Mapping Surface Water Fraction Over the Pan-Tropical Region Using CYGNSS Data

Qingyun Yan, *Member, IEEE*, Shuci Liu, Tiexi Chen, Shuanggen Jin, *Senior Member, IEEE*, Tao Xie, and Weimin Huang, *Senior Member, IEEE*

***Abstract*— A new method, which integrates multivariable consisting of soil moisture (SM) active passive (SMAP)-derived SM and vegetation optical depth, the water seasonality, geolocation, digital elevation model (DEM), slope, and biomass as inputs and adopts the technique of bootstrap aggregation of regression trees (BARTs), is proposed for retrieving monthly surface water fraction (SWF) at a spatial resolution of 0.025° from cyclone global navigation satellite system (CYGNSS) data. The model is trained using surface water microwave product series (SWAMPS) data with a coarser resolution of 25 km and then applied to CYGNSS data with an enhanced resolution of 0.025° to generate high-resolution water maps. The resulting CYGNSS SWF (CSWF) maps are evaluated by comparing them with other water data sources, namely, SWAMPS, global surface water (GSW), and global surface water dynamics (GLADs), as well as ground measurements. A quadruple collocation (QC) analysis indicates that the CSWF results exhibit the lowest error variance among the four SWF datasets. Furthermore, additional testing with water level (WL) measurements demonstrates a strong correlation with station data and clear seasonal patterns. Notably, the CSWF estimates significantly improve spatial coverage compared to both optical data (GSW and GLAD) with enhanced spatial resolution and the coarser SWAMPS data. This study underscores the effectiveness and efficiency of CSWF estimates, highlighting their potential as a valuable complement to existing microwave- and optical-based surface water products.**

***Index Terms*— Cyclone global navigation satellite system (CYGNSS), global navigation satellite system-reflectometry (GNSS-R), global surface water (GSW), global surface water dynamic (GLAD), surface water.**

Manuscript received 14 September 2023; revised 23 January 2024 and 29 March 2024; accepted 25 April 2024. Date of publication 29 April 2024; date of current version 8 May 2024. The work of Qingyun Yan was supported by the Anhui Province Key Laboratory of Physical Geographic Environment under Grant 2023PGE012. *(Corresponding author: Tiexi Chen.)*

Qingyun Yan is with the School of Environmental Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China, and also with Anhui Province Key Laboratory of Physical Geographic Environment, Anhui Engineering Research Center of Remote Sensing and Geoinformatics, and Anhui Center for Collaborative Innovation in Geographical Information Integration and Application, Chuzhou University, Chuzhou 239000, China.

Shuci Liu is with the Department of Environment and Science, Queensland Government, Queensland, QLD 4102, Australia.

Tiexi Chen is with the School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: txchen@nuist.edu.cn).

Shuanggen Jin and Tao Xie are with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China.

Weimin Huang is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1B 3X5, Canada.

Digital Object Identifier 10.1109/TGRS.2024.3394744

I. Introduction

SURFACE water is of utter importance due to its tight conjunction with greenhouse gas emissions [1], biodiversity [2], and various forms of lives [3]. However, it is vulnerable to climate changes and human activities [4]. In addition, the presence, extent, and amount of surface water can be highly variable in both space and time. Therefore, it is critical but challenging to monitor surface water and its spatiotemporal evolution in an accurate, effective, and prompt manner.

Remote sensing (RS) provides an efficient means of continuously observing surface water data across regional and global scales. In general, there are two types of RS sensors primarily used to map surface water: optical and microwave. The former is characterized by high spatial resolution. For instance, datasets such as global surface water (GSW) [5] and Global surface water dynamics (GLAD) [6] that are based on Landsat optical imagery have a spatial resolution of 30 m. However, optical RS is hampered by clouds and vegetation, which results in fewer cloud-free satellite images accessible during wet seasons than dry seasons [5]. This leads to a lack of data and limited applicability for the tropical region, especially during the wet seasons. Fig. 1 demonstrates the temporal coverage rate at a monthly scale for the aforementioned two datasets and shows a relatively low revisit frequency over the tropical regions and South Asia. Moreover, the spectral properties of waterbody surfaces are complex at the global scale as they vary with the depths, dissolved material, and amount of chlorophyll [5]. In contrast, microwave RS is less affected by these constraints, offering all-day and all-weather surveillance. Passive radiometers are typically associated with coarse spatial resolutions (25–50 km). Conversely, active microwave platforms, such as synthetic aperture radar (SAR) and radar altimeters, have suitable spatial resolutions, but their revisit periods are usually lengthy. Taking Sentinel-1 as an example, its spatial resolution ranges from several to tens of meters depending on the operating mode, and its revisit time is six days at the equator and increases at higher latitudes [7]. The surface water microwave product series (SWAMPS) dataset [8] is one of the longest (over 20 years) records of global water coverage obtained through combining passive and active microwave observations. The revisit time and spatial resolution of the data are approximately three days and 25 km, respectively. Although it has found various successful applications, its low level of detail makes it unsuitable for fine-scale or regional water mapping.

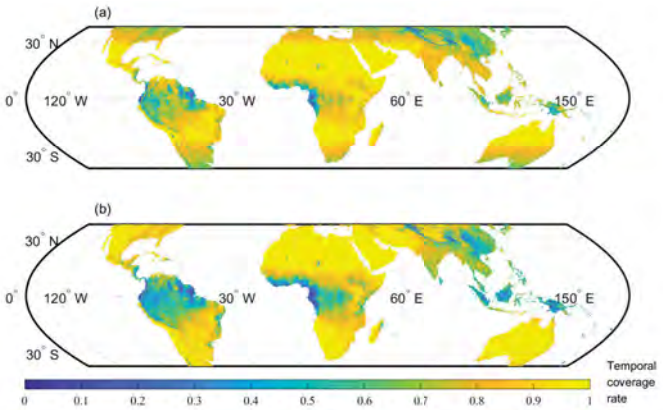


Fig. 1. Temporal coverage rates (i.e., monthly revisit rates) of (a) GSW [5] and (b) GLAD [6] data for the pan-tropical regions over the period from August 2018 to December 2021.

Recently, global navigation satellite system-reflectometry (GNSS-R) has gained widespread recognition as a valuable RS tool. Working in a multistatic mode, spaceborne GNSS-R opportunistically collects forward scattered signals from a surface area surrounding the specular point (SP). The ground track of SPs is quasi-random due to the orbiting motion of both the transmitter and receiver. Unlike traditional RS satellites that sample data in large swaths, GNSS-R provides overall wider coverage although small spatial gaps may exist within certain regions. Taking the first GNSS-R constellation mission—cyclone GNSS (CYGNSS)—as an example, it can provide nearly full coverage over the pan-tropical regions at a spatial resolution of 0.025° and at a monthly time scale (see Fig. 1 as a contrast of incomplete coverage, and more details can be found in Section IV-C). Moreover, GNSS-R has been successful in many global applications, e.g., ocean wind measurement [9], [10], altimetry [11], [12], soil moisture (SM) estimation [13], [14], and ice sensing [15], [16]. Recently, GNSS-R also exhibited its extraordinary capacity in surface water mapping. Notably, GNSS-R adopts L-band signals that provide a better capability in penetrating through canopies than higher frequencies. This is evidenced by the sensitivity of CYGNSS observations to the presence of small Amazonian tributaries even when water is under dense vegetation [17]. Following this, hydrological applications such as flood detection [18], [19], [20], wetland observation [21], [22], [23], and inundation/inland water body mapping [24], [25], [26], [27], [28], [29], [30], [31] have been extensively carried out. More importantly, it is worth undertaking a comprehensive comparison between surface water retrieval products from CYGNSS and those obtained from alternative instruments. This investigation will elucidate the distinct advantages inherent in CYGNSS-based products. Chapman et al. [32] evaluated SAR and CYGNSS-derived entropy and confirmed the significant sensitivity of GNSS-reflected signal to inundation, even for wetlands with vegetation. Downs et al. [33] assessed the sensitivity of CYGNSS normalized signal-to-noise ratio to the inundation extent along with data from other instruments such as visible infrared imaging radiometer suite (VIIRS), SAR, and optical sensors. They argued that CYGNSS presented an oversensitivity to small water bodies, showing that a simple thresholding approach was not robust for accurate inland water mapping.

Though there were successful applications with promising results, most existing studies primarily focused on classification tasks, i.e., labeling pixels into distinct categories (e.g., water or land). These investigations exhibit two noteworthy limitations. First, the classification accuracy can be influenced by biases due to errors in class labeling [6]. Second, these classification-based methods typically lack the capacity to incorporate additional information about a pixel except categorical designation. Therefore, a more comprehensive approach is required to not only quantify the type of surface feature but also incorporate its associated characteristics, particularly those related to surface water properties. This more holistic approach will allow us to comprehensively understand the studied area. In particular, Loria et al. [34] studied the scattering characteristics of CYGNSS signals from inland water bodies and estimated surface water fraction (SWF) for a part of the Amazon River Basin as a case study. Unfortunately, their methods have not yet been examined for a global study, and the accuracy can be further assessed and improved. Furthermore, Chew et al. [35] developed a dielectric model to map SWF by considering soil and water conditions. The authors showed examples of flood maps for four different places (Amazon, Mozambique, Mali, and Australia) and compared them with other data sources (e.g. SWAMPS, Dartmouth Flood Observatory, and GSW). Still, the retrievals have some uncertainties that affect their accuracy mainly because of the parameterizations of soil and water surface roughness, causing an underestimate of SWF, especially over the areas with high SWF.

