https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

GWL_FCS30: global 30 m wetland map with fine classification system using multi-sourced and time-series remote sensing imagery in 2020

Xiao Zhang[1, 2], Liangyun Liu[1, 2, 3, *], Tingting Zhao[1, 4], Xidong Chen[5], Shangrong Lin[6], Jinqing Wang[1, 2, 3], Jun Mi[1, 2, 3] and Wendi Liu[1, 2, 3]

1 International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China
2 Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
3 School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
4 College of Geomatics, Xi'an University of Science and Technology, Xi'an 710054, China
5 North China University of Water Resources and Electric Power, Zhengzhou 450046, China
6 School of Atmospheric Sciences, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Sun Yat-sen University, Zhuhai 519082, Guangdong, China
* Corresponding author Email: liuly@radi.ac.cn

Abstract

Wetlands, often called the "kidneys of the earth", play an important role in maintaining ecological balance, conserving water resources, replenishing groundwater, and controlling soil erosion. Wetland mapping is very challenging because of its complicated temporal dynamics and large spatial and spectral heterogeneity. An accurate global 30-m wetland dataset that can simultaneously cover inland and coastal zones is lacking. This study proposes a novel method for wetland mapping by combining an automatic sample extraction method, multisource existing products, time-series satellite images, and a stratified classification strategy. This approach allowed for the generation of the first global 30-m wetland map with a fine classification system (GWL_FCS30), including four inland wetland sub-categories (swamp, marsh, flooded flat, and saline) and three coastal wetland sub-categories (mangrove, salt marsh, and tidal flats), which was developed using Google Earth Engine platform. We first combined existing multi-sourced global wetland products, expert knowledge, training sample refinement rules, and visual interpretation to generate a large and geographically distributed wetland training samples. Second, we integrated the time-series Landsat reflectance products and Sentinel-1 SAR imagery to generate water-level and phenological information to capture the complicated temporal dynamics and spectral heterogeneity of wetlands. Third, we applied a stratified classification strategy and the local adaptive random forest classification models to produce the wetland dataset with a fine classification system at each 5°×5° geographical tile in 2020. Lastly, the GWL_FCS30, mosaicked by 961 5°×5° regional wetland maps, was validated using 18,701 validation samples, which achieved an overall accuracy of 87.7% and a kappa coefficient of 0.810. The cross-comparisons with other global wetland products demonstrated that the GWL_FCS30 dataset performed better in capturing the spatial patterns of wetlands and had significant advantages over the diversity of wetland subcategories. The statistical analysis showed that the global wetland area reached 3.57 million km^2, including 3.10 million km^2 of inland wetlands and 0.47 million km^2 of coastal wetlands, approximately 62.3% of which were distributed poleward of 40°N. Therefore, we can conclude that the proposed method is suitable for large-area wetland mapping and that the GWL_FCS30 dataset is an accurate wetland mapping product that has the potential to provide vital support for wetland management. The GWL_FCS30 dataset in 2020 is freely available at https://doi.org/10.5281/zenodo.6575731 (Liu et al. 2022).

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

1. Introduction

The RAMSAR Convention defines a wetland as an "areas of marsh, fen, peatland or water, whether natural or artificial, permanent or temporary, with water that is static or flowing, fresh, brackish or salt, including areas of marine water the depth of which at low tide does not exceed six meters" (Gardner and Davidson, 2011). Wetlands not only provide humans with a large amount of food, raw materials and water resources (Ludwig et al., 2019; Zhang et al., 2022b) but also play an important role in maintaining ecological balance, conserving water resources, replenishing groundwater, and controlling soil erosion (Hu et al., 2017a; Mao et al., 2021; Wang et al., 2020; Zhu and Gong, 2014). Therefore, they are also called the "kidneys of the earth" (Guo et al., 2017). However, due to increasing human activities, including agriculturalization, industrialization and urbanization (McCarthy et al., 2018; Xi et al., 2020), and climatic changes such as sea-level rise and coastal erosion (Cao et al., 2020; Wang et al., 2021), wetlands have been seriously degraded and threatened over the past few decades (Mao et al., 2020). Thus, having access to timely and accurate wetland mapping information is pivotal for protecting biodiversity and supporting the sustainable development goals.

Along with the rapid development of remote sensing techniques and computing abilities, a variety of regional or global wetland datasets have been produced with spatial resolutions ranging from 30 m to 1° (~112 km) (Chen et al., 2022; Gumbricht et al., 2017; Lehner and Döll, 2004; Mao et al., 2020; Matthews and Fung, 1987; Tootchi et al., 2019). Tootchi et al. (2019) and Hu et al. (2017a) have systematically reviewed the generation process of global wetland datasets with various spatial and temporal resolutions and wetland categories and found significant uncertainties and inconsistencies among these datasets. For example, the global total wetland area reviewed by Hu et al. (2017a) ranged from 2.12 to 7.17 million km^2 based on remote sensing products. Therefore, great uncertainties among global wetland datasets directly hindered wetland applications and analysis. Furthermore, from the perspective of spatial resolution, although many wetland products have been produced, at regional or global scales, using various remote sensing imagery and different methods (Guo et al., 2017; Tootchi et al., 2019), most of them were coarse spatial resolution datasets, ranging from 100 m to 25 km. Recently, with the improvement of computing power and storage abilities, three global 30-m land-cover products (including GlobeLand30 (Chen et al., 2015), FROM_GLC (Gong et al., 2013) and GLC_FCS30 (Zhang et al., 2021b)), containing an independent wetland layer, were produced, but their classification algorithms were not specifically designed for the wetland environment, so the wetland usually suffered from low accuracy in these products. In addition, several global coastal wetland products have been developed, including the global mangrove extent (Bunting et al., 2018; Hamilton and Casey, 2016) and global 30 m tidal flat datasets from 1984 to 2016 (Murray et al., 2019), but these only covered the intertidal zones. Thus, an accurate global 30 m wetland dataset, with fine wetland categories and covering both inland and coastal zones, is still lacking.

One of the largest challenges of current state-of-the-art methods for large-area wetland mapping is to collect massive amount of training samples (Liu et al., 2021; Ludwig et al., 2019). Zhang et al. (2021b) mentioned two options for collecting training samples, including the visual interpretation method and deriving training samples from pre-existing products. First, since the visual interpretation method had significant advantage over the confidence of training samples, it was widely used for local or regional wetland mapping (Amani et al., 2019; Wang et al., 2020). However, collecting accurate and sufficient training samples is usually a time-consuming process and involves a large amount of manual work, so it was impractical and nearly impossible to use the visual interpretation for collecting global wetland samples. Comparatively, deriving training samples from

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

existing products and applying some rules or refinement methods to identify these high confidence samples from existing products shows promise (Zhang et al., 2021b). So this approach is practical in that it could quickly large and geographically diverse distribution of training samples without much manual effort. Thus, the second option attached increasing attention and has been successfully used for large-area land-cover mapping (Zhang and Roy, 2017; Zhang et al., 2021b; Zhang et al., 2020). For example, Zhang et al. (2021b) used derived global training samples from the combination of the CCI_LC and MCD43A4 NBAR datasets to produce a global 30-m land-cover product with a fine classification system in 2015 and 2020 (GLC_FCS30) with an overall accuracy of 82.5%. Therefore, if we take effective measures to fuse these existing products and then derive high confidence training samples using some refinement rules, the deriving approach would exude great potential for global wetland mapping.

Another major challenge inherent to wetland mapping is the complicated temporal dynamics and spatial and spectral heterogeneity. The spectral characteristics of the wetlands would quickly change with the seasonal or daily water levels of the underlying surface (Ludwig et al., 2019; Mahdianpari et al., 2020). Therefore, many studies proposed to combine multi-sourced, time-series remote sensing imagery for capturing the spatial extent and temporal dynamics of wetlands (LaRocque et al., 2020; Ludwig et al., 2019; Murray et al., 2019; Wang et al., 2021; Zhang et al., 2022b). For example, Zhang et al. (2022b) and Murray et al. (2019) used the time-series Landsat imagery to generate tidal-level and phenological features for identifying coastal wetlands and successfully produced the coastal wetlands in China with an overall accuracy of 97.2% (Zhang et al., 2022b) and global trajectory tidal flats with the overall map accuracy of 82.3% (Murray et al., 2019). Except for optical imagery, synthetic aperture radar (SAR) data, which was sensitive to soil moisture, vegetation structure, and inundation, enabled data acquisition regardless of solar illumination, clouds, or haze and was also widely used for wetland mapping, especially after the open-access of Sentinel-1 data became available (Li et al., 2020; Slagter et al., 2020; Zhang et al., 2018). For example, Li et al. (2020) used the time-series Sentinel-1 imagery to discriminate wetlands with and without trees and achieved an overall accuracy of 86.0±0.2%. Therefore, the fusion of multi-sourced and time-series remote sensing imagery is vital for accurate wetland mapping.

Due to the complicated temporal dynamics and spatial and spectral heterogeneity of wetlands, global 30 m wetland mapping remains a challenging task. Consequently, there is no global 30-m dataset covering both inland and coastal wetlands until now. In this study, we combined several existing wetland products and multi-sourced time-series remote sensing imagery to (1) derive a large and geographically distributed wetland training samples from multi-sourced pre-existing global wetland products to minimize the manual participation; (2) develop a robust method to capture the temporal dynamics of wetlands and then produce the first global 30-m wetland dataset with a fine classification system (GWL_FCS30); (3) quantitatively analyze the spatial distribution of different wetland categories and assess the accuracy of the GWL_FCS30 in 2020.

2. Datasets

2.1 Multi-sourced remote sensing imagery

Three types of remote sensing imagery were collected to capture the temporal dynamics and spatial and spectral heterogeneity of wetlands. These include Landsat optical data, Sentinel-1 SAR, and ASTER GDEM topographical data. First, all available Landsat imagery during 2019–2021 was obtained via the Google Earth Engine platform. To minimize the effect of atmosphere, each Landsat image was atmospherically corrected to the surface reflectance by the United States Geological Survey using Land Surface Reflectance Code (LaSRC)

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

method (Vermote et al., 2016) and then archived on the GEE platform. And these 'bad quality' observations (shadow, cloud, snow, and saturated pixels) in Landsat imagery were masked using CFmask cloud detection method (Zhu et al., 2015; Zhu and Woodcock, 2012). In this study, 764,239 Landsat scenes were collected to extract the water-level and phenological features presented in Section 4. Figure 1a illustrates the spatial distribution of all clear-sky Landsat imagery. It can be seen that there were more than 10 Landsat observations at each scene, including the tropics. Then, the Sentinel-1 SAR data, which was demonstrated to be sensitive to the soil moisture, vegetation structure, and inundation information (Li et al., 2020), used dual-polarization C-band backscatter coefficients to measure the incident microwave radiation scattered by the land surface (Torres et al., 2012). This study obtained the time-series Sentinel-1 imagery archived on the GEE platform in 2020 in Interferometric Wide Swath mode with a dual-polarization of VV and VH. Notably, all Sentinel-1 SAR imagery on the GEE platform has been pre-processed by the Sentinel-1 Toolbox with thermal noise removal, radiometric calibration, and terrain correction using 30-m elevation data (Veci et al., 2014). Figure 1b also illustrates the spatial distribution of all available Sentinel-1 SAR imagery. We found that there were enough Sentinel-1 SAR observations in each area to capture the temporal dynamics of wetlands. Lastly, as many studies have demonstrated that the topography would directly affect the spatial distribution of wetlands, which are mainly distributed in low-lying areas (Hu et al., 2017b; Ludwig et al., 2019; Tootchi et al., 2019), the ASTER GDEM elevation and derived slope and aspect were used as auxiliary information for wetland mapping. It had a spatial resolution of 30 m and covered the entire global land area (Tachikawa et al., 2011a). Quantitative assessment indicated that the GDEM achieved an absolute vertical accuracy of 0.7 m over bare areas and 7.4 m over forested areas (Tachikawa et al., 2011b).

