach, E., & Camps, A. (2014). Tutorial on remote sensing using GNSS bistatic radar of opportunity. *IEEE Geoscience and Remote Sensing Magazine*, *2*, 8–45. doi:10.1109/MGRS.2014.2374220. Publisher: IEEE.