This study aims to achieve pan-tropical SWF estimation with a spatial resolution of 0.025° and a monthly interval using CYGNSS data. A retrieval model based on bootstrap aggregation of regression trees (BARTs) is developed for this task. This model is first trained according to SWAMPS with a spatial resolution of 25 km and then extended to the inputs rescaled to 0.025° for fine-resolution SWF maps. Results are comprehensively evaluated and compared with the microwave SWAMPS dataset [8] and two optical datasets (see [5], [6]) along with in situ measurements. The remainder of this article is organized as follows. The CYGNSS, reference, and ancillary datasets are described in Section II. The developed SWF retrieving model based on a BART algorithm is presented in Section III. The experimental validation and discussions are shown in Section IV, and the main conclusions are summarized in Section V.

II. Datasets

In this section, CYGNSS RS data, referred surface water data, and other auxiliary inputs are described.

A. CYGNSS RS Data

CYGNSS provides measurement over the region from 38°S to 38°N. The 41-month long (from August 2018 to December 2021) CYGNSS Level 1 ($L1$) Version 3.1 datasets

(accessible at https://cmr.earthdata.nasa.gov/virtual-directory/collections/C2146321631-POCLOUD) are adopted in this work. In practice, the CYGNSS-derived surface reflectivity Γ can be obtained by assuming that coherent reflections dominate over land (e.g., [13], [14], [36], [37], and [38])

$$\Gamma = \frac{\sigma (R_t + R_r)^2}{4\pi (R_t R_r)^2} \tag{1}$$

where σ is the CYGNSS bistatic radar cross section, and R_t and R_r are the distances from SP to the transmitter and receiver, respectively. Such data along with ancillary information about the observation and its geometry are provided in CYGNSS $L1$ data. Moreover, a delay-Doppler map observable (pixel number observable, hereafter, PN) that describes surface roughness [15] is employed as well. The data quality control scheme follows that in [14]. In addition, data with the quality flag "SP in the sidelobe" are excluded due to their low confidence in the antenna gain. CYGNSS data are aggregated into monthly scale with a spatial resolution of $0.025° \times 0.025°$ grids over regions between 37.5°S and 37.5°N according to the geolocation of SPs.

B. Reference Data

1) SWAMPS Data: SWAMPS data [8] are adopted as training references. SWAMPS is a gridded dataset of inundation fraction that is derived from data collected by several different microwave sensors. It stands out as one of the longest publicly available microwave RS datasets, with daily data files spanning the time period 1992–2020. The temporal repeat of the data is approximately three days, and the data are gridded with a resolution of 25 km. This dataset is employed as the training target and serves as a reference for coarse-resolution evaluation. In addition, SWAMPS data are aggregated from daily to monthly averages.

2) Optical-Based Data: For fine-resolution assessment, two different surface water datasets (GSW [5] and GLAD [6]) produced from Landsat optical imagery are employed in this work. These datasets are among the most comprehensive sources of surface water data, spanning long time periods (1984 to 2021 for GSW and 1999 to 2021 for GLAD), with spatial and temporal resolutions of 30 m and one month, respectively. However, they may face challenges such as insufficient coverage due to cloud contamination or heavy canopies. Data from August 2018 to December 2021 (41 months in total) are selected for analysis. In addition, the number of valid observations (excluding NaN in the original data) is counted for each spatial grid, and the ratio of this number to 41 represents the temporal coverage rate, as shown in Fig. 1. Slight differences in temporal coverage between the two datasets are noticeable in Fig. 1. Furthermore, it is reported that the GLAD dataset identifies more water than the GSW dataset [6]. Due to the absence of a consensus regarding which dataset performs better, both GSW and GLAD are utilized as references in this study.

3) Hydro Station Data: For local-scale validation, two types of water level (WL) measurement data are utilized. First, daily WL observations from 1491 gauging stations across Australia were obtained from the Bureau of Meteorology Water data online (http://www.bom.gov.au/waterdata/). Although the data have a high sampling frequency of one day, the measurements are restricted at specific points.

Second, WL data from 2173 virtual stations (derived from altimeter) provided by Copernicus global land service for the Amazon River regions over [(15°S–5°N), (80°W–30°W)] are also used, and they can be downloaded at https://land.copernicus.eu/global/products/wl. The satellites can only measure the WL of water bodies that are directly below them (nadir position). The quality of the measurement also depends on the size of the water body and its surrounding landscape, and the revisit time of virtual stations is determined by the orbital period of the satellite that observes them. In the Amazon basin, the median correlation between the time series of the virtual and in situ stations is 0.95, and the mean temporal sampling rate is 27.9 days. Despite the good correlation with in situ data, low revisit frequency is a limitation of virtual station measurements.

C. Ancillary Data

To account for the impacts of geolocation, topography, canopy attenuation, and surface SM, latitude/longitude, digital elevation model (DEM)/slope, biomass/vegetation water content (VWC), and SM data are, respectively, adopted as auxiliary data, and they are described as follows.

1) DEM: Global land one-kilometer base elevation (GLOBE) DEM data (available at https://www.ngdc.noaa.gov/mgg/fliers/globedem.html) are employed. In addition, the surface slope calculated from such DEM data is also utilized. Both datasets have a spatial resolution of 1 km.

2) Water Seasonality: The GSW seasonality map, downloadable from https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/GSWE/Aggregated/LATEST/seasonality, has a spatial resolution of 30 m. It provides the number of months water was present within a year and is derived from the GSW data [5]. This map may underestimate the extent of semipermanent water beneath vegetation canopies, but it offers valuable a priori knowledge of where surface water usually exists.

3) Biomass: European Space Agency's (ESA's) climate change initiative (CCI) biomass (available at https://data.ceda.ac.uk/neodc/esacci/biomass/data/agb/maps/v3.0/) dataset for the year of 2018 is utilized. These abovementioned ancillary data are static data that cannot describe actual temporal variation. The biomass maps have a global coverage but may be biased at regional scales. In addition, the input data source, primarily radar backscatter, may not accurately reflect the complex structure of forest biomass, leading to potential error in areas with dense understory or significant deadwood.

4) VWC and SM: SM active passive (SMAP) enhanced $L3$ radiometer global and polar grid daily 9-km EASE-Grid SM, Version 5 [39], is also employed. The time span between August 2018 and December 2021 is considered. Google Earth Engine (GEE), a cloud computing platform [40], is used to query and process SMAP data, i.e., daily AM and PM SM and VWC from August 1, 2018, to December 31, 2021. The daily AM and PM collections for both variables are aggregated to a daily scale and then further aggregated to a monthly scale,

TABLE I
RESOLUTIONS OF ANCILLARY INPUT DATA SOURCES AND REFERENCES

Data Source	Native Resolution	Resolution During Training	Resolution During Retrieval
SWAMPS	25 km	25 km	0.025°
DEM	1 km		
Water Seasonality	30 m		
Biomass	100 m		
VWC	9 km		
SM	9 km		
GSW	30 m		
GLAD	30 m		

with a spatial resolution of 9 km. The GEE Python application programming interface (API) is employed to accomplish data collection and processing [41].

D. Data Collocation

1) Training: During the training stage, all the involved parameters are projected into the geo-framework provided by the SWAMPS dataset that is of a spatial resolution of 25 km. Data are aggregated and averaged on a monthly basis.

2) Fine-Resolution Retrieval: This work aims at providing SWF retrieval at a spatial resolution of 0.025° and a monthly time step; therefore, the aforementioned data are matched on such basis accordingly. The fine-resolution DEM, slope, and biomass data, as well as GSW reference data, are spatially averaged into grids of 0.025°. The coarse-resolution SMAP SM and VWC data are linearly interpolated and reprojected into the adopted spatial grids. The monthly mean water percent from GLAD data [6] is directly used. The water mask data from GSW data [5] are converted to an equivalent parameter based on pixel counting. These two data are utilized for comparison. The native resolutions of all the inputs and reference optical SWF products are summarized in Table I.

III. SWF ESTIMATION METHOD

Regression trees (RTs) [42] can recursively partition the input space, which, in this context, involves the CYGNSS Γ and PN, SMAP SM and VWC, the water seasonality, geolocation, DEM, slope, and biomass data and, subsequently, maps each partition to the desired output, i.e., SWF. However, a single RT usually overfits the data. As a solution, Bagging [43] is utilized to improve both the stability and accuracy of RTs by reducing their variance and avoiding overfitting. It can also handle high-dimensional data and complex interactions among features as the input here contains ten elements.

As such, the bootstrap aggregation (bagging) of regression trees are adopted here to quantify the relationship between SWF and devised input (CYGNSS data along with considered influencing factors). It is worth noting that the BART model has been successfully applied to several RS applications, e.g., downscaling SMAP SM products [44] and SM retrieval [45]. To achieve SWF estimation, given the training set $T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$ with x_i being the ith input containing ten elements, y_i being the ith desired SWF value, and N being the number of total training samples, this work implements the BARTs as follows. The overall goal is to build a model $f(x)$ such that the loss function is minimum as follows:

$$\text{object} = \min(\text{Loss}(f(x_i), y_i)). \tag{2}$$

Generally, the loss function is defined as the mean square error (MSE) between the estimated and targeted outputs

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - y_i)^2. \tag{3}$$

Assuming that the designed RTs are allocated with M leaves, consequently, there will be at most M varying estimates (or equivalent M units $R_1, \ldots, R_m$), and (3) can be updated with

$$\min \frac{1}{n}\sum_{m=1}^{M}\sum_{x_i \in R_m}(c_m - y_i)^2 \tag{4}$$

where c_m is the prediction from the mth leaf. As such, the splitting parameter and point for each leaf that minimizes (4) should be recorded. This is achieved through traversing a parameter p, scanning the splitting point q, and determining the (p, q) pair that meets

$$\min\left[\min_{c_1}\sum_{x_i \in R_1\{p,q\}}(y_i - c_1)^2 + \min_{c_2}\sum_{x_i \in R_2\{p,q\}}(y_i - c_2)^2\right]. \tag{5}$$

Based on the determined (p, q) pair, one can further calculate the outputs

$$\hat{c_m} = \frac{1}{N_m}\sum_{x_i \in R_m(j,s)} y_i, \quad x \in R_m, \ m = 1, 2 \tag{6}$$

with

$$\begin{aligned} R_1\{j, s\} &= \{x \mid x^{(j)} \le s\} \\ R_2\{j, s\} &= \{x \mid x^{(j)} > s\} \end{aligned} \tag{7}$$

and N_m is the number of elements for R_m. When the total M leaves are determined, the overall RT is, thus, in the following form:

$$f(x) = \begin{cases} \hat{c_1}, & x_i \in R_1(j, s) \\ \hat{c_2}, & x_i \in R_2(j, s) \\ \hat{c_3}, & x_i \in R_3(j, s) \\ \quad \cdots & \\ \hat{c_M}, & x_i \in R_M(j, s). \end{cases} \tag{8}$$

For details of the implementation of Bagging, one can refer to [43]; the formula can be given as

$$\hat{h}(x) = \frac{1}{T}\sum_{t=1}^{T}\hat{f}_t(x) \tag{9}$$

where T is the number of new sample sets and $\hat{f}_t$ is the predication from the tth tree.