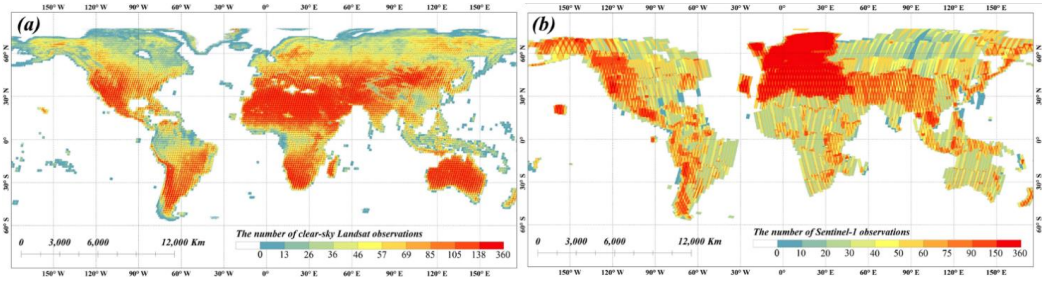


Figure1. The availability of clear-sky Landsat observations (a) and Sentinel-1 SAR imagery (b).

2.2 Global prior wetland datasets

To achieve the goal of deriving a large and geographically diverse distribution of training samples with minimum manual labor, we propose combining various prior global wetland datasets for generating high-confidence training samples. Table 1 lists the characteristics of several global wetland datasets. Specifically, we collected five global mangrove forest products with different spatial resolutions and time spans, and all of them achieved desirable accuracy. For example, the Global Mangrove Watch (GMW) was validated to reach an overall accuracy of 95.25%, and the user and producer accuracies of mangrove forest were 97.5% and 94.0%, respectively (Thomas et al., 2017). Furthermore, to derive the samples of salt marsh and tidal flats, we collected the time-series global 30-m tidal flats products from 1984 to 2016 with an interval of three years, achieving an overall map accuracy of 82.3% (Murray et al., 2019). The global salt marsh dataset, containing 350,985 individual occurrence polygon shapefiles, helped generate the global salt marsh estimation (McOwen et al., 2017).

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

Except for the coastal wetland products, two thematic wetland products (TROP-SUBTROP Wetland and GLWD contained various wetland sub-categories), three global land-cover products (GlobeLand30, GLC_FCS30, and CCI_LC contained an independent layer), and the time-series 30-m water dynamic dataset (JRC_GSW) were combined to determine the inland maximum wetland extents and generate the wetland training samples after using a series of refinement rules given in Section 3. Specifically, the TROP-SUBTROP was produced by combining the hydrological model and annual time series of soil moisture, mainly covering the tropics and sub-tropics (40°N ~ 60°S) with a resolution of 231 m (Gumbricht, 2015). The GLWD, combining the GIS functionality and a variety of existing maps and information, was developed with 12 wetland sub-categories at a resolution of 1 km (Lehner and Döll, 2004). The JRC_GSW dynamic water dataset achieved a producer accuracy of 98.5% for these seasonal waters (Pekel et al., 2016) and was used to capture those wetlands around rivers, ponds, etc. Furthermore, three global land-cover products, simultaneously containing wetland layer and non-wetland land-cover types, were used to determine the non-wetland samples and then served as the auxiliary datasets to improve the confidence of inland wetland samples.

Table 1. The characteristics of 13 global wetland products with various spatiotemporal resolutions.

Dataset name	Wetland categories	Year	Resolution	Reference
World atlas of mangroves (WAM)	Mangrove	2010	1:1,000,000	Spalding (2010)
Global mangrove watch (GWM)		1996-2016	0.8 seconds	Thomas et al. (2017)
A global biophysical typology of mangroves (GBTM)		1996-2016	0.8 seconds	Worthington et al. (2020)
Continuous global mangrove forest cover (CGMFC)		2000-2010	30 m	Hamilton and Casey (2016)
Global distribution of mangroves USGS (GDM_USGS)		2011	30 m	Giri et al. (2011)
Global distribution of tidal flat ecosystems	Tidal flat	1984-2016	30 m	Murray et al. (2019)
Global distribution of saltmarsh	Salt marsh	1973-2015	1:10,000	McOwen et al. (2017)
Tropical and subtropical wetland distribution (TROP-SUBTROP Wetland)	Open water, mangrove, swamps, fens, riverine, floodplains, and marshes	2011	~231 m	Gumbricht (2015)
Global lakes and wetlands database (GLWD)	Lake, reservoir, river, freshwater marsh, swamps, coastal wetland, saline wetland, and Peatland	2004	30 second (~1 km)	Lehner and Döll (2004)
JRC-GSW	Water	1984-2021	30 m	Pekel et al. (2016)
CCI_LC	Swamps, mangrove, and Shrub or herbaceous cover wetlands	1992-2020	300 m	Defourny et al. (2018)

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

GlobeLand30	Wetland	2000-2020	30 m	Chen et al. (2015)
GLC_FCS30	Wetland	2015, 2020	30 m	Zhang et al. (2021b)

2.3 Global 30 m tree cover product

The global 30-m forest cover change in tree cover (GFCC30TC) data in 2015 was produced by downscaling the 250-m MODIS VCF (Vegetation Continuous Fields) tree cover product using Landsat imagery and then incorporating the MODIS cropland layer to guarantee the tree cover accuracy in agricultural areas (Sexton et al., 2016; Sexton et al., 2013). This product was used to accurately distinguish between inland swamp and marsh wetlands because both of them reflected obvious vegetation spectra characteristics. It was validated to achieve an overall accuracy of 91%; the average producer and user accuracy for stable forests were 92.5% and 95.4%, respectively (Sexton et al., 2016; Townshend et al., 2012).

2.4 Global wetland validation dataset

To quantitatively analyze the performance of our GWL_FCS30 wetland map, a total of 18,701 validation samples, including 10,346 non-wetland points and 8,355 wetland points, were collected by combining high-resolution imagery, time-series Landsat and Sentinel observations and visual interpretation method. Firstly, as the wetland was sparse land-cover type compared to the non-wetlands (forest, cropland, grassland and bare land), the stratified random strategy was applied to randomly derive validation points at each strata. Then, as the wetlands had significant correlation with the water levels (Zhang et al., 2022b), the time-series optical observations archived on the GEE cloud platform were used as the auxiliary dataset to interpret these water-level sensitive wetlands such as: tidal flat and flooded flat. It should be noted that the visual interpretation was implemented on the GEE cloud platform because it archives a large amount of satellites imagery with various time spans and spatiotemporal resolution (Zhang et al., 2022a). Meanwhile, each validation point is independently interpreted by five experts for minimizing the effect of expert's subjective knowledge, and only these complete agreement points were retained otherwise they were discarded. Figure 2 intuitively illustrated the spatial distribution of global wetland validation points, it can be found that the distribution of wetland points accurately revealed the spatial patterns of global wetlands.

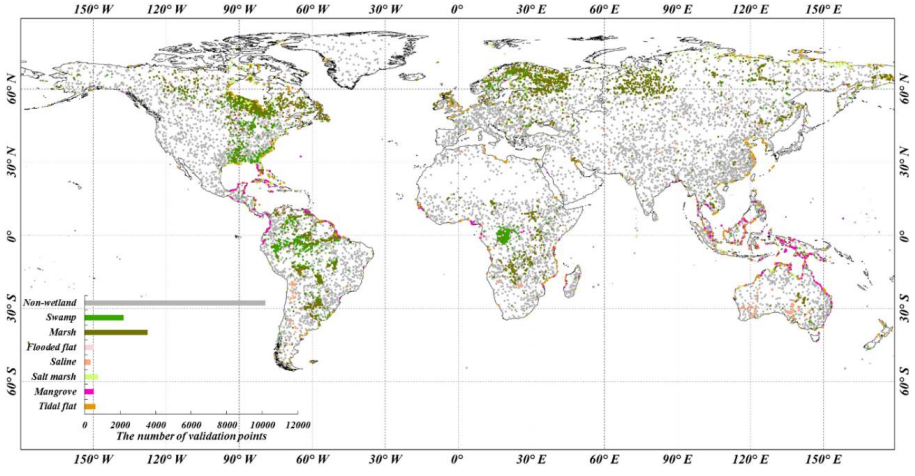


Figure 2. The spatial distribution of 18,701 global wetland validation samples using stratified sampling strategy.

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

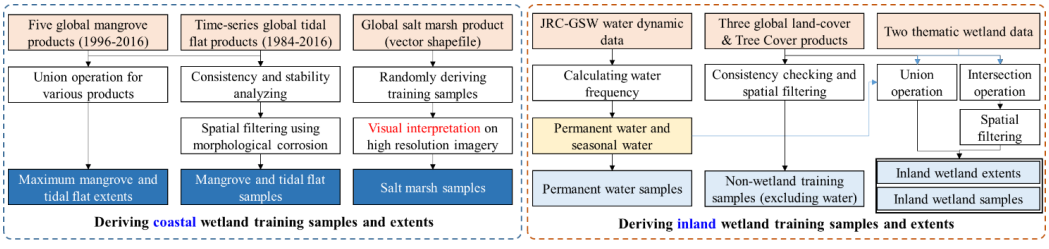
3. Collecting training samples and determining maximum wetland extents

In this study, after considering the applicability of moderate resolution (10–30 m) imagery, their practical use for ecosystem management, and the available pre-existing global wetland dataset, the fine wetland classification system, containing seven sub-categories (three coastal sub-categories and four inland sub-categories), was proposed to comprehensively depict the spatial patterns of global wetlands (Table 2). Specifically, the sub-categories of coastal wetlands consist of mangroves, salt marshes, and tidal flats. By importing the vegetation and water cover information associated with this land cover, these categories were widely recognized in many previous studies (Wang et al., 2021; Zhang et al., 2022b). The inland wetland types shared similar characteristics and were grouped into swamp, marsh, and flooded flat. Except for the freshwater-related wetlands, the inland saline wetland, inherited from the Global Lakes and Wetlands Dataset (GLWD) (Lehner and Döll, 2004), was also imported to capture saline soils and halophytic plant species along saline lakes.