In summary, the input of the BART model consists of CYGNSS Γ and PN, SMAP SM and VWC, the water seasonality, geolocation, DEM, slope, and biomass, while the output

TABLE II
ACCURACY OF SWF ESTIMATES AT 25 km-RESOLUTION
(* INDICATES THAT p-VALUE IS <0.05)

	Measure	R^2	RMSE
CSWF v.s. SWAMPS	Training	0.976*	0.0147
	Test	0.964*	0.0179
	Overall	0.974*	0.0153
GSW v.s. SWAMPS		0.532*	0.0618
GLAD v.s. SWAMPS		0.546*	0.0552
CIF v.s. SWAMPS		0.583*	0.0612
CSWF v.s. GLAD		0.523*	0.0485
GSW v.s. GLAD		0.964*	0.0099
CIF v.s. GLAD		0.842*	0.0196

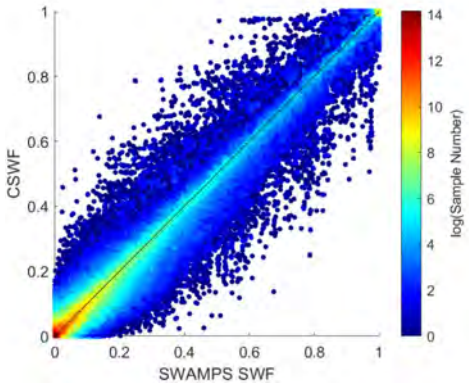


Fig. 2. Scatter plot of SWF results: CSWF versus SWAMPS.

TABLE III
DECREASE OF PERFORMANCE BY REMOVING VARIABLES

Remove	CYGNSS	SMAP	Seasonality	Biomass	Lat/lon	Dem/slope
ΔRMSE	0.0379	0.0317	0.0490	0.0114	0.0267	0.0337
ΔR^2	0.196	0.182	0.327	0.042	0.153	0.201

is SWF. It should be clarified that variables of different types (e.g., SWAMPS inundation fraction and CYGNSS Γ) can be used interchangeably as inputs to the BART model during the training and application stages. The model construction and analyses here are performed based on the regression learner toolbox of MATLAB R2021b software.

IV. EXPERIMENTS AND EVALUATION

Here, the proposed scheme for retrieving CYGNSS SWF (CSWF) is implemented and evaluated using multisource data, including SWAMPS [8], rescaled GSW [5], and GLAD [6] global products, as well as field measurements.

A. Model Training

Although the SWAMPS data have a coarse resolution of 25 km, it is able to penetrate clouds providing observation absent in the GSW/GLAD datasets. As such, the SWAMPS dataset was adopted as the training target here. However, it should be noted that SWAMPS may underestimate SWF for areas with dense canopy and have limited seasonal changes [8], [23], [46], potentially affecting the retrieval accuracy in this study. By the date when this study began, the SWAMPS dataset up to December 2020 was available, and its overlap with CYGNSS data was 29 months. To ensure the model's robustness in capturing different seasonal dynamics, the first 24-month data were selected as a training set and the rest as a testing set for the BART model (see details in Section III). To reduce the overall computational complexity, the pan-tropical regions under investigation (a latitudinal span of $\pm 37.5°$) were divided subject to the subbasin breakdowns level 2 (see https://www.hydrosheds.org/products/hydrobasins), and the training process was operated for each subarea individually. The training and test performance in terms of root MSE (RMSE) and coefficient of determination (R^2) between all of the CSWF and SWAMPS results is summarized in Table II. For both sets, R^2 exceeds 0.96, and RMSE is lower than 0.018. Slight variations between the training and test sets indicate the proposed model's generalizability. The heat map showing the overall good consistency is displayed in Fig. 2.

B. Coarse-Resolution Test

First, the retrieval results in the SWAMPS geo-frameworks (25-km resolution) are compared. The high-resolution GSW and GLAD products are upscaled accordingly. An example of these four datasets for December 2020 is illustrated in Fig. 3. Overall, there are consistent spatial patterns across all four products, and the accuracy statistics are summarized in Table II. To assess the errors of these four SWF products, the quadruple collocation (QC) analysis [47] was adopted. Their spatial distribution (see Fig. 4) indicates that the CYGNSS-derived result has the lowest error variance globally.

In addition, to evaluate the importance of the time-evolving parameters associated with CYGNSS and SMAP satellites, feature selection was also performed for input sets without CYGNSS Γ and PN and for that excluding SMAP SM and VWC. Both schemes experienced obvious performance degradation (original RMSE $= 0.0153$ and $R^2 = 0.974$), and the former (without CYGNSS data, generating an RMSE $=$ 0.0532 and $R^2 = 0.778$) dropped more than the latter (RMSE $= 0.0470$ and $R^2 = 0.792$), indicating the higher impacts of CYGNSS data. For the contribution of static data, we also implemented a variable importance analysis by excluding the DEM/slope, latitude/longitude, biomass, and water seasonality data from the inputs one by one and measuring the decrease in the accuracy of retrieved results for each model. The overall contribution of each input product can be evaluated from the associated performance degradation shown in Table III, from which it can be seen all variables made improvements to the performance, and the water seasonality, CYGNSS, and topographic data had the highest influences.

C. Fine-Resolution Examination

Through applying the devised BARTs to the 0.025°-gridded data, SWF maps with increased detail were generated. Here, the equivalent products computed from the GSW and GLAD datasets were adopted for comparison. The degree of agreement among them regarding R^2 and RMSE is tabulated in Table IV. The CYGNSS-based results demonstrated

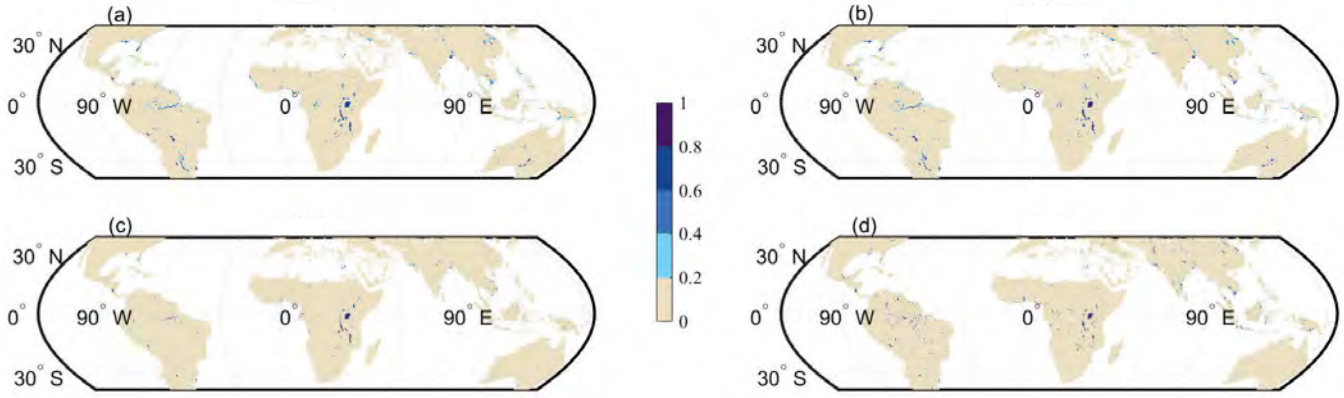


Fig. 3. 25 km-resolution SWF for December 2020. (a) CSWF. (b) SWAMPS. (c) GSW. (d) GLAD.

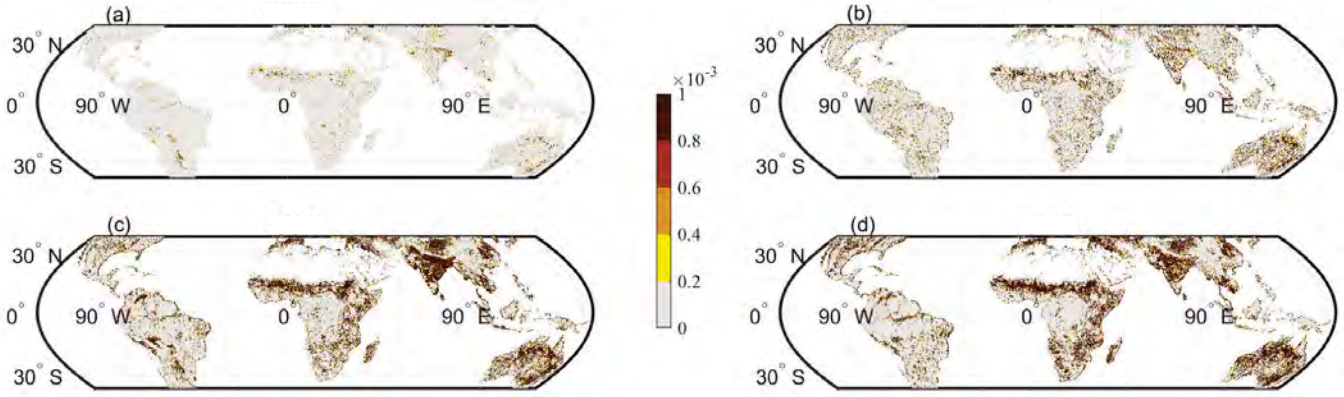
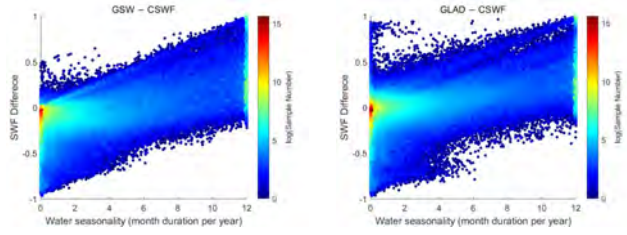


Fig. 4. Error variance of SWF based on (a) CSWF, (b) SWAMPS, (c) GSW, and (d) GLAD data from August 2018 to December 2020.

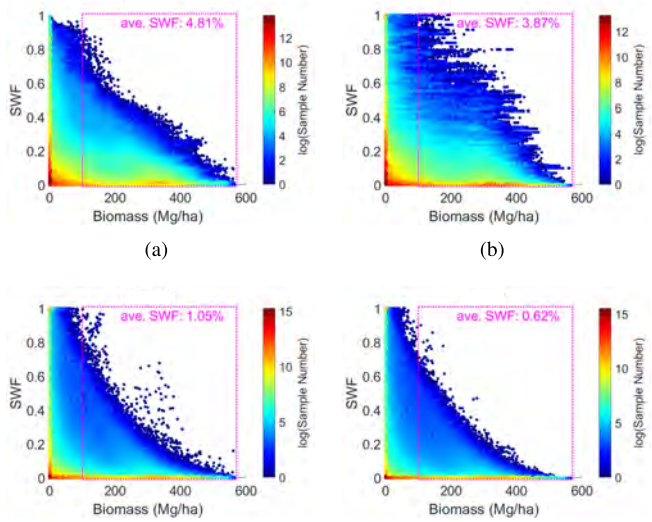
TABLE IV
CONSISTENCY BETWEEN 0.025°-GRIDDED SWF MAPS
(* INDICATES THAT p-VALUE IS <0.05)

Measure	R^2	RMSE
CYGNSS (bef. int.) v.s. GLAD	0.494*	0.0607
CYGNSS (aft. int.) v.s. GLAD	0.495*	0.0608
GSW v.s. GLAD	0.829*	0.0193
CIF v.s. GLAD	0.695*	0.0442



(a) (b)

Fig. 5. Heat maps of SWF differences in terms of water seasonality. (a) GSW. (b) GLAD versus CSWF.



(a) (b) (c) (d)

Fig. 6. Heat maps of SWF with respect to biomass. (a) CSWF. (b) SWAMPS. (c) GLAD. (d) GSW. The mean SWFs for regions with high biomass (over 100 Mg/Ha) are 4.81%, 3.87%, 1.05%, and 0.62%, respectively.

reasonable consistency with the optical-based ones, despite the differences in data sources. Furthermore, the achieved RMSE and R^2 between the CYGNSS and optical SWFs show similar values at two different spatial scales, indicating the stability of CYGNSS results downscaled from 25 km to 0.025°.

Differences between the optical- and CYGNSS-based SWFs are calculated with respect to the water seasonality (introduced in Section II-C2) are presented in Fig. 5. Compared with both optical results, CYGNSS tends to underestimate SWF

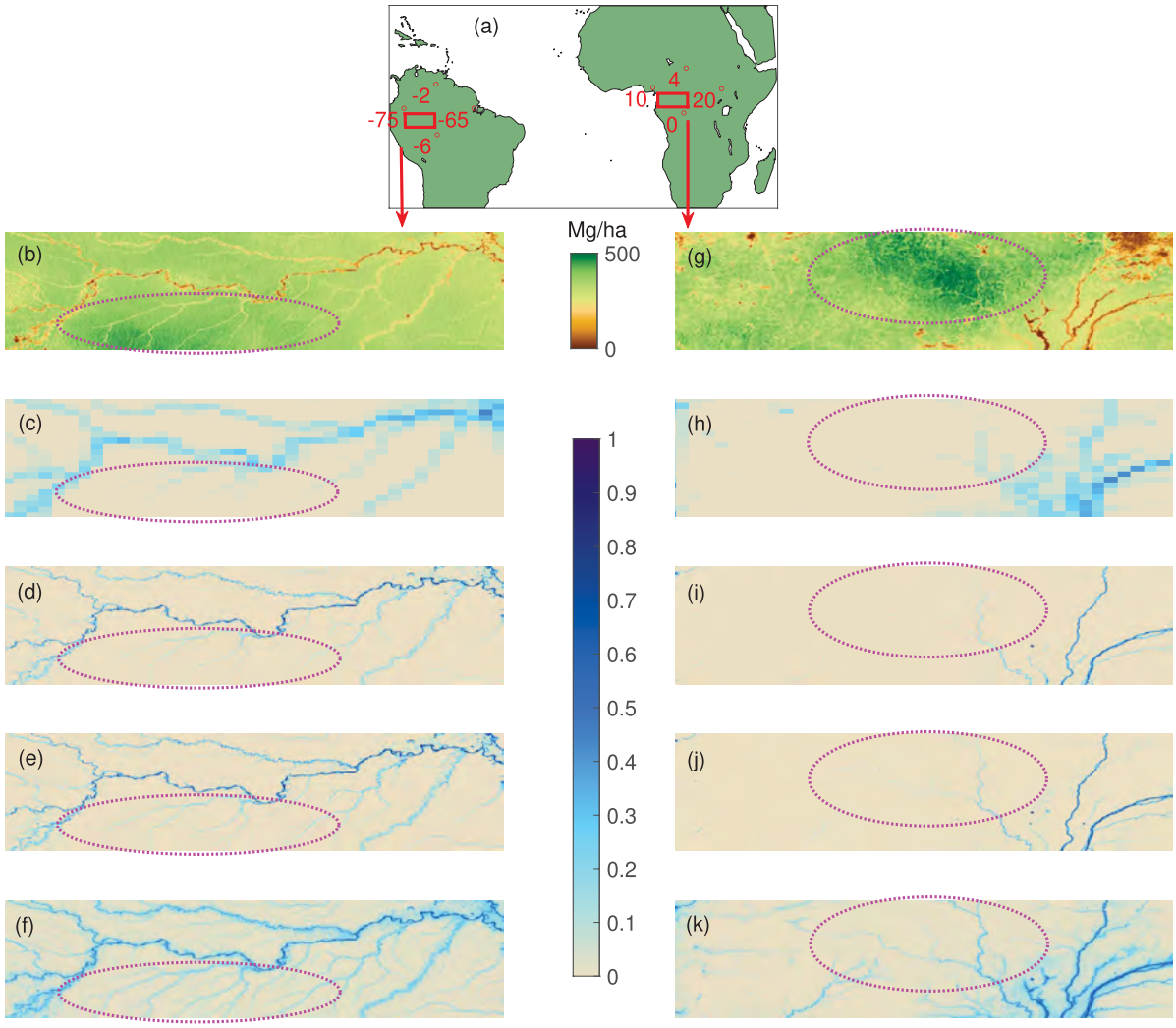


Fig. 7. Examples of maximum SWF and biomass over two regions: (a) geolocation, (b) biomass, (c) SWAMPS, (d) GSW, (e) GLAD, and (f) CSWF in Amazon, and (g) biomass, (h) SWAMPS, (i) GSW, (j) GLAD, and (k) CSWF in Congo.

for permanent water whose seasonality is 12 months. This can be due to imperfectly resolved surface roughness issues over these regions as in [35]. It is worth mentioning that the zero seasonality did not impede CYGNSS from producing nonzero SWF values implying the capability of CYGNSS in detecting semipermanent water bodies that may not be discovered by optical sensors under high biomass. To further assess this, the maximum SWFs are calculated, and the density plots with respect to biomass are presented in Fig. 6, from which higher SWF estimates are observed from CSWF. The averaged SWF values over regions with biomass above 100 Mg/ha are 4.81%, 3.87%, 1.05%, and 0.62% for CSWF, SWAMPS, GLAD, and GSW, respectively. Two examples of SWF estimates over regions with high biomass (see Fig. 7) also demonstrate the capability of CSWF in mapping surface water under dense canopy.

It is worth reiterating that the GSW and GLAD data are based on Landsat optical imagery and can have numerous data gaps due to cloud contamination, as indicated by the overall temporal coverage shown in Fig. 1. Furthermore, it is illustrated that data gaps are particularly prominent during flood seasons. Taking the GLAD data as an example, the missing rates over time in the four selected regions, i.e., Ease Asia, South Asia, Nigeria, and Amazon, are plotted in Fig. 8.