Table 2. The description of wetland classification system in this study

Category I	Category II	Description
Coastal wetland	Mangrove	The forest or shrubs which grow in the coastal blackish or saline water
	Salt marsh	Herbaceous vegetation (grasses, herbs and low shrubs) in the upper coastal intertidal zone
	Tidal flat	The tidal flooded zones between the coastal high and low tide levels including mudflats and sandflats.
Inland wetland	Swamp	The forest or shrubs which grow in the inland freshwater
	Marsh	Herbaceous vegetation (grasses, herbs and low shrubs) grows in the freshwater
	Flooded flat	The non-vegetated flooded areas along the rivers and lakes
	Saline	Characterized by saline soils and halophytic (salt tolerant) plant species along saline lakes

Many studies have explained that the quality and confidence of training samples directly affected the classification performance (Zhang et al., 2021b; Zhu et al., 2016). The previously mentioned process of collecting sufficient training samples via visual interpretation was time-consuming and involved a lot of manual labor. Fortunately, a variety of regional and global wetland products have been developed and released over the past few decades (Table 1), and many studies have demonstrated that deriving training samples from existing products could be used for large-area classification and mapping (Huang et al., 2021; Zhang et al., 2021b). Therefore, we propose to combine existing global wetland datasets to independently derive coastal/inland wetland training samples and their maximum distribution extents (Figure 3).



https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

Figure 3. The flowchart of deriving coastal and inland wetland samples from multiple pre-existing datasets

3.1 Deriving coastal wetland training samples and maximum extents

This study divided the coastal wetlands into three sub-categories: mangrove forest, salt marsh, and tidal flat. The previously existing products have been collected in Table 1. For the mangrove training samples, we collected five global mangrove products with different spatiotemporal resolutions, all of which achieved fulfilling performances. For example, Hamilton and Casey (2016) stated that their continuous mangrove forest cover (CGMFC) dataset could cover 99% of all mangrove forests from 2000 to 2012, and Thomas et al. (2017) validated their Global Mangrove Watch (GMW) products from 1996 to 2016 and reached an overall accuracy of 95.25%. Therefore, we first measure the temporal consistency of the three time-series mangrove forest products (CGMFC, GMW, and GBTM mangroves), and only these temporally stable mangrove forest pixels were selected as the primary candidate points ($P_{mangrove}^{Tstable}$). Meanwhile, to minimize the influence of classification error in each mangrove forest product, the cross-consistency of five mangrove products was analyzed, and only the pixel, simultaneously identified as mangrove forest in all five products, was labeled as stable and consistent candidate points ($P_{mangrove}^{Tstable,Scons}$). Furthermore, considering that there was a temporal interval between prior mangrove products and our study, and that mangrove deforestation usually followed the pattern of edge-to-center contraction, a morphological erosion filter with a local window of 3×3 was applied to the $P_{mangrove}^{Tstable,Scons}$ points to further ensure the confidence of mangrove training samples. Lastly, as for the maximum mangrove forest extents ($MaxExtent_{mangrove}$), the union operation was applied to five global mangrove products as shown in Eq. (1).

$$MaxExtent_{mangrove} = M_{WAM} \cup M_{GMW} \cup M_{GBTM} \cup M_{CGMFC} \cup M_{GDM\_USGS} \quad (1)$$

where $[M_{WAM}, M_{GMW}, M_{GBTM}, M_{CGMFC}, M_{GDM\_USGS}]$ are the spatial distributions of five global mangrove forest products listed in Table 1. It should be noted that these prior mangrove products were demonstrated to cover almost all mangroves over the world, so the $MaxExtent_{mangrove}$ can be used as the boundary for mangrove mapping; namely, only the pixel within the maximum mangrove extent was labeled as mangrove forest.

Regarding the collection of tidal flat samples, the prior time-series global 30 m tidal flat products ($Gtidalflat$) from 1984 to 2016 were validated to achieve an overall map accuracy of 82.3%, and user accuracies for the non-tidal and tidal flat of 83.3% and 81.1%, respectively (Murray et al., 2019). To ensure the accuracy of tidal flat samples, we first applied temporal consistency analysis to the time series of tidal flat datasets from 2000 to 2016 and identified the temporally stable tidal flat pixels ($P_{tidal}^{Tstable}$) during 16 consecutive years. The reason why we discarded the tidal flat datasets before 2000 was that the available Landsat imagery was sparse and could not accurately capture the high-tidal and low-tidal information, and suffered lower monitoring accuracy. Next, Radoux et al. (2014) found that transition zones between two different land-cover types are likely to be misclassified; therefore, the candidate tidal flat samples $P_{tidal}^{Tstable}$ were further refined by the morphological erosion filter with a local window of 3×3. Furthermore, as a tidal flat is a non-vegetated coastal wetland, we combined the empirical rule (EVI ≥ 0.1, NDVI ≥ 0.2, and LSWI > 0) proposed by Wang et al. (2020) and time-series Landsat imagery in 2020 to exclude all vegetated pixels from tidal flat training samples. Lastly, to derive the maximum tidal flat extents ($MaxExtent_{tidalflat}$), the union operation was applied to the time-series tidal flat products from 1984 to 2016. It should be noted that the Murray's global 30 m tidal flat datasets only covered the regions of 60°N~60°S (Murray et al., 2019), therefore, we used the coastal

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

shorelines ($Line_{coastal}$) to create a 50 km buffer (applied by the Wang et al. (2020) and (Murray et al., 2019)) as the potential tidal flat zones in the high latitude regions (>60°N) as in Eq. (2).

$$MaxExtent_{tidalflat} = \begin{cases} \cup_{t=1984}^{2016} Gtidalflat_{t,s}, & s \in [60°S,\ 60°N] \\ Line_{coastal} \pm 50km, & s \in [60°N,\ 90°N] \end{cases} \quad (2)$$

Compared with the mangrove forest and tidal flat, the pre-existing global or regional salt marsh products were relatively sparse. The global distribution of the salt marsh dataset contained 350,985 individual vector polygons and was the most complete dataset on salt marsh occurrence and extent at the global scale (McOwen et al., 2017). However, after careful review, we found some mislabeled salt marsh polygons, so this dataset cannot be used directly to derive training samples. This study first used the random sampling method to generate 35,099 salt marsh points (approximately 10% of the total polygons) based on prior datasets. We combined the visual interpretation method and high-resolution imagery to check each salt marsh point. After discarding the incorrect and uncertain samples, a total of 32,712 salt marsh points were retained. However, the prior dataset only captured the extent of salt marshes in 99 countries worldwide (McOwen et al., 2017), further noting that the distribution of salt marshes was spatially correlated with tidal flat and mangrove forest (Wang et al., 2021). Consequently, the maximum extents of tidal flat and mangrove forest, in addition to the prior salt marsh extent were used for salt marsh mapping. Meanwhile, as the wetland layer in the global land-cover products (GLC_FCS30, GlobeLand30, and CCI_LC) also covered some coastal wetlands, the wetland layers in these land-cover products over coastal regions were also imported.

3.2 Deriving inland wetland training samples and maximum extents

The pre-existing inland wetland datasets usually suffered from lower accuracy compared to coastal wetland products; for example, the wetland layer in the GlobeLand30-2010 and GLC_FCS30-2015 was validated to achieve a user accuracy of 74.9% (Chen et al., 2015) and 43.4% (Zhang et al., 2021b), respectively. Therefore, we first generated high-confidence inland wetland samples and then determined their sub-categories (swamp, marsh, inland flat, and saline wetland). Specifically, the consistency analysis of five global wetland datasets (TROP-SUBTROP Wetland, GLWD, CCI_LC, GlobeLand30, and GLC_FCS30) and the temporal stability checking for CCI_LC (1992–2020), GlobeLand30 (2000-2020) and GLC_FCS30 (2015-2020) were applied to identify these temporally stable and high cross-consistency wetland points ($P_{inlandWet}^{Tstable,Scons}$). Namely, only the pixel identified as inland wetland (excluding permanent water bodies) in all five products was retained. Then, the morphological erosion filter with a local window of 3 × 3 was also used to decrease the sampling uncertainty over these land-cover transition areas because the transition zones between two different land-cover types are likely to be misclassified (Lu and Wang, 2021; Radoux et al., 2014).

Afterward, to determine the wetland sub-category for each inland wetland sample, we first used the empirical vegetation rule (EVI ≥ 0.1, NDVI ≥ 0.2, and LSWI > 0) proposed by Wang et al. (2020) and time-series Landsat imagery to split candidate samples into two parts: vegetated wetland samples (swamp and marsh) and non-vegetated wetland samples (flooded flat and saline). Then, as the swamp was defined as the forest or shrubs which grow in the inland freshwater, the global 30-m tree cover dataset (GFCC30TC) was adopted to distinguish the swamp and marsh from vegetated wetland samples. Specifically, if the tree cover of the sample was greater than 30% (Hansen et al., 2013), it was labeled as swamp, and the remaining vegetated wetland samples were labeled as marsh. Furthermore, to distinguish between the inland flat and saline samples from these non-vegetated wetland samples, the saline blocks in the prior GLWD products were first checked by visual

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.


interpretation and then imported as the reference dataset to identify all saline wetland samples. The remaining non-wetland samples were labeled as flooded flat. Lastly, as for determining the maximum inland wetland extents ($Mextent_{inWet}$), the union operation was conducted to six pre-existing global datasets as in Eq. (3).

$$Mextent_{inWet} = W_{\mathrm{TROP-SUBTROP}} \cup W_{GLWD} \cup W_{CCI\_LC} \cup W_{GLC\_FCS30} \cup W_{Globeland30} \cup W_{JRC\_GSW} \quad (3)$$

Here, $[W_{\mathrm{TROP-SUBTROP}}, W_{GLWD}, W_{CCI\_LC}, W_{GLC\_FCS30}, W_{Globeland30}]$ were wetland distributions of five pre-existing global wetland products. To comprehensively capture these fragmented and small river and lake wetlands, the seasonal water extents derived from the JRC-GSW time-series water dynamic datasets (W_{JRC_GSW}) were also added to $Mextent_{inWet}$. Specifically, as the time series of the JRC-GSW datasets provided the water probability at a monthly history for 1984–2021 (Pekel et al., 2016), the seasonal water body could be separated by the water probability using the threshold of 0.95 suggested by Wang et al. (2020).

3.3 Deriving non-wetland training samples from prior land-cover products

Except for coastal and inland wetland samples, the non-wetland samples were also necessary because some non-wetland land-cover types were shown to have a similar spectrum to wetlands. For example, swamp and forest or shrubs exhibited the same vegetation reflectance characteristics in optical imagery, and marsh and grassland shared similar spectra curves during the growing season (Zhang et al., 2022b). This study divided the non-wetlands into forest/shrubland, grassland, cropland, permanent water, and others (bare land, impervious surfaces, and snow). To automatically derive these non-wetland samples, the multi-epochs GlobeLand30, GLC_FCS30, CCI_LC global land-cover products, and the JRC-GSW water dynamic dataset were integrated. Specifically, the temporal stability and cross-consistency analysis were applied to three land-cover products to identify temporally stable forest/shrubland, grassland, cropland, and other candidate samples. Furthermore, the morphological erosion filter with the local window of 3 × 3 was also adopted to decrease the sampling uncertainty over land-cover transition areas. Regarding the permanent water samples, the JRC_GSW water dynamic dataset was validated and achieved producer's and user's accuracies of 99.7% and 99.1% for permanent water (Pekel et al., 2016). The permanent water training samples were directly derived from the JRC_GSW dataset without any refinement rules.

Except for the confidence of training samples, many studies also found that the size and distribution of training samples also affected classification performances (Jin et al., 2014; Zhu et al., 2016). As this study aimed to identify wetlands instead of all land-cover types, the equal allocation sample distribution would perform better than the proportional distribution (the sample size determined by the area) (Jin et al., 2014; Zhang et al., 2020). Namely, the approximate proportion of inland wetland, coastal wetland, and non-wetland samples was 4:3:5 in the coexisting areas because the classification system was composed of four inland and three coastal wetland sub-categories and five non-wetland land-cover types. Regarding the sample size, Zhu et al. (2016) had analyzed the quantitative relationships of sample size and the mapping accuracy and found that the mapping accuracies slowly increased and then remained stable with any further increase in the number of samples and suggested using a total of 20,000 samples in the Landsat scene. In this study, we used the stratified random sampling strategy to collect the training samples (excluding salt marsh because it was collected globally using visual interpretation in Section 3.1) at each 2° × 2° geographical grid using an approximate sample size of 2000 for each category. According to our statistics, this study derived exceeding 20 million training samples for mapping global fine wetlands.

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

4. Mapping wetland using the water-level and phenological features and the stratified classification strategy

Considering that the spectral characteristics of the wetlands would quickly change with the seasonal or daily water levels of the underlying surface, the time-series Landsat-8 and Sentinel-1 SAR observations, and ASTER DEM topographical image were combined to capture the complicated temporal dynamics and spectral heterogeneity. Figure 4 illustrates the flowchart of the proposed method for generating the water-level and phenological features and developing a global 30-m fine wetland map using a stratified random forest modeling strategy.

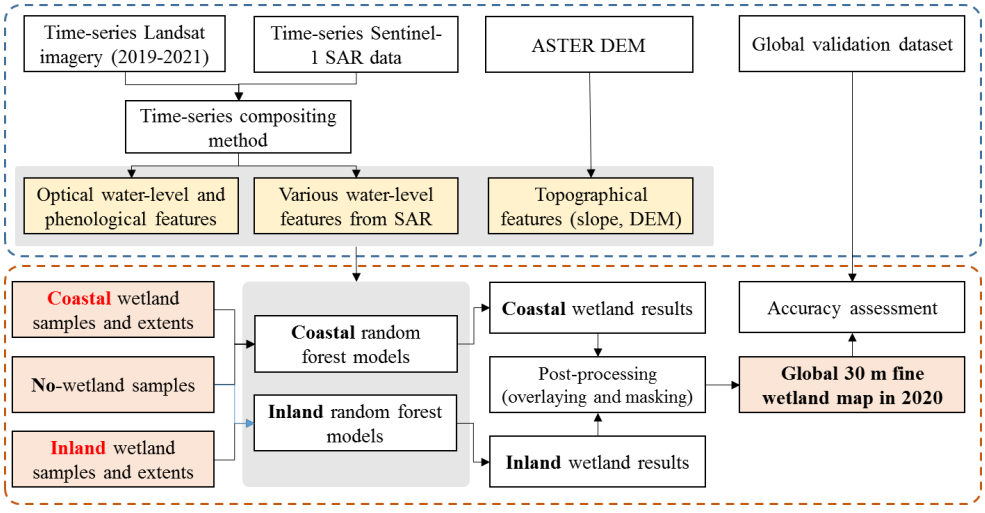


Figure 4. The flowchart of wetland mapping using water-level and phenological features and a stratified classification strategy.

4.1 Generating the water-level and phenological features

The spectral characteristics of the wetlands would quickly change along with the seasonal or daily water levels of the underlying surface. For example, the tidal flat was the status of seawater at the high tidal stage and mud or sand flats at low tidal stages (Wang et al., 2021); therefore, it was necessary to extract the high- and low-water-level features to completely capture these water-level sensitive wetlands. Over the past several years, the time-series compositing strategy has been widely used to capture phenological and cloud-free composites (Jia et al., 2020; Ludwig et al., 2019; Murray et al., 2019; Zhang et al., 2021a). For example, Murray et al. (2019) used the quantile compositing method to extract different tidal stage information, and successfully produced the global distribution of tidal flats. However, Zhang et al. (2022b) explained that the percentile composting method fails to capture the highest- and lowest-water stages and further proposed to use the maximum normalized index for compositing the highest- and lowest-water features. Meanwhile, a multi-temporal phenology was also essential for classifying the vegetated wetlands and excluding these non-wetland land-cover types (Li et al., 2020; Ludwig et al., 2019). Before generating various water-level and phenological features, four spectral indexes including normalized difference water index (NDWI), land surface water index (LSWI), normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) were imported because many studies have demonstrated that they were of great help in wetland mapping (Mao et al., 2020; Wang et al., 2020),

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

$$LSWI = \frac{\rho_{nir}-\rho_{swir1}}{\rho_{nir}+\rho_{swir1}}, NDWI = \frac{\rho_{green}-\rho_{swir1}}{\rho_{green}+\rho_{swir1}}, NDVI = \frac{\rho_{nir}-\rho_{red}}{\rho_{nir}+\rho_{red}}, \ EVI = 2.5 \times \frac{\rho_{nir}-\rho_{red}}{\rho_{nir}+6\times\rho_{red}-7.5\times\rho_{blue}+1} \quad (4)$$

where ρ_{blue}, ρ_{green}, ρ_{red}, ρ_{nir}, ρ_{swir1} were the blue, green, red, near-infrared and shortwave infrared bands of Landsat imagery, respectively.

Regarding the highest and lowest water-level features, considering that NDWI was sensitive to open surface water and that Zhang et al. (2022b) found a positive relationship between tidal height and NDWI using field survey data, the maximum NDWI compositing was applied to the time-series clear-sky Landsat imagery to capture the optical highest water-level composites illustrated in Figure 5b. As for the lowest water-level features, considering that the tidal/flooded flat or marsh usually reflected higher NDVI and EVI values than water bodies and that Zhang et al. (2022b) also used the field data to demonstrate that there was a negative relationship between tidal-level height and NDVI, the maximum NDVI composite was applied to capture the optical lowest water-level information illustrated in Figure 5a. Considering that optical observations were usually contaminated by clouds, especially during the rainy seasons, and that the SAR back coefficients had a great advantage in the presence of cloud coverage and were found to be sensitive to the soil moisture, vegetation structure, and inundation information, the time-series Sentinel-1 SAR imagery could be used as a complementary dataset for capturing the highest and lowest water-level features (DeVries et al., 2020; Li et al., 2020; Mahdianpari et al., 2018). Specifically, as the SAR active transmitting signals were heavily absorbed when they reached the water body, the corresponding SAR back coefficients in the water body had lower values compared to other land-cover types. To capture the high water-level features from the time-series Sentinel-1 imagery, the percentile compositing method using the 5th percentile was applied, as illustrated in Figure 5d. Conversely, the 95th percentile of Sentinel-1 VV and VH were generated to capture the lowest water-level information (Figure 5c). It should be noted that the minimum and maximum percentiles were not used because the time-series Sentinel-1 imagery still contained the residual errors caused by the quantitative processing.

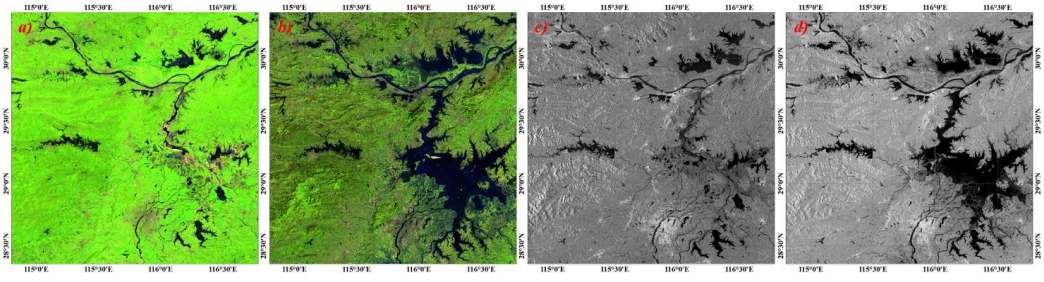


Figure 5. The lowest and highest water-level features derived from (a-b) time-series Landsat optical reflectance data and (c-d) the Sentinel-1 SAR imagery using the time-series compositing method in Poyang Lake, China.

There were usually two options for capturing phenological features from time-series Landsat imagery. These included seasonal-based compositing (Zhang et al., 2021a; Zhang et al., 2022a) and percentile-based compositing (Hansen et al., 2014; Zhang and Roy, 2017; Zhang et al., 2021b). The former used the phenological calendar for selecting time-matched imagery. It then adopted the compositing rule to capture the seasonal features, while the latter directly used the statistical distributions to select various percentiles. Azzari and Lobell (2017) quantitatively analyzed the performance of two compositing methods and found that both of them had similar mapping accuracy for land-cover mapping. Meanwhile, the seasonal-based compositing method needed the prior phenological calendar, while the percentile compositing method did not require any prior knowledge or explicit assumptions regarding the timing of the season; therefore, the percentile compositing method was

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.


more suitable to generate phenological features. This study composited time-series Landsat reflectance bands and four spectral indexes into five percentiles (15th, 30th, 50th, 70th and 85th). It should be noted that the minimum and maximum percentiles were excluded because they were usually affected by residual clouds, shadows, and saturated observations.

Lastly, the topographical variables were also important factors for determining the spatial distribution of wetlands (Ludwig et al., 2019; Tootchi et al., 2019). For example, the widely used topographical wetness index (TWI) uses the local slope to reveal soil wetness, which improves wetland classification performance and reduces commission errors within upland areas (Ludwig et al., 2019). Therefore, the elevation, aspect, and slope, calculated from the ASTER GDEM dataset, were included in the multi-sourced features.

4.2 The stratified classification strategy for wetland mapping

Since we have simultaneously extracted the maximum coastal and inland wetland extents when deriving training samples from prior wetland datasets, the stratified classification strategy was adopted to fully use the maximum extent constraint. Namely, if a pixel was classified as a coastal wetland outside the maximum coastal wetland extents, it would be identified as a misclassification. Furthermore, there were two approaches, the large-area land-cover mapping, which included global classification modeling (using one universal model for the whole areas) and local adaptive modeling (using various models for different zones) (Zhang et al., 2020). For example, Zhang and Roy (2017) demonstrated that local adaptive modeling outperformed the global classification modeling strategy. Therefore, the global land surface was first divided into 961 5° × 5° geographical tiles, which were inherited from the global 30 m land-cover mapping by (Zhang et al., 2021b), and then independently trained the local adaption models using training samples from adjacent 3×3 tiles for ensuring the classification consistency across neighboring geographical tiles.

Afterward, as the random forest (RF) classifier was demonstrated to have obvious advantages in dealing with high-dimensional data, robustness for training noise and feature selection, as well as achieving higher classification when compared to other widely used machine learning classifiers (e.g., support vector machines, neural networks, decision trees, etc.) (Belgiu and Drăguţ, 2016; Gislason et al., 2006) , the RF classifier was selected for mapping inland and coastal wetlands using multi-sourced features. It should be noted that the RF classifier had two key parameters: the number of selected prediction variables (Mtry) and the number of decision trees (Ntree). Belgiu and Drăguţ (2016) and Zhang et al. (2022b) have demonstrated the quantitative relationship of Ntree against classification accuracy and found that the classification accuracy stabilized when Ntree was greater than 100. Meanwhile, Belgiu and Drăguţ (2016) suggested that the Mtry should take its default value of the square root of the number of all input features. Therefore, the Ntree and Mtry took 100 and the square root of the number of all input features, respectively.

The inland and coastal wetland maps were produced by combining water-label and phenological features, the stratified classification strategy, local adaptive modeling, and the derived wetland and non-wetland training samples. As the inland and coastal wetlands were independently produced, some pixels in the overlapping area of maximum inland and coastal wetland extents were simultaneously labeled as inland wetlands and coastal wetlands. However, as the final global wetland map was a hard classification, these pixels should be post-processed into one label. As the random forest classifier could provide the posterior probability for each pixel, we determined the labels of the confused pixels by comparing the posterior probabilities.

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.


4.3 Accuracy assessment

To quantitatively analyze the accuracy of the proposed method and corresponding GWL_FCS30 wetland maps, we employed four metrics typically used to evaluate accuracy, which include the kappa coefficient, overall accuracy, user's accuracy (measuring the commission error), and producer's accuracy (measuring the omission error) (Gómez et al., 2016; Olofsson et al., 2014), were calculated using 18,701 global wetland validation samples in Section 2.4. Further, to intuitively understand the performance of the produced map, four existing global wetland products (GlobeLand30 wetland layer (Chen et al., 2015), GLC_FCS30-2020 wetland layer (Zhang et al., 2021b), CCI LC wetland layer (Defourny et al., 2018) and GLWD dataset) were collected to analyze the over-estimation and under-estimation problems in the inland regions, and three widely used mangrove forest datasets (Atlas mangrove, Global Mangrove Watch, and USGS Mangrove, 1isted in Table 1) were imported to assess the performance of the developed GWL_FCS30 wetland map in coastal areas.