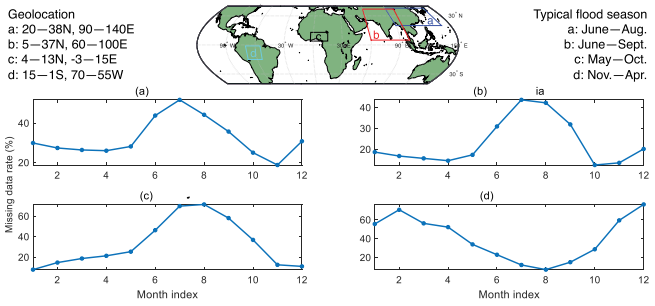


Fig. 8. Missing data rates (i.e., the average spatial fraction of areas that do not contain data during the period from August 2018 to December 2021) calculated from the GLAD [6] dataset for the pan-tropical regions over four regions. (a) East Asia. (b) South Asia. (c) Nigeria. (d) Amazon.

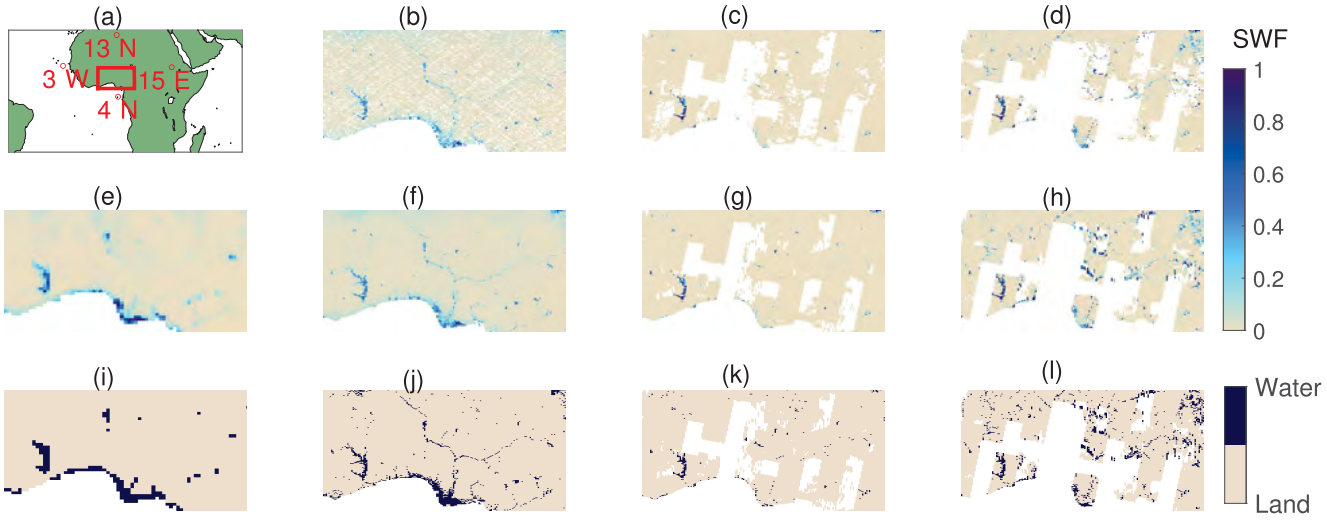


Fig. 9. SWF maps. (a) Study area, (b) CSWF (0.025°) before interpolation (BI), (c) GSW (0.025°) BI, (d) GLAD (0.025°) BI, (e) SWAMPS (25 km), (f) CSWF (0.025°) after interpolation (AI), (g) GSW (0.025°) AI, and (h) GLAD (0.025°) AI and corresponding water masks: (i) SWAMPS, (j) CSWF, (k) GSW, and (l) GLAD for August 2018.

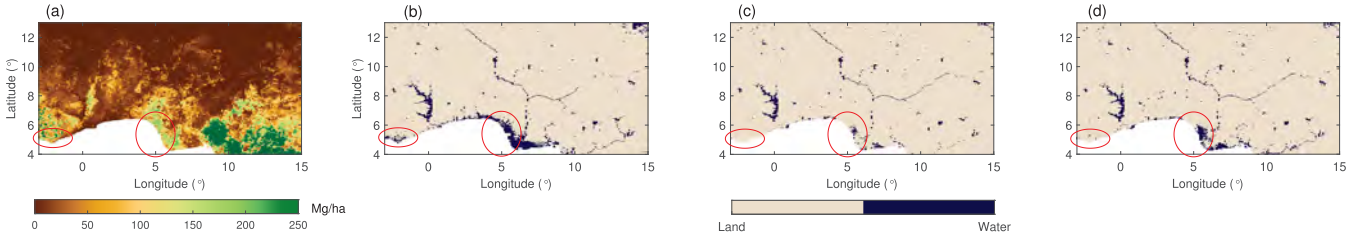


Fig. 10. (a) Biomass and water masks from (b) CSWF, (c) GSW, and (d) GLAD for January 2019.

Without loss of generality, the Nigeria region (Area C in Fig. 8) is treated as a case study here. Fig. 9(e)–(h) displays the four SWF maps of August 2018 (flood season). It is evident that both the GSW and GLAD results are affected by data gaps, whereas the microwave CYGNSS and SWAMPS data remain immune to such issues. In addition, in comparison with the 25-km resolution SWAMPS data, the 0.025°-gridded CYGNSS retrieval results offer more detailed information with clearer textures. Fig. 9(i)–(l) presents the water masks through thresholding the SWF results with a threshold of 0.1. Fig. 10(b)–(d) depicts the derived water masks (water: SWF $\geq$ 0.1) for January 2019 (nonflood season), during which the optical-based products obtained good spatial coverage but captured less water extents than those based on microwave sensors (see red circles; the water mask from SWAMPS is similar to Fig. 9(b) and, thus, is neglected here). This may be due to the high biomass that leads to an underestimation in optical-based SWF.

It should also be noted that the initial spatial coverage rate of all four products was around 90%. However, employing a simple and effective 10 $\times$ 10 moving window for gap filling resulted in a coverage rate of approximately 100% for CSWF. In contrast, the improvement for the GSW and GLAD datasets was relatively limited. This discrepancy can be attributed to the large gaps between swaths and the presence of clouds in Landsat imagery, while CYGNSS data randomly sample the Earth's surface with a wide extent and small gaps. An example depicting the difference between the results obtained using data with/without gap filling can be seen in Fig. 9(b)–(d) and (f)–(h), and the comparison of overall coverage is tabulated in Table V. Furthermore, the filled CYGNSS-based data showed a similar performance to the original ones (see Table IV).

TABLE V
COVERAGE RATE FOR OF 0.025° SWF PRODUCTS

Data	Original Coverage	Coverage after gap filling
GSW	90.30%	91.97%
GLAD	93.54%	95.87%
CYGNSS (before 201907)	73.27%	99.41%
CYGNSS (after 201907)	91.35%	99.94%

D. Case Studies in Amazon and Australia: Seasonality and Validation With WL Measurement

For illustration, seasonal patterns of retrieved CSWF in the Amazon Basin and Australia are displayed in Figs. 11 and 12 for January, April, July, and October 2019. The results, when subtracted by their mean across multiple years, are also presented. Notably, distinct seasonal cycles are evident, and different regions exhibit varying patterns. For example, the Orinoco River shows an opposite trend against the Madeira River, while the Amazon River displays a relative lag (see Fig. 11), which can be caused by the oscillation of Inter