5. Results

5.1 The spatial pattern of global wetlands in 2020

Figure 6 illustrates the spatial distributions of our GWL_FCS30 wetland map and their area statistics in latitudinal and longitudinal directions in 2020. Overall, the GWL_FCS30 map accurately captured the spatial patterns of wetlands. It mainly concentrated on the high latitude areas in North Hemisphere and the rainforest areas (Congo Basin and Amazon rainforest in South America). Quantitatively, according to the latitudinal statistics, approximately 62.3% of wetlands were distributed poleward of 40°N (a large number of wetlands are located in Canada and Russia), and 19.9% of wetlands were located in equatorial areas, between 10°S~10°N, within which the Congo and Amazon rainforest wetlands are located. As for the longitudinal direction, there were mainly four statistical peak intervals: 100°W~80°W (Canada wetlands), 75°W~50°W (Amazon wetlands), 15°E~25°E (Congo wetlands), and 60°E~90°E (Russia wetlands). Afterward, to more intuitively understand the performance of our GWL_FCS30 wetland map, four local enlargements in Florida, the Congo Basin, Sundarbans, and Poyang Lake were also illustrated. All of them comprehensively captured the wetland patterns in these local areas. For example, there was significant consistency between our results and Hansen's regional wetland maps in the Congo Basin (Bwangoy et al., 2010) ; both results indicated that the wetlands occurred closer to major rivers and floodplains. Next, according to the lowest and highest water-level features derived from Sentinel-1 SAR and Landsat optical imagery in Figure 5, the inland wetlands having various water levels were also comprehensively identified in the Poyang wetland map (Figure 6d). Figure 6c illustrates the spatial distributions of the world's largest mangrove forest in the Sundarbans (Figure 6c), and the cross-comparison in Figure 11 also demonstrates the great performance of the GWL_FCS30 dataset. Lastly, the Florida wetlands simultaneously contained five sub-categories (mangrove, tidal flat, salt marsh, marsh, and swamp). These were distributed along the coastlines and rivers and are accurately captured in Figure 6a.

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

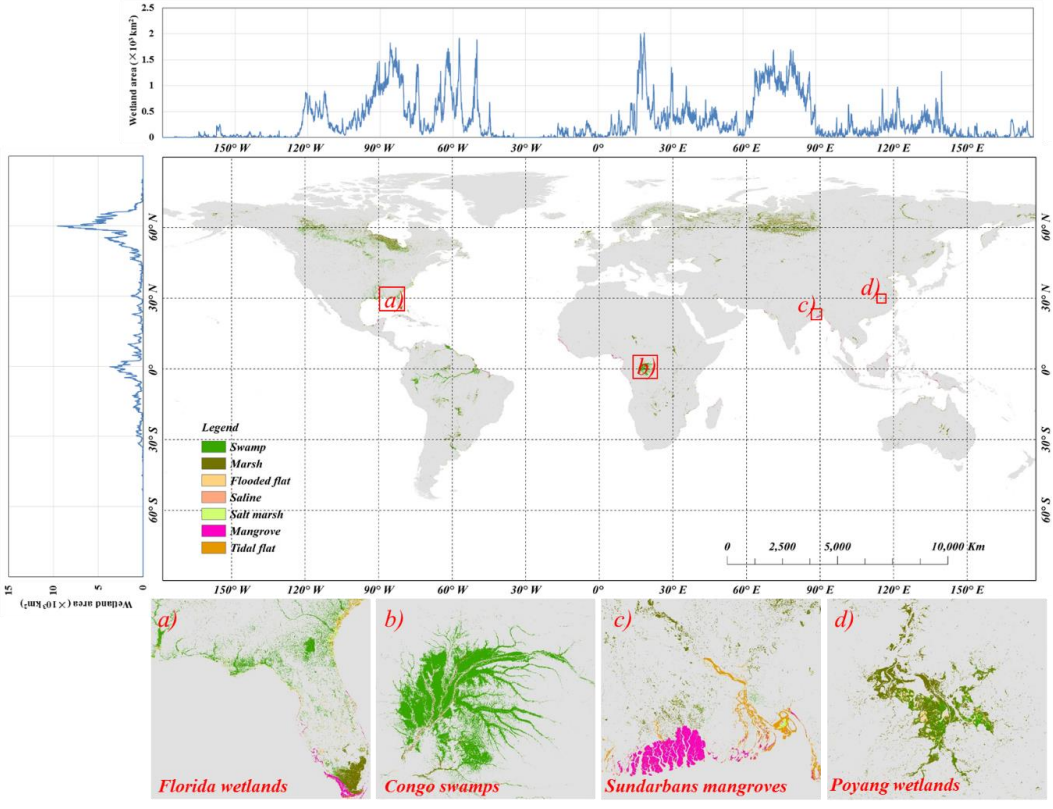


Figure 6. The overview of global 30-m fine wetland maps and their area statistics in latitudinal and longitudinal directions in 2020. Four local enlargements in (a) Florida, (b) Congo Basin, (c) Sundarbans, and (d) Poyang Lake were also illustrated.

Figure 7 illustrates the spatial distribution of seven sub-category wetlands after aggregating to the 0.5° × 0.5° grid cell. Intuitively, swamp and marsh accounted for most inland wetlands. In contrast, flooded tidal wetlands had obviously lower proportions, and the inland saline type was only distributed along the surroundings of several saline lakes. In terms of the spatial distribution, it can be found that: 1) the swamp wetlands mainly were concentrated in the Congo and Amazon rainforests, Southern United States, and Northern Canada; 2) most marsh wetlands were located in high latitude areas in the Northern Hemisphere including Northern Canada, Russia, and Sweden; 3) there were significant coexistent relationships between flooded flat, swamp, and marsh wetlands, and flooded flat wetlands were sparse land-cover types compared to the other two wetlands. Similar to coastal wetlands, the mangrove forests were only found in coastal areas below 30°N and were mainly concentrated in regions between 30°N ~ 30°S, including Southeast Asia, West Africa, and the east coast of South America. The salt marshes and tidal flats shared similar spatial distributions. They were widely distributed globally and can be observed along most coastlines. In addition, the tidal flat distributions were closely related to the slope of coastlines, tidal ranges, and sediment inflows. For example, the tidal flats in Asia and Europe usually were located in the tide-dominated estuaries and deltas. Similarly, Murray et al. (2019) also demonstrated that there were often more tidal flats where the river flowed into the sea.

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

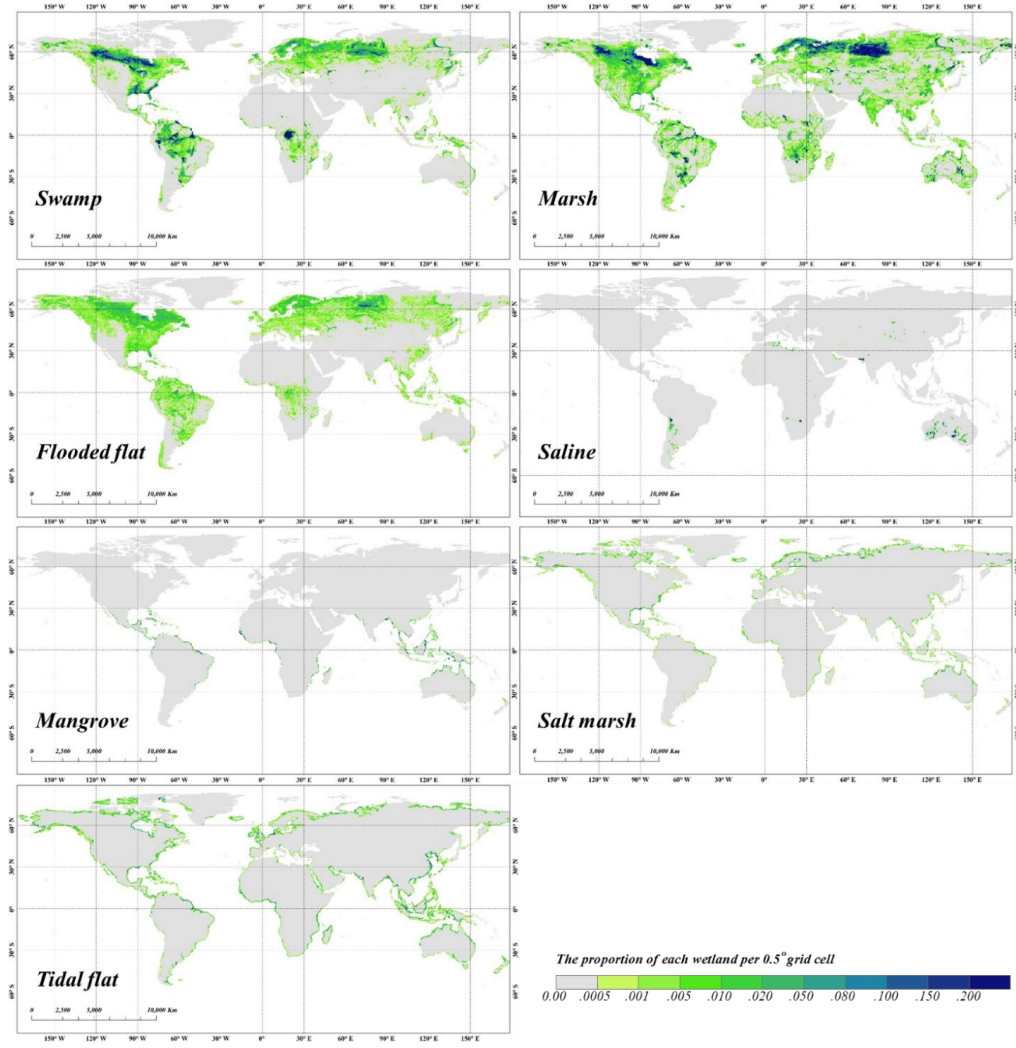


Figure 7. The spatial distributions of the seven wetland sub-categories after aggregating them to a resolution of 0.5° × 0.5°.

To quantitatively summarize the distribution of the seven sub-category wetlands, the total area and area percentages of seven fine wetlands over each continent were calculated in Figure 8 and Table 3. The total wetland area was 3.57 million km^2, including 3.10 million km^2 of inland wetlands and 0.47 million km^2 of coastal wetlands, and the distribution of wetlands varied across different continents. Intuitively, approximately 60% of coastal wetlands (tidal flat, salt marsh, and mangrove) and 70% of flooded flat and marsh wetlands were distributed in the Northern Hemisphere, especially in the Asian and North American continents. More than 85% of saline wetlands were located in the Southern Hemisphere, especially the Oceania continent. Then, in terms of specific wetland sub-categories, the swamp was mainly distributed on the North American, African, and South American continents, which contained many rainforest wetlands, with corresponding swamp areas of 0.35, 0.18, and 0.32 million km^2, respectively. Swamp areas in the Oceania continent were the smallest, covering

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

only 6599 km^2, mainly because the forest cover in Oceania was smaller than in other continents. The marsh and flooded flats shared similar areal proportions in all six continents and were mainly concentrated in the North Hemisphere (exceeding 70%), where many lakes and rivers were distributed. Next, the mangrove forests only covered regions south of 30°N and were mostly concentrated in tropical regions near the equator, such as Southeast Asia, East Africa, and Central America, so this sub-category was absent in the Europe continent.

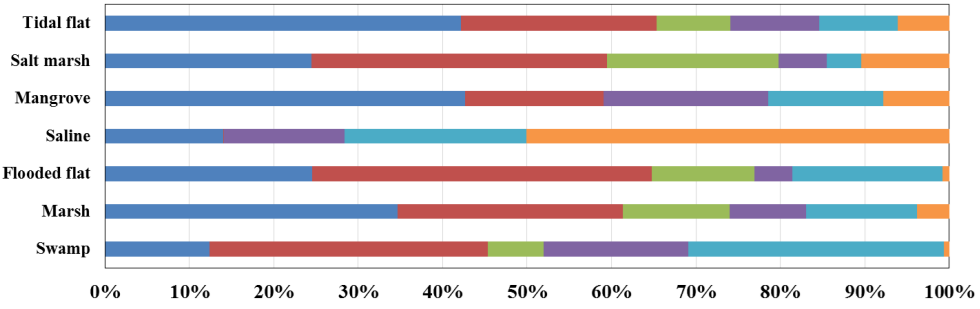


Figure 8. The area proportions of seven wetland sub-categories over each continent.

Table 3. The total wetland area (unit: km^2) of seven wetland sub-categories at six continents and globe.

	Swamp	*Marsh*	*Flooded flat*	*Saline*	*Mangrove*	*Salt marsh*	*Tidal flat*
Asia	132229.8	570929.4	72801.2	12252.1	68220.6	25356.3	88928.7
North America	352780.8	438726.7	119137.9	82.4	26242.9	36262.2	48948.9
Europe	70084.2	208957.2	36138.9	48.2	0.0	21035.5	18347.1
Africa	183727.1	149138.1	13188.7	12492.8	31130.7	5952.4	22278.9
South America	323164.9	216470.4	52643.9	18881.4	21809.5	4218.4	19622.1
Oceania	6599.3	62513.7	2403.9	43904.2	12463.9	10759.4	12826.6
Total	1068586.1	1646735.4	296314.5	87661.1	159867.6	103584.1	210952.4

5.2 Accuracy assessment of global 30 m fine wetland map

Using 18,701 global validation samples, the confusion matrix of the novel GMW_FCS30 wetland map was calculated in Table 4. Overall, our wetland map achieved an overall accuracy of 87.7% and a kappa coefficient of 0.810 across the fine wetland classification system. In terms of the producer and user accuracies, the non-wetlands achieved the highest performance with a producer's accuracy of 93.1% and a user's accuracy of 96.1%, mainly because we combined multi-sourced pre-existing wetland datasets to determine the maximum wetland boundary and further used multi-sourced and time-series imagery to distinguish between wetlands and non-wetlands. As for the coastal wetlands, mangrove forest and tidal flat achieved higher accuracies than other sub-categories, with producer's accuracies of 84.3% and 83.1% and user's accuracies of 94.0% and 88.8%, respectively. The misclassification of mangrove mainly focused on the confusion between mangrove, swamp, and salt marsh because they all shared similar vegetation spectral characteristics. The tidal flat also suffered from confusion with the salt marsh, flooded flat, and the non-wetlands, especially for the water bodies, because this land-surface type reflected complicated temporal dynamics at various water levels. The salt marsh had a lower producer accuracy of 75.4% than mangrove and tidal flat because its reflectance spectra were affected by both water levels and vegetation cover with considerable spatiotemporal heterogeneity. Furthermore, in terms

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

of the four inland sub-categories, the swamp and marsh obviously performed better than the flooded flat and saline wetland, with producer accuracies of 82.6% and 85.9%, respectively. It can be seen that the confusion between swamp and marsh was the main source of the misclassification error of swamp and that the marsh was simultaneously confused with non-wetland, swamp, and flooded flat because the spectra of marsh changed along with the water levels. For example, the marsh in Poyang Lake, shown in Figure 5b, was flooded at its highest water levels. Then, the flooded flat achieved a low producer accuracy of 51.1% because it usually coexisted with the marsh and shared similar spectral characteristics, so approximately 24.3% of flooded flat points were labeled as the marsh in our wetland map. The saline wetland was mainly concentrated along the edge of salt lakes and demonstrated great performance in our mapping, with producer's and user's accuracies of 74.7% and 92.5%, respectively.

Table 4. The confusion matrix of the global 30 m fine wetland map using 18701 validation points.

	NWT	**SWP**	**MSH**	**FFT**	**SAL**	**SMH**	**MGV**	**TFT**	**Total**	**P.A.**
NWT	9627	246	313	65	0	44	11	40	10346	93.1
SWP	24	1950	306	73	3	1	5	0	2362	82.6
MSH	122	172	2856	172	1	0	1	0	3324	85.9
FFT	89	50	149	312	1	6	1	3	611	51.1
SAL	30	13	37	2	271	3	0	7	363	74.7
SMH	61	4	3	23	0	356	13	12	472	75.4
MGV	39	32	1	1	0	16	498	4	591	84.3
TFT	28	22	11	11	17	17	1	525	632	83.1
Total	10020	2489	3676	659	293	443	530	591	18701	
U.A.	96.1	78.3	77.7	47.3	92.5	80.4	94.0	88.8		
O.A.	87.7									
Kappa	0.810									

Note: **NWT**: non-wetlands, **SWP**: swamp, **MSH**: marsh, **FFT**: flooded flat, **SAL**: saline, **SMH**: salt marsh, **MGV**: mangrove forest, **TFT**: tidal flat, **O.A.**: overall accuracy, **P.A.**: producer's accuracy, **U.A.**: user's accuracy.

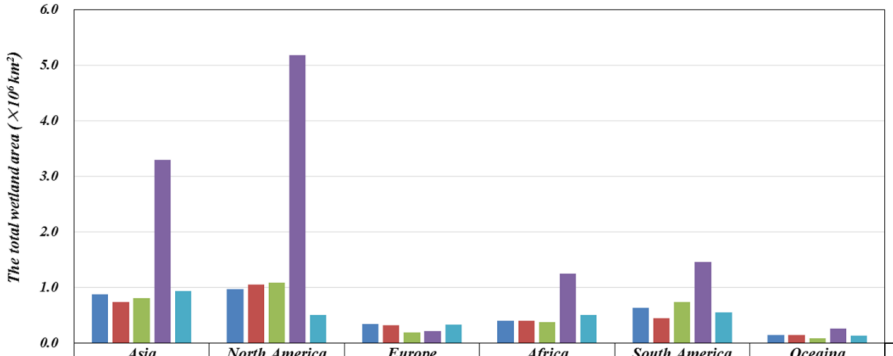
5.3 Cross-comparisons with other global wetland maps

To comprehensively understand the performance of the GWL_FCS30 wetland maps, four existing global wetland datasets (GLC_FCS30, GlobeLand30, CCI_LC, and GLWD), listed in Table 1, were selected. Figure 9 quantitatively illustrates the total wetland area of five products over each continent. Notably, the total wetland area of four existing wetland datasets estimated from our study differed from previous studies (Hu et al., 2017a; Tootchi et al., 2019) because we excluded the water bodies when calculating the total wetland area. Specifically, the estimated total wetland area in this study was more reasonable because permanent water bodies with depths of more than six meters were not considered wetlands, according to the RAMSAR Convention (Gardner and Davidson, 2011).

Specifically, the total wetland area of different wetland products varied. The GLWD obviously overestimated the wetland area on each continent mainly because it was derived from the compilation model instead of actual remote sensing observations (Lehner and Döll, 2004). Namely, the GLWD classified a large amount of non-wetlands as potential wetlands. The remaining four wetland products, derived from the Landsat and PROBE-V remote sensing imagery, shared a similar total wetland area of approximately 3.0 million km^2,

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

and our GWL_FCS30 wetland dataset had the largest total area of 3.574 million km^2 among these datasets. The CCI LC wetland layer contained the smallest wetland area of 2.955 million km^{2,} and the estimated area in North America was profoundly lower than the other datasets, mainly because the CCI LC heavily underestimated the wetland distribution in Canada after a comparison with the Canadian Wetland Inventory (Amani et al., 2019). Next, the total wetland area in GlobeLand30 and GLC_FCS30 wetland layer was lower than the GWL_FCS30 wetland dataset because some water-level sensitive wetlands cannot be comprehensively captured in these two datasets (Figure 10).



	Asia	North America	Europe	Africa	South America	Oceaina	Total
GWM_FCS30	0.881	0.972	0.338	0.399	0.633	0.138	3.574
GLC_FCS30	0.739	1.054	0.316	0.400	0.452	0.149	3.110
GlobeLand30	0.803	1.090	0.196	0.382	0.734	0.083	3.287
GLWD	3.294	5.183	0.213	1.253	1.453	0.264	11.660
CCI_LC	0.931	0.505	0.332	0.508	0.550	0.129	2.955

Figure 9. The total wetland area (unit: million km^2) of five global wetland products on six continents.

Figure 10 illustrates the performances of five wetland products for two typical wetland regions (Poyang Lake in China and Pantanal wetland in Brazil). The reasons for choosing these two regions were that the wetlands in Poyang Lake quickly changed with water levels, and the Pantanal wetland was the largest wetland in the world. Intuitively, the GWL_FCS30 wetland maps had the greatest performance in capturing the spatial patterns of various wetland sub-categories. Comparatively, the GLC_FCS30 wetland layer seriously underestimated the wetland area in both regions, which obviously overestimated many water-sensitive wetlands as water bodies in Poyang Lake and also missed a large number of wetlands in the Pantanal wetland. Zhang et al. (2021b) also stated that the wetland in the GLC_FCS30 suffered from low accuracy because of a lack of enough wetland samples and multi-sourced wetland sensitive features. Then, the GlobeLand30 wetland layer performed better in the Pantanal wetland than in Poyang Lake, which obviously misclassified many water-sensitive wetlands as water bodies in the Poyang Lake mainly because the low water-level features were not captured during the development of the GlobeLand30 (Chen et al., 2015). In addition, the wetland layer of GlobeLand30 in Pantanal still suffered from the over-estimation problem, and some non-wetlands in Pantanal Wetland Park were mislabeled as wetland, so the wetland layer in the GlobeLand30 only achieved a user's accuracy of 74.87% (Chen et al., 2015). The CCI LC was highly consistent with the GWL_FCS30 wetland maps in spatial distribution. Details show that the wetlands in the CCI LC were still underestimated in the Poyang Lake wetland and overestimated in the Pantanal wetland based on the highest and lowest water-level composites. Lastly, the GLWD dataset significantly overestimated the wetlands in two regions, namely, the mapped marsh

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

area was obviously greater than its actual area and it also misclassified these water-sensitive wetlands as water bodies near Poyang Lake.

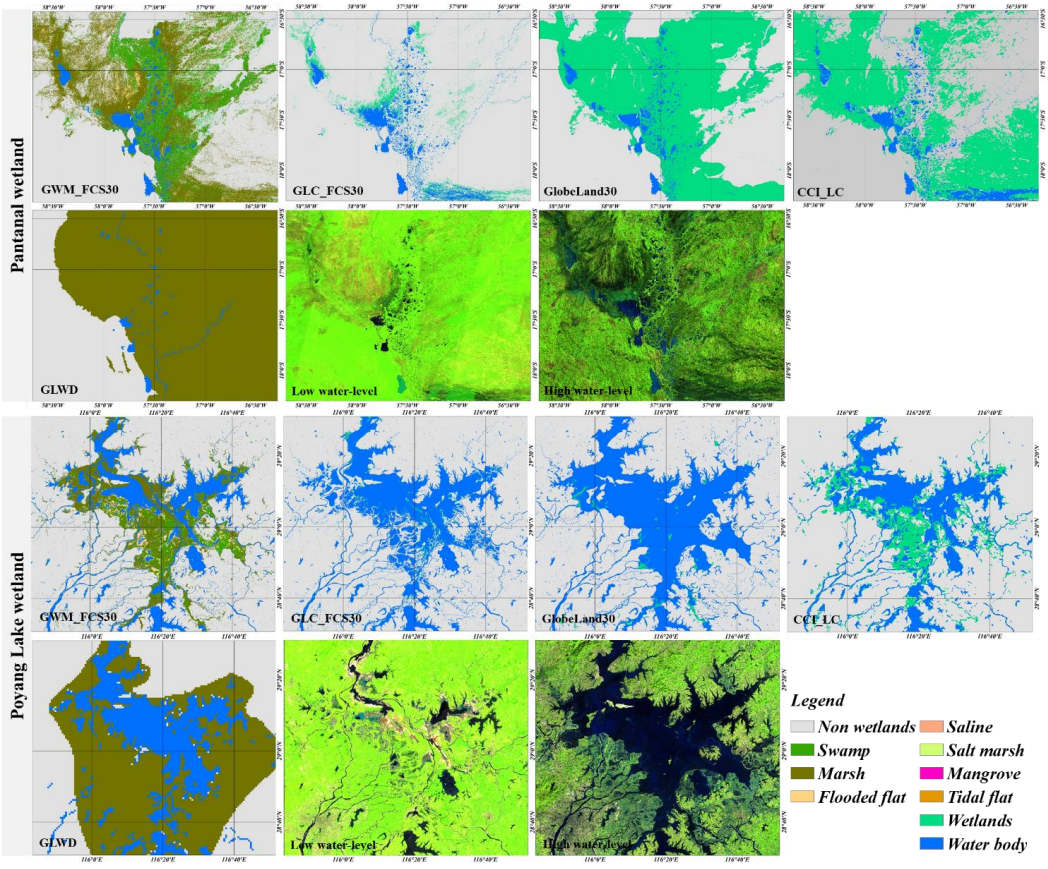
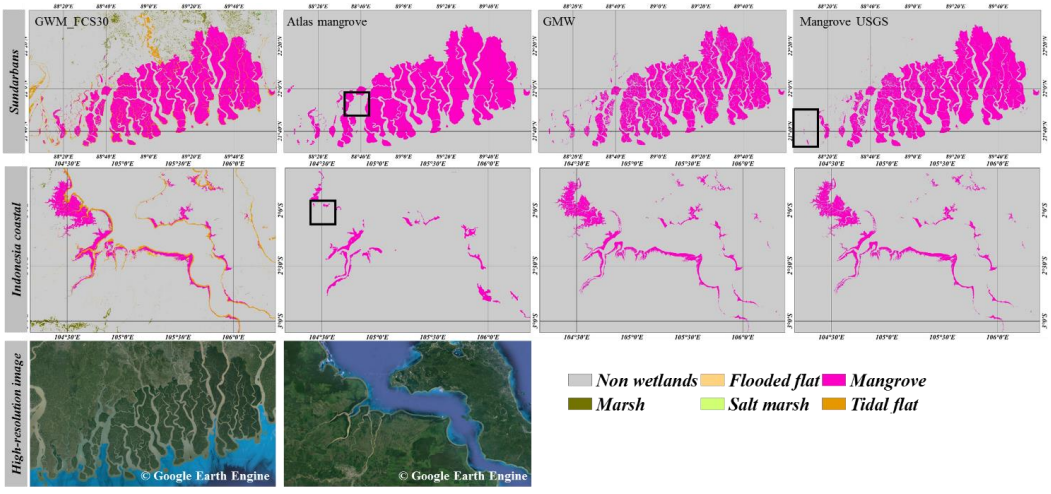