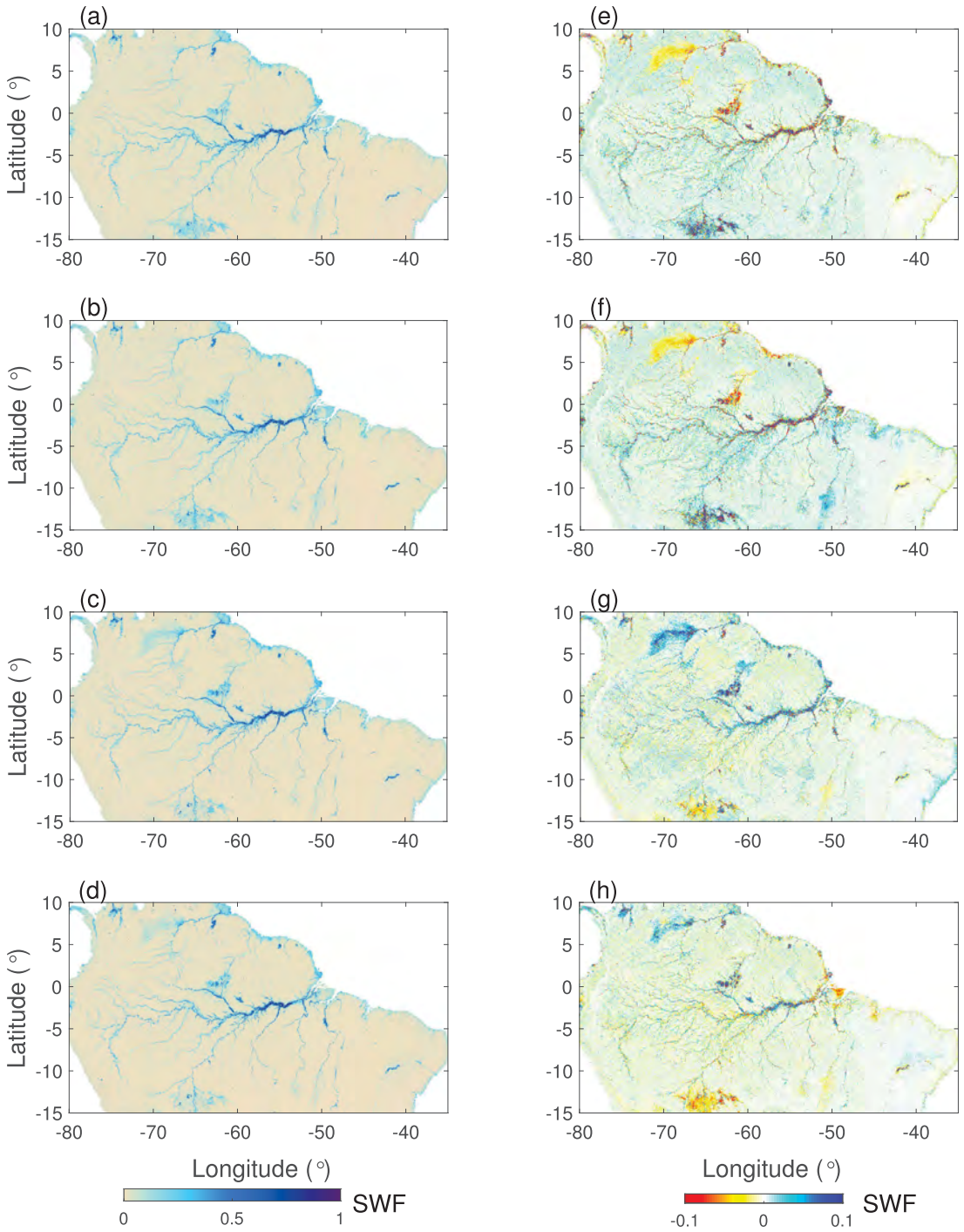


Fig. 11. Retrieved monthly CSWF maps in Amazon Basin for (a) 2019.01, (b) 2019.04, (c) 2019.07, and (d) 2019.10 along with their corresponding values being subtracted by the yearly mean: (e)–(h).

Tropical Convergence Zone. The seasonal difference in SWF across the Australian continent (see Fig. 12) clearly indicates that there are considerable spatial and temporal variabilities in SWFs in regions such as the Great Artesian Basin (GAB). This can be explained by the nonperennial nature of the water bodies across the GAB, as a result of diverse climates and variable connections between surface and groundwater storages. For instance, the tropical climate in the northern part of the GAB (e.g., the Cape York Peninsula) leads to most rainfall occurring in summer, leading to an above average SWF [see Fig. 12(b)]. For the arid southern part of the GAB (e.g., Lake Eyre and the Cooper River), the groundwater recharge creates a water source for surface water even during dry periods. Nonetheless, the seasonal variation presented by the proposed CSWF can be underestimated due to the limitation of SWAMPS data in representing seasonal changes [23].

To verify the seasonality of SWF data and its application at the regional scale, the WL records collected by stations were taken as another reference. Although WL is not a direct measurement of SWF, they are highly correlated [48]. Two

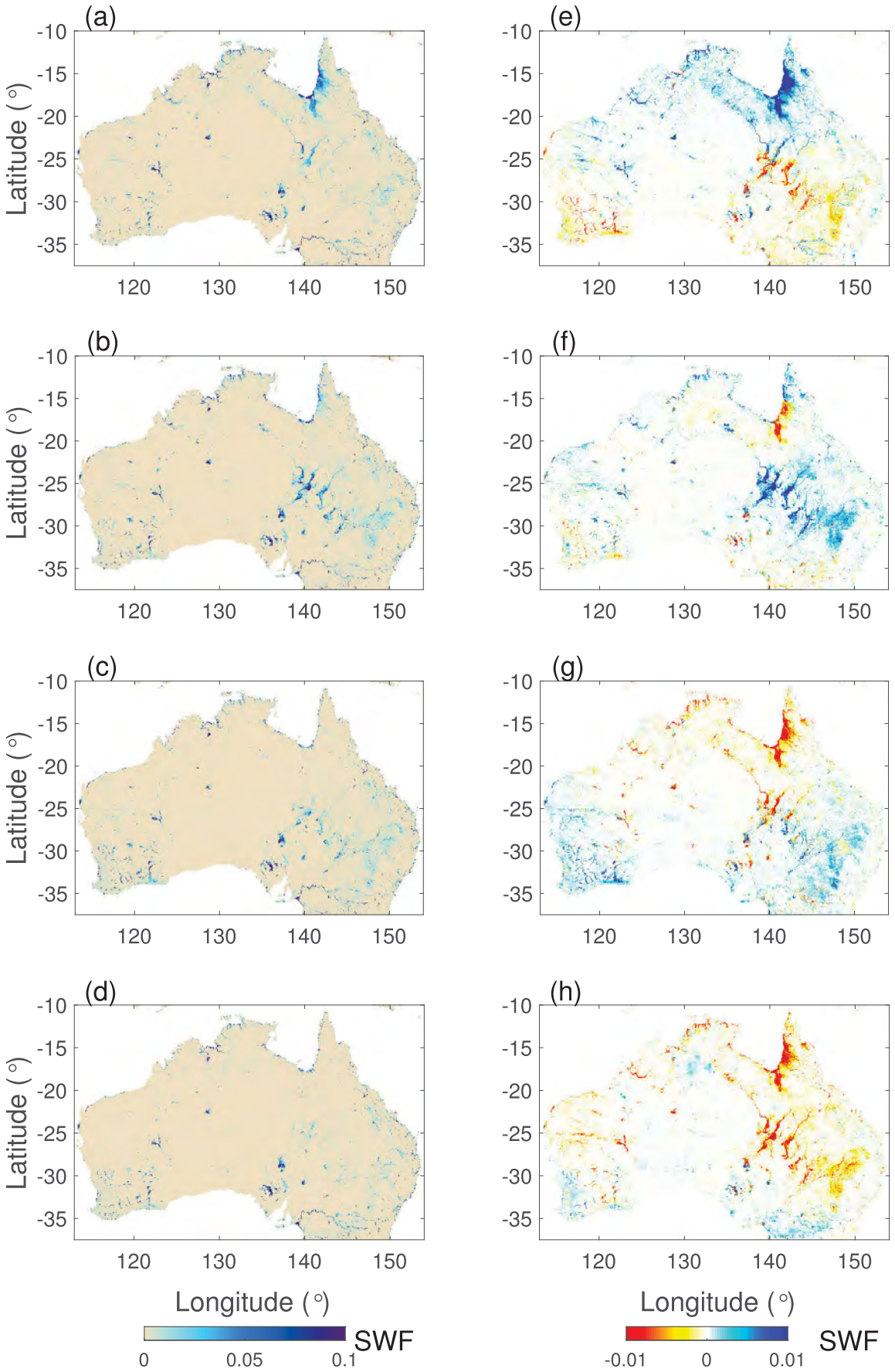


Fig. 12. Monthly CSWF retrievals in Australia for (a) 2019.01, (b) 2019.04, (c) 2019.07, and (d) 2019.10 along with their values being subtracted by the yearly mean: (e)–(h).

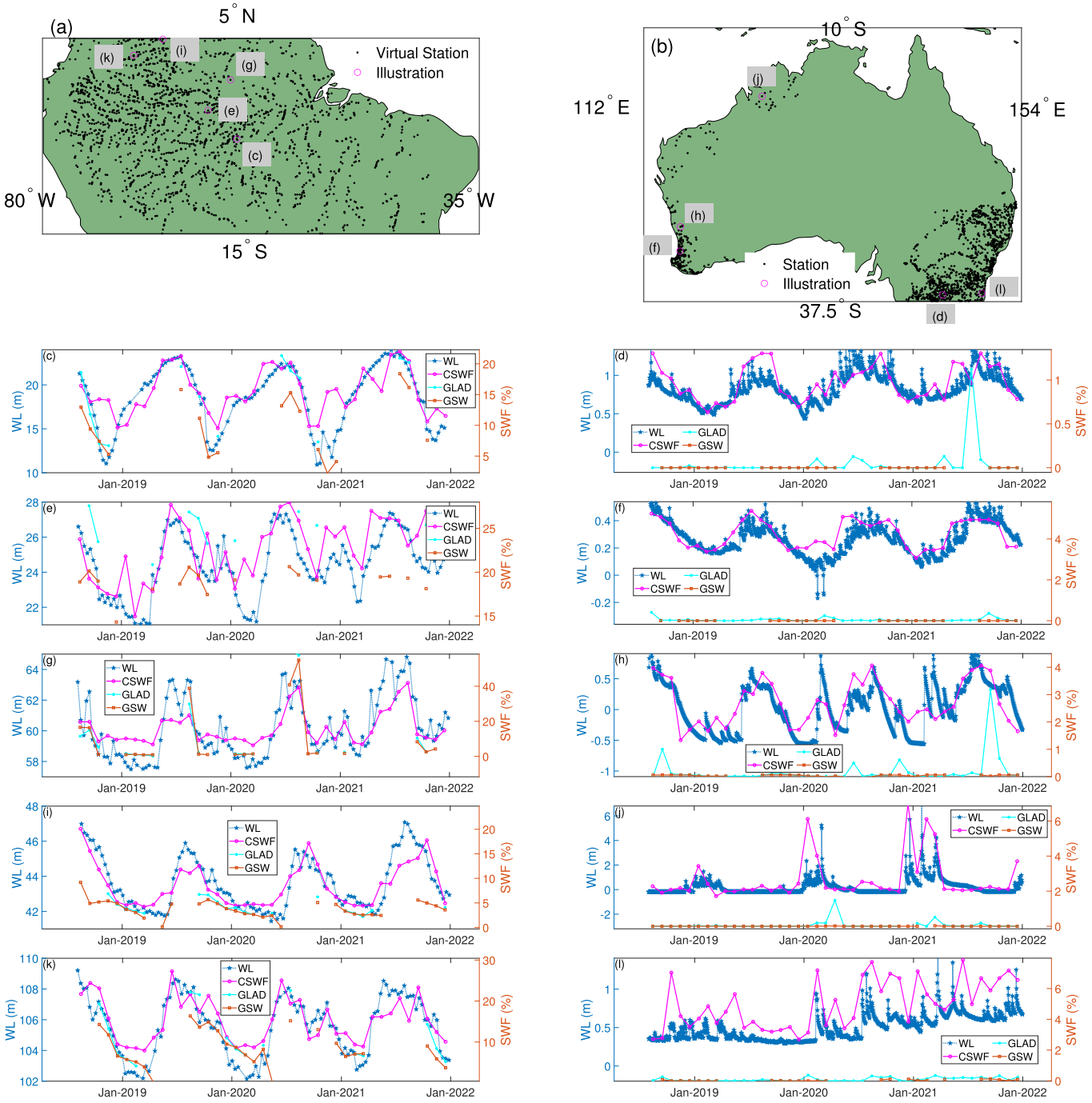


Fig. 13. Distribution of WL (virtual) stations in (a) Amazon River and (b) Australia and examples of WL/SWF data in both regions. Locations of (c)–(l) are circled in (a) and (b).

field data sources from the Amazon River and Australia were employed, the former is provided by a radar altimeter (which measures water surface height above the reference datum), and the latter is derived from continuous monitoring stations. The distribution of these stations (marked by dots) and examples (labeled by circles) of WL and SWF measurements for 41 months in both regions are demonstrated in Fig. 13. Each mark in the time-series plots represents a sample of WL. The sampling rates of WL stations in Australia and Amazon are 1 day and tens of days, respectively. Consequently, the WL records in Australia appear denser than those in Amazon [see Fig. 13(c)–(l)]. Amazon experiences seasonal variations in WL, specifically with a swelling (i.e., increase in WL) in river extent during the wet season, and a shrinkage (i.e., decrease in WL) in the dry season. Australia also undergoes seasonal cycles in WL but with water scarcity during the dry season, particularly in the inland arid regions. Regarding the monthly SWF time series, SWAMPS results are not presented here due to its coarse spatial resolution (25 km) and a short concurrent period (29 months) with CSWF. CSWF provides continuous observations, while GLAD and GSW exhibit data gaps in both regions. For examples in Amazon, CSWF exhibits

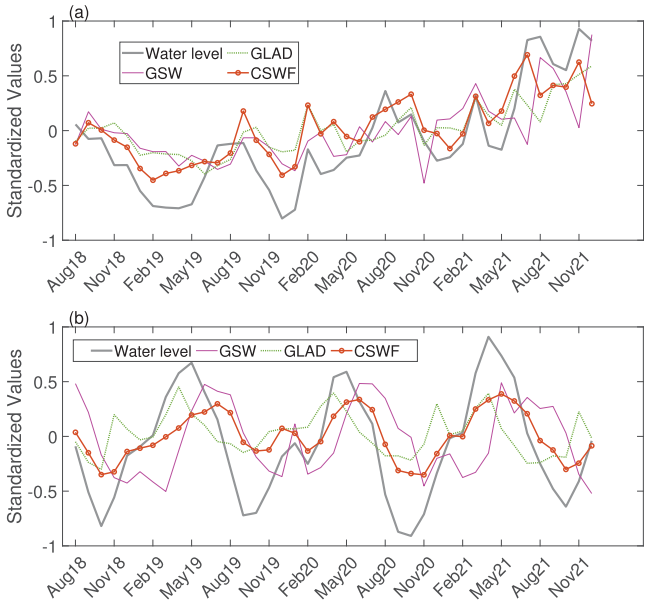


Fig. 14. Standardized WL measurements and SWF estimates. (a) Australia. (b) Amazon river.

TABLE VI

DETERMINATION COEFFICIENT R^2 BETWEEN SWF AND WL (* INDICATES THAT p-VALUE IS <0.05)

Region	Dataset	R^2
Australia	GSW	0.739*
	GLAD	0.592*
	CYGNSS	0.810*
Amazon River	GSW	0.369*
	GLAD	0.795*
	CYGNSS	0.845*

clear seasonal variations similar to those of WL. For some examples in Australia, seasonal patterns of WL may not be so clear, but CSWF provides good consistency. It should be noted that using data from only two annual cycles to fully capture seasonality in SWF dynamics may be challenging. This is because some regions may exhibit consistent seasonal patterns over consecutive years, but some can experience significant interannual variability due to a multitude of factors such as climatic anomalies, land use changes, and extreme weather events. In addition, by taking monthly means, the obtained SWF products may lose some short-term variability presented in WL data.

To further quantify the agreement between SWF and WL, the WL measurements were rescaled to monthly data with a spatial resolution of 0.025° for both Amazon and Australia. The derived SWF products have a spatial resolution of 0.025° at a monthly interval. Therefore, the point measurements by gauge data are aggregated into monthly time series and mapped into the 0.025°-grid as a representative value for corresponding pixels. In addition, there may be multiple stations within a grid cell, and then, data from those gauges are averaged under such conditions. In addition, both SWF and WL were normalized by subtracting their mean values and then divided by the corresponding standard deviation. The spatially averaged time series of normalized SWF and WL

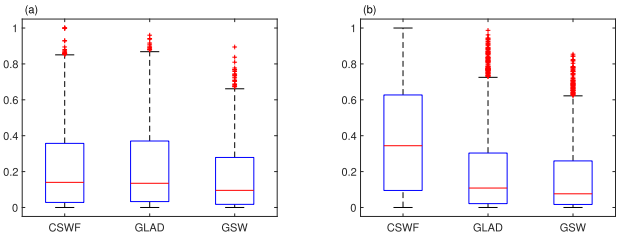


Fig. 15. Box plot of R^2 between SWF estimates and WL measurements in (a) Australia and (b) Amazon River.

during the 41-month period for both regions with WL station data is displayed in Fig. 14, and the R^2 values are summarized in Table VI. CSWF shows the best correlation with WL. In Fig. 14, it is noted that between August 2018 to mid-2020, both WL and SWF showed negative values, indicating severe drought conditions over Australia, which is in line with the findings in [49]. On the other hand, an upward trend in both WL and SWF has been observed since the beginning of the year of 2021. This is also supported by the fact that 2021 is the wettest year since 2016 with a national total rainfall of 9% higher than the 1961–1990 average [50]. The pixel-level comparison was also conducted for each 0.025° grid by calculating R^2 between corresponding WL and SWF. The box plots of R^2 based on each individual 0.025° pixel for both regions are shown in Fig. 15. It is clear that the performance in Australia is satisfactory for all three SWF estimates, while in the Amazon River region where vegetation and cloud covers are dense, CSWF (microwave) exhibits enhanced agreement than the Landsat-based (optical) GLAD and GSW.

V. CONCLUSION

In this work, a new scheme is proposed for estimating SWF from the CYGNSS $L1$ data via BARTs. The developed model incorporates DEM, slope, latitude, longitude, biomass, water seasonality, SMAP SM and VWC, CYGNSS Γ, and observable. The model is trained with SWAMPS data, which has a low level of detail (25 km), and then applied to CYGNSS data that have a higher resolution (0.025°) to make finer SWF maps. Comprehensive evaluation demonstrated the robustness and effectiveness of this method. The overall agreement between the global reference and the corresponding derivatives was satisfactory, with an overall r of up to 0.987 and an RMSE of 1.53% compared with the 25-km SWAMP set. Intercomparison among the GSW, GLAD, and CYGNSS results at both 25-km and 0.025° shows reasonable consistency, thus ensuring the robustness of the fine-resolution retrievals. Further investigation using hydrological station data demonstrates that the CYGNSS-based dataset achieves the best correlation with WL measurements based on Australian water station records and satellite altimetry over the Amazon River. In summary, the advantages of CYGNSS-based SWF in terms of an improved spatial coverage rate over the optical-based data and an enhanced spatial resolution against the SWAMPS are also illustrated. i The accuracy of the dataset relies on the quality and accuracy of the training reference (SWAMPS), which may underestimate SWF for areas with dense canopies and present limited seasonal variation. Ensuring reliable and

accurate validation is a significant challenge in these regions. In the future, a better suitable reference can be considered, and assessment with more ground-truth data and independent validation sources and other alternative existing methods (such as the radiometric model approach [35]) will be performed. As a future direction, this method will be improved for even finer spatiotemporal resolutions so that the SWF estimates can be better employed in hydrological modeling and so on. This can be achieved through integrating data from other GNSS-R missions since involving more data allows enhanced spatiotemporal coverage. In addition, this work employed and interpolated the 9-km SMAP data as auxiliary data; it may introduce some uncertainties. Using SM/VWC data with improved resolution may be helpful. While the current study relies on static data (e.g., biomass) due to the constraint of data availability, future research can benefit from using higher temporal resolution data when they become accessible.

ACKNOWLEDGMENT

The authors are grateful to the reviewers for their constructive feedback and insightful comments on this article.

REFERENCES

[1] M. A. Holgerson and P. A. Raymond, "Large contribution to inland water CO_2 and CH_4 emissions from very small ponds," *Nature Geosci.*, vol. 9, no. 3, pp. 222–226, Mar. 2016.

[2] S. M. Haig, S. P. Murphy, J. H. Matthews, I. Arismendi, and M. Safeeq, "Climate-altered wetlands challenge waterbird use and migratory connectivity in arid landscapes," *Sci. Rep.*, vol. 9, no. 1, Mar. 2019, Art. no. 4666.

[3] A. Karpatne, A. Khandelwal, X. Chen, V. Mithal, J. Faghmous, and V. Kumar, "Global monitoring of inland water dynamics: State-of-the-art, challenges, and opportunities," in *Computational Sustainability*. Cham, Switzerland: Springer, 2016, pp. 121–147.