Figure 10. The cross-comparisons between our GWL_FCS30 wetland maps with four existing wetland products: GLC_FCS30 generated by Zhang et al. (2021b), GlobeLand30 generated by Chen et al. (2015), CCI LC generated by Defourny et al. (2018) and GLWD generated by Lehner and Döll (2004) at Pantanal and Poyang Lake wetland. The false-color composited Landsat imagery (SWIR1, NIR, and Red bands) at the highest and lowest water levels were also illustrated.

Figure 11 illustrates the comparisons between our fine wetland maps with three widely used global mangrove forest products (Atlas mangrove, GMW (Global Mangrove Watch), and USGS Mangrove) listed in Table 1 in two typical mangrove regions (coastal Indonesia and Sundarbans). Intuitively, there was great consistency over four mangrove datasets because the mangrove forest reflected obvious and strong vegetation reflectance characteristics and was easier to identify than other wetland sub-categories. Specifically, the Atlas mangrove dataset suffers from the underestimation problem; namely, the mangrove area in the Atlas mangrove dataset was obviously lower than the other three products, especially in coastal Indonesia (black rectangles). The USGS mangrove product can comprehensively capture the spatial distribution of mangroves over two regions. Still, it missed isolated fragments of mangrove forests in the Sundarbans (black rectangle) based on

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

590 high-resolution imagery. The GMW dataset was validated to achieve an overall accuracy of 95.25%, with user and producer accuracies of mangrove forests of 97.5% and 94.0%, respectively (Bunting et al., 2018; Thomas et al., 2017), which shows great agreement with our fine wetland maps and confirms that this dataset accurately identified the spatial patterns of mangrove forest in both regions.



595 **Figure 11**. The cross-comparisons between our GWL_FCS30 wetland maps with three mangrove products (Atlas mangrove developed by Spalding (2010), GMW (Global mangrove watch) developed by Thomas et al. (2017) and Mangrove USGS developed by Giri et al. (2011)) in Sundarbans and coastal Indonesia. The high-resolution imagery came from the © Google Earth Engine platform (https://earthengine.google.com; last access: 16 May 2022).

600 Figure 12 illustrates the scatterplot between the tidal flat of the GWL_FCS30 map in 2020 against Murray's tidal flat in 2016 (Murray et al., 2019) after aggregating to a spatial resolution of 0.01°. Overall, two the tidal flat maps shows significant agreement and achieves a determination coefficient of 0.717 and RMSE of 0.151. However, in terms of the spatial distribution of the scatterplot, the tidal flat in Murray's dataset was significantly lower than that of the GWL_FCS30 mainly because the highest and lowest tidal levels cannot be
605 comprehensively captured in Murray's tidal flat dataset (Murray et al., 2019), as these tidal flats over deeper water were easier to miss in Murray's products. Similarly, Zhang et al. (2022b) also demonstrated that Murray's tidal flat dataset suffered from an underestimation problem. To intuitively understand the performance of the two tidal flat datasets, two local regions and their corresponding highest and lowest tidal-level composites are illustrated in Figure 13. It is evident that the actual tidal flat in Murray's dataset was clearly underestimated
610 (black rectangles), especially in the first region.

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

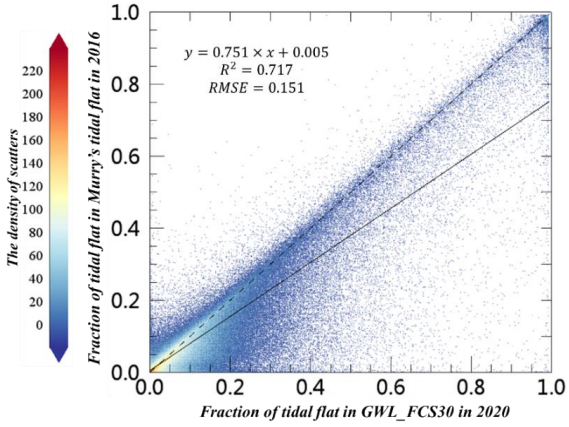


Figure 12. The scatterplot between the tidal-flat fraction in GWL_FCS30 in 2020 (x-axis) against that of Murray's tidal-flat fraction in 2016 (y-axis) (Murray et al., 2019) after aggregating to a spatial resolution of 0.01°.

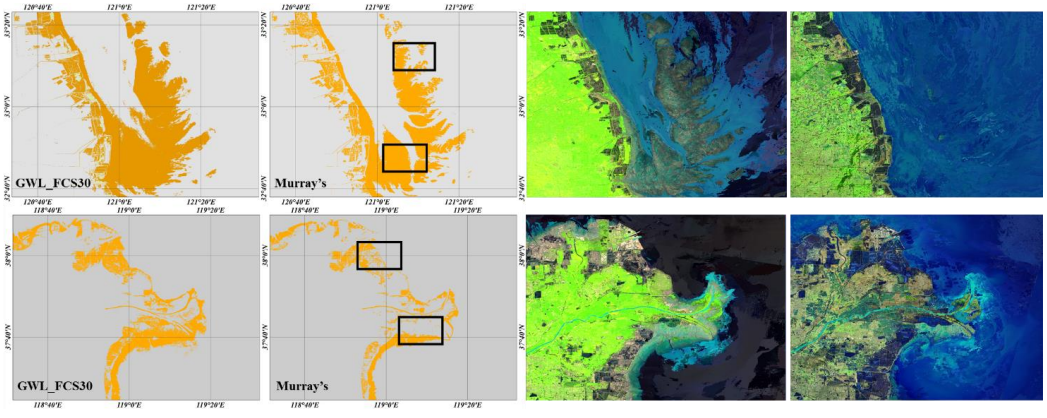


Figure 13. The comparisons between the tidal flat of GWL_FCS30 in 2020 and Murray's tidal flat in 2016 for two local regions. In each case, the highest and lowest tidal-level composites, composited by SWIR1, NIR, and red bands, are illustrated.

6. Discussion

6.1 The feasibility of derived training samples for wetland mapping

Previous studies found that the confidence of training samples directly affected the final classification accuracy (Mellor et al., 2015; Radoux et al., 2014). However, collecting global training samples via visual interpretation was highly time-consuming and involved a large amount of manual work, so it was impossible to use the visual interpretation for collecting global wetland samples. This study proposed combining multi-sourced pre-existing wetland products, refinement rules, and expert knowledge to automatically derive these massive inland and coastal wetland training samples globally. To demonstrate the reliability of the derived training samples for wetland mapping, we randomly selected approximately 10,000 points from the sample pool and checked their confidence using visual interpretation. It should be noted that we cannot check all the training

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

samples because the number of derived samples was massive (exceeding 20 million training samples in Section 3). After a point-to-point inspection, these selected training samples achieved an overall accuracy of 91.53% in 2020. Meanwhile, we also used 10,000 selected wetland training samples and many non-wetland samples to analyze overall and producer's accuracies of coastal and inland wetlands versus number of erroneous training samples. Specifically, we gradually increased the "contaminated" samples by randomly altering the label of a certain percentage of training samples in steps of 0.01, and then used these "contaminated" samples to build the RF classification model. After repeating the process 100 times, the quantitative relationship between mapping accuracies and erroneous samples is illustrated in Fig. 14. Obviously, the overall accuracy and producer's accuracy of wetlands (merging seven sub-categories into one wetland) was insensitive to the erroneous training samples when the percentage of erroneous samples was controlled within 20%. Beyond this, the accuracies slowly decreased along with the increase of erroneous training samples. Similarly, previous studies by Zhang et al. (2021b) and Zhang et al. (2022a) quantitatively analyzed the relationship between overall accuracy and the erroneous training samples size. They found that the overall accuracy stabilized when the percentage of erroneous training samples was controlled within the threshold and then rapidly decreased after exceeding the threshold. Gong et al. (2019) also demonstrated the random forest classification model was resistant to the erroneous training samples when the percentage of erroneous training samples remained below 20%. Therefore, the derived training samples in Section 3 were accurate enough to support large-area fine wetland mapping.

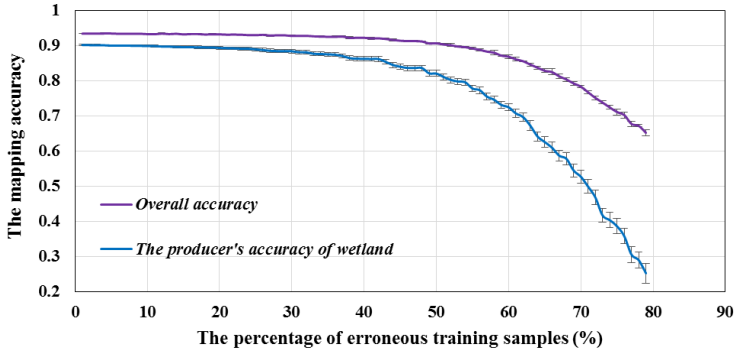


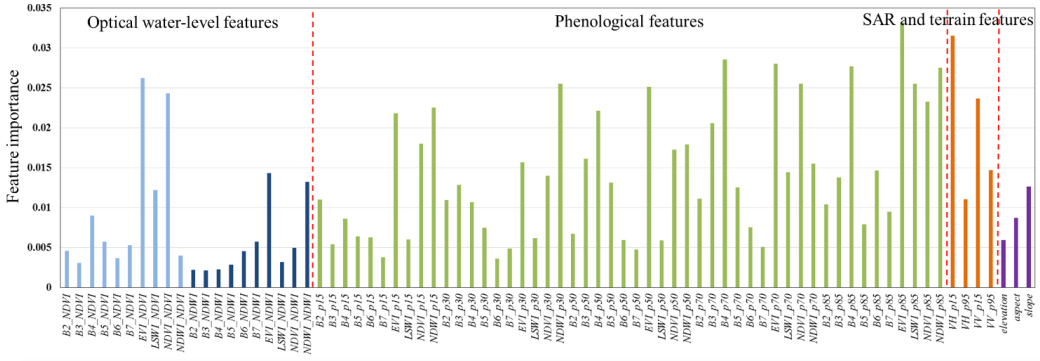
Figure 14. The relationship between mapping accuracies with the percentage of erroneous training samples with a step of 1%.