[4] C. J. Vörösmarty, P. Green, J. Salisbury, and R. B. Lammers, "Global water resources: Vulnerability from climate change and population growth," *Science*, vol. 289, no. 5477, pp. 284–288, Jul. 2000.

[5] J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward, "High-resolution mapping of global surface water and its long-term changes," *Nature*, vol. 540, pp. 418–422, Dec. 2016. [Online]. Available: https://www.nature.com/articles/nature20584

[6] A. H. Pickens et al., "Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series," *Remote Sens. Environ.*, vol. 243, Jun. 2020, Art. no. 111792.

[7] *User Guides—Sentinel-1 SAR—Revisit and Coverage—Sentinel Online—Sentinel Online*. Accessed: Feb. 1, 2024. [Online]. Available: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/revisit-and-coverage

[8] K. Jensen and K. Mcdonald, "Surface water microwave product series version 3: A near-real time and 25-year historical global inundated area fraction time series from active and passive microwave remote sensing," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1402–1406, Sep. 2019.

[9] M. P. Clarizia, C. S. Ruf, P. Jales, and C. Gommenginger, "Spaceborne GNSS-R minimum variance wind speed estimator," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6829–6843, Nov. 2014.

[10] G. Foti et al., "Spaceborne GNSS reflectometry for ocean winds: First results from the UK TechDemoSat-1 mission," *Geophys. Res. Lett.*, vol. 42, no. 13, pp. 5435–5441, Jul. 2015.

[11] E. Cardellach et al., "Consolidating the precision of interferometric GNSS-R ocean altimetry using airborne experimental data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4992–5004, Aug. 2014.

[12] W. Li, E. Cardellach, F. Fabra, S. Ribo, and A. Rius, "Assessment of spaceborne GNSS-R ocean altimetry performance using CYGNSS mission raw data," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 238–250, Jan. 2020.

[13] C. C. Chew and E. E. Small, "Soil moisture sensing using spaceborne GNSS reflections: Comparison of CYGNSS reflectivity to SMAP soil moisture," *Geophys. Res. Lett.*, vol. 45, no. 9, pp. 4049–4057, May 2018.

[14] Q. Yan, W. Huang, S. Jin, and Y. Jia, "Pan-tropical soil moisture mapping based on a three-layer model from CYGNSS GNSS-R data," *Remote Sens. Environ.*, vol. 247, Sep. 2020, Art. no. 111944.

[15] Q. Yan and W. Huang, "Spaceborne GNSS-R sea ice detection using delay-Doppler maps: First results from the UK TechDemoSat-1 mission," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 10, pp. 4795–4801, Oct. 2016.

[16] Q. Yan and W. Huang, "Sea ice thickness measurement using spaceborne GNSS-R: First results with TechDemoSat-1 data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 577–587, 2020.

[17] C. S. Ruf et al., "A new paradigm in Earth environmental monitoring with the CYGNSS small satellite constellation," *Sci. Rep.*, vol. 8, no. 1, pp. 1–13, Dec. 2018. [Online]. Available: https://www.nature.com/articles/s41598-018-27127-4/

[18] C. Chew, J. T. Reager, and E. Small, "CYGNSS data map flood inundation during the 2017 Atlantic Hurricane season," *Sci. Rep.*, vol. 8, no. 1, pp. 1–8, Jun. 2018, doi: 10.1038/s41598-018-27673-x.

[19] P. Ghasemigoudarzi, W. Huang, O. De Silva, Q. Yan, and D. T. Power, "Flash flood detection from CYGNSS data using the RUSBoost algorithm," *IEEE Access*, vol. 8, pp. 171864–171881, 2020.

[20] S. Zhang et al., "POBI interpolation algorithm for CYGNSS near real time flood detection research: A case study of extreme precipitation events in Henan, China in 2021," *Adv. Space Res.*, vol. 71, no. 6, pp. 2862–2878, Mar. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0273117722010389

[21] N. Rodriguez-Alvarez, E. Podest, K. Jensen, and K. C. McDonald, "Classifying inundation in a tropical wetlands complex with GNSS-R," *Remote Sens.*, vol. 11, no. 9, p. 1053, May 2019.

[22] M. Morris, C. Chew, J. T. Reager, R. Shah, and C. Zuffada, "A novel approach to monitoring wetland dynamics using CYGNSS: Everglades case study," *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111417.

[23] C. Gerlein-Safdi, A. A. Bloom, G. Plant, E. A. Kort, and C. S. Ruf, "Improving representation of tropical wetland methane emissions with CYGNSS inundation maps," *Global Biogeochem. Cycles*, vol. 35, no. 12, pp. 1–14, Dec. 2021.

[24] C. Gerlein-Safdi and C. S. Ruf, "A CYGNSS-based algorithm for the detection of inland waterbodies," *Geophys. Res. Lett.*, vol. 46, no. 21, pp. 12065–12072, Nov. 2019.

[25] P. Ghasemigoudarzi, W. Huang, O. De Silva, Q. Yan, and D. Power, "A machine learning method for inland water detection using CYGNSS data," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[26] M. M. Al-Khaldi et al., "Inland water body mapping using CYGNSS coherence detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7385–7394, Sep. 2021.

[27] W. Li, E. Cardellach, S. Ribó, A. Rius, and B. Zhou, "First spaceborne demonstration of BeiDou-3 signals for GNSS reflectometry from CYGNSS constellation," *Chin. J. Aeronaut.*, vol. 34, no. 9, pp. 1–10, Sep. 2021.

[28] P. Zeiger, F. Frappart, J. Darrozes, C. Prigent, and C. Jiménez, "Analysis of CYGNSS coherent reflectivity over land for the characterization of pan-tropical inundation dynamics," *Remote Sens. Environ.*, vol. 282, Dec. 2022, Art. no. 113278. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425722003844

[29] B. Liu et al., "Statistical analysis of CyGNSS speckle and its applications to surface water mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5803915.

[30] Q. Yan et al., "Inland water mapping based on GA-LinkNet from CyGNSS data," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[31] E. Loria et al., "Comparison of GNSS-R coherent reflection detection algorithms using simulated and measured CYGNSS data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5105216.

[32] B. D. Chapman et al., "Comparison of SAR and CYGNSS surface water extent metrics," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3235–3245, 2022.

[33] B. Downs, A. J. Kettner, B. D. Chapman, G. R. Brakenridge, A. J. O'Brien, and C. Zuffada, "Assessing the relative performance of GNSS-R flood extent observations: Case study in South Sudan," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4201213.

[34] E. Loria, A. O'Brien, V. Zavorotny, B. Downs, and C. Zuffada, "Analysis of scattering characteristics from inland bodies of water observed by CYGNSS," *Remote Sens. Environ.*, vol. 245, Aug. 2020, Art. no. 111825.

[35] C. Chew, E. Small, and H. Huelsing, "Flooding and inundation maps using interpolated CYGNSS reflectivity observations," *Remote Sens. Environ.*, vol. 293, Aug. 2023, Art. no. 113598.

[36] H. Carreno-Luengo, A. Camps, J. Querol, and G. Forte, "First results of a GNSS-R experiment from a stratospheric balloon over boreal forests," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2652–2663, May 2016.

[37] M. P. Clarizia, N. Pierdicca, F. Costantini, and N. Floury, "Analysis of CYGNSS data for soil moisture retrieval," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2227–2235, Jul. 2019.

[38] O. Eroglu, M. Kurum, D. Boyd, and A. C. Gurbuz, "High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks," *Remote Sens.*, vol. 11, no. 19, p. 2272, Sep. 2019.

[39] P. E. O'Neill et al., *SMAP Enhanced L3 Radiometer Global and Polar Grid Daily 9 km EASE-Grid Soil Moisture*, document Version 5, 2021. [Online]. Available: https://nsidc.org/data/SPL3SMP_E/versions/5

[40] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, Dec. 2017.

[41] Q. Wu, "Geemap: A Python package for interactive mapping with Google Earth Engine," *J. Open Source Softw.*, vol. 5, no. 51, p. 2305, Jul. 2020.

[42] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Evanston, IL, USA: Routledge, Oct. 1984.

[43] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.

[44] S. Chakrabarti, J. Judge, T. Bongiovanni, A. Rangarajan, and S. Ranka, "Spatial scaling using temporal correlations and ensemble learning to obtain high-resolution soil moisture," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1238–1250, Mar. 2018.

[45] Q. Yan, S. Gong, S. Jin, W. Huang, and C. Zhang, "Near real-time soil moisture in China retrieved from CyGNSS reflectivity," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[46] Z. Zhang et al., "Development of the global dataset of wetland area and dynamics for methane modeling (WAD2M)," *Earth Syst. Sci. Data*, vol. 13, no. 5, pp. 2001–2023, May 2021.

[47] N. Pierdicca, F. Fascetti, L. Pulvirenti, R. Crapolicchio, and J. Muñoz-Sabater, "Quadruple collocation analysis for soil moisture product assessment," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 8, pp. 1595–1599, Aug. 2015.

[48] C. Prigent, C. Jimenez, and P. Bousquet, "Satellite-derived global surface water extent and dynamics over the last 25 years (GIEMS-2)," *J. Geophys. Res., Atmos.*, vol. 125, no. 3, pp. 1–18, Feb. 2020.

[49] B. Fang, P. Kansara, C. Dandridge, and V. Lakshmi, "Drought monitoring using high spatial resolution soil moisture data over Australia in 2015–2019," *J. Hydrol.*, vol. 594, Mar. 2021, Art. no. 125960.

[50] (2021). *Annual Climate Statement 2021: Wetter and Warmer Than Average—Social Media Blog—Bureau of Meteorology*. [Online]. Available: https://media.bom.gov.au/social/blog/2669/annual-climate-statement-2021-wetter-and-warmer-than-average/