6.2 The importance of multi-sourced phenological features for wetland mapping

The complicated temporal dynamics and spectral heterogeneity caused great uncertainties in wetland mapping because their spectral characteristics quickly changed with the seasonal or daily water levels of the underlying surface (Ludwig et al., 2019). Single-date optical or SAR observations often failed to capture the spatiotemporal variability of wetlands, which led to the commission and omission errors in wetland mapping, so many studies have demonstrated that using multi-temporal data was an effective way to achieve high-precision wetland mapping, especially for the water-sensitive sub-categories (tidal flat and marsh) (Jia et al., 2020; Zhang et al., 2022b). This study combined the time-series Landsat reflectance and Sentinel-1 SAR products to capture the various water levels and phenological features for comprehensively depicting their temporal dynamics and spectral characteristics, as discussed in Section 4.1. To quantitatively analyze the importance of these multi-sourced and multi-temporal features, we used the random forest classification model,

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

660 which calculated the increased mean squared error by permuting the out-of-bag data of a variable while keeping remaining variables constant (Breiman, 2001; Zhang et al., 2020), in an effort to compute their importance. Figure 15 illustrates the importance of all multi-sourced and phenological features, and it can be found that the phenological features which made the most significant contribution mainly did so because they used the multi-temporal percentiles to comprehensively capture vegetation phenology (EVI and NDVI) and water-level
665 dynamics (NDWI and LSWI) for the various land-cover types. Then, the combination of optical and Sentinel-1 SAR water-level features ranked as the second-most important role in distinguishing the fine wetlands and non-wetlands. Based on the low and high water-level features in Fig. 5, the highest and lowest water-level features greatly contributed to determining these water-sensitive wetlands (marsh, tidal flat, and flooded flat). For example, Zhang et al. (2022b) quantitatively analyzed the contribution of multi-sourced features to mapping
670 accuracy. They found that importing water-level features significantly improved the ability to separate tidal flats from non-wetlands. Lastly, three topographical variables also contributed to wetland mapping because the spatial distribution of wetlands had a significant relationship with topography and was mainly distributed in low-lying areas (Zhu and Gong, 2014).



675 **Figure 15**. The importance of multi-sourced and multi-temporal features derived from the random forest classification model.

6.3 The limitations and prospects of our global fine wetland map

Using pre-existing global wetland products, multi-sourced and time-series remote sensing imagery, stratified classification strategy, and local adaptive classification methods, the first global 30-m fine wetland
680 maps were produced with an overall accuracy of 85.5% and a kappa coefficient of 0.776. Meanwhile, the training sample reliability analysis and multi-sourced feature importance evaluation also demonstrated that the proposed method was suitable for large-area fine wetland mapping. However, it should be noted there were still many uncertainties and limitations to the proposed method and global wetland maps. First, the proposed method used continuous Landsat reflectance and Sentinel-1 SAR imagery to capture various water-level information.
685 Still, it might fail when the available Landsat observations were sparse and lacked the aid of Sentinel-1 SAR data, especially before 2000. Thus, our future work would focus on combining a richer multi-sourced data source, including MODIS, Sentinel-2, SPOT, and PALSAR imagery, to develop a more robust wetland mapping method. For example, Chen et al. (2018) integrated Landsat and MODIS observations to successfully monitor the wetland dynamics from 2000 to 2014 using a spatiotemporal adaptive fusion model.

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

We then combined the pre-existing global wetland products to derive the training samples; however, the salt marsh and saline samples still used the visual interpretation method to ensure their reliability because of lacking sufficient pre-existing global products. Additionally, it was found that the producer accuracy of salt marsh and saline in Table 4 was relatively poor compared with other sub-categories mainly because visual interpretation cannot provide massive and geographically distributed salt marsh and saline training samples. Namely, this study cannot comprehensively capture the regional adaptive reflectance characteristics of salt marsh and saline. Fortunately, many studies have built expert knowledge of these sub-categories over recent years. For example Mao et al. (2020) combined multi-scale segmentation, multiple normalized indices, and rule-based classification methods to develop a wetland map of China with an overall classification accuracy of 95.1%. Similarly, Wang et al. (2020) used the four widely used spectral indices to successfully identify three sub-categories within coastal wetlands. Thence, our further work would attach more effort on the spectral characteristics of salt marsh and saline wetlands and build expert knowledge of them for automatically deriving their training samples.

7. Data availability

The GWL_FCS30 wetland dataset in 2020 was freely available at https://doi.org/10.5281/zenodo.6575731 (Liu et al. 2022). It was composed of 961 5°×5° geographical grid tiled files, and each tiled file was stored using the geographical projection system with a spatial resolution of 30-meter in the GeoTIFF format. The fine wetland subcategory information was labeled as 0, 181, 182, 183, 184, 185 186 and 187, representing the non-wetland, swamp, marsh, flooded flat, saline, mangrove forest, salt marsh and tidal flat, respectively. The validation samples are available upon request.

8. Conclusions

Over the past few decades, many global and regional wetland products have been developed; however, an accurate global 30-m wetland dataset, with fine wetland categories and coverage of both inland and coastal zones, is still lacking. In this study, the time-series Landsat reflectance and Sentinel-1 SAR imagery, together with the stratified classification strategy and local adaptive random forest classification algorithm, were successfully integrated to produce the first global 30-m wetland product with a fine classification system in 2020. The wetlands were classified into four inland wetlands (swamp, marsh, flooded flat, and saline) and three coastal wetlands (mangrove, salt marsh, and tidal flat). The produced wetland dataset, GWL_FCS30, accurately captured the spatial patterns of seven wetland sub-categories with an overall accuracy of 87.7% and a kappa coefficient of 0.810 for the fine wetland classification system with lower omission and commission errors compared to other global products. The quantitative statistical analysis showed that the global wetland area reached 3.57 million km^2, including 3.10 million km^2 of inland wetlands and 0.47 million km^2 of coastal wetlands. Approximately 62.3% of wetlands were distributed poleward of 40°N. Therefore, the proposed method is suitable for large-area fine wetland mapping, and the GWL_FCS30 dataset can serve as an accurate wetland map that could potentially provide vital support for wetland management.

Acknowledgements

This research has been supported by the National Natural Science Foundation of China (grant no. 41825002) and National Earth System Science Data Sharing Infrastructure (grant no. 2005DKA32300).

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

References

Amani, M., Mahdavi, S., Afshar, M., Brisco, B., Huang, W., Mohammad Javad Mirzadeh, S., White, L., Banks, S., Montgomery, J., and Hopkinson, C.: Canadian Wetland Inventory using Google Earth Engine: The First Map and Preliminary Results, Remote Sensing, 11, 842, https://doi.org/10.3390/rs11070842, 2019.

Azzari, G. and Lobell, D. B.: Landsat-based classification in the cloud: An opportunity for a paradigm shift in land cover monitoring, Remote Sensing of Environment, 202, 64-74, https://doi.org/10.1016/j.rse.2017.05.025, 2017.

Belgiu, M. and Drăguț, L.: Random forest in remote sensing: A review of applications and future directions, ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24-31, https://doi.org/10.1016/j.isprsjprs.2016.01.011, 2016.

Breiman, L.: Random Forests, Machine Learning, 45, 5-32, https://doi.org/10.1023/a:1010933404324, 2001.

Bunting, P., Rosenqvist, A., Lucas, R., Rebelo, L.-M., Hilarides, L., Thomas, N., Hardy, A., Itoh, T., Shimada, M., and Finlayson, C.: The Global Mangrove Watch—A New 2010 Global Baseline of Mangrove Extent, Remote Sensing, 10, 1669, https://doi.org/10.3390/rs10101669, 2018.

Bwangoy, J.-R. B., Hansen, M. C., Roy, D. P., Grandi, G. D., and Justice, C. O.: Wetland mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices, Remote Sensing of Environment, 114, 73-86, https://doi.org/10.1016/j.rse.2009.08.004, 2010.

Cao, W., Zhou, Y., Li, R., and Li, X.: Mapping changes in coastlines and tidal flats in developing islands using the full time series of Landsat images, Remote Sensing of Environment, 239, 111665, https://doi.org/10.1016/j.rse.2020.111665, 2020.

Chen, B., Chen, L., Huang, B., Michishita, R., and Xu, B.: Dynamic monitoring of the Poyang Lake wetland by integrating Landsat and MODIS observations, ISPRS Journal of Photogrammetry and Remote Sensing, 139, 75-87, https://doi.org/10.1016/j.isprsjprs.2018.02.021, 2018.

Chen, G., Jin, R., Ye, Z., Li, Q., Gu, J., Luo, M., Luo, Y., Christakos, G., Morris, J., He, J., Li, D., Wang, H., Song, L., Wang, Q., and Wu, J.: Spatiotemporal Mapping of Salt Marshes in the Intertidal Zone of China during 1985–2019, Journal of Remote Sensing, 2022, 1-15, https://doi.org/10.34133/2022/9793626, 2022.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., and Mills, J.: Global land cover mapping at 30m resolution: A POK-based operational approach, ISPRS Journal of Photogrammetry and Remote Sensing, 103, 7-27, https://doi.org/10.1016/j.isprsjprs.2014.09.002, 2015.

Defourny, P., Kirches, G., Brockmann, C., Boettcher, M., Peters, M., Bontemps, S., Lamarche, C., Schlerf, M., and M., S.: Land Cover CCI: Product User Guide Version 2, available at: https://www.esa-landcover-cci.org/?q=webfm_send/84 (last access: 25 May 2022) 2018.

DeVries, B., Huang, C., Armston, J., Huang, W., Jones, J. W., and Lang, M. W.: Rapid and robust monitoring of flood events using Sentinel-1 and Landsat data on the Google Earth Engine, Remote Sensing of Environment, 240, 111664, https://doi.org/10.1016/j.rse.2020.111664, 2020.

Gómez, C., White, J. C., and Wulder, M. A.: Optical remotely sensed time series data for land cover classification: A review, ISPRS Journal of Photogrammetry and Remote Sensing, 116, 55-72, https://doi.org/10.1016/j.isprsjprs.2016.03.008, 2016.

Gardner, R. C. and Davidson, N. C.: The ramsar convention. In: Wetlands, Springer, https://doi.org/10.1007/978-94-007-0551-7_11, 2011.

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

Giri, C., Ochieng, E., Tieszen, L. L., Zhu, Z., Singh, A., Loveland, T., Masek, J., and Duke, N.: Status and distribution of mangrove forests of the world using earth observation satellite data, Global Ecology and Biogeography, 20, 154-159, https://doi.org/10.1111/j.1466-8238.2010.00584.x, 2011.

Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R.: Random Forests for land cover classification, Pattern Recognition Letters, 27, 294-300, https://doi.org/10.1016/j.patrec.2005.08.011, 2006.

Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, W., Bai, Y., Chen, B., Xu, B., Zhu, Z., Yuan, C., Ping Suen, H., Guo, J., Xu, N., Li, W., Zhao, Y., Yang, J., Yu, C., Wang, X., Fu, H., Yu, L., Dronova, I., Hui, F., Cheng, X., Shi, X., Xiao, F., Liu, Q., and Song, L.: Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017, Science Bulletin, 64, 370-373, https://doi.org/10.1016/j.scib.2019.03.002, 2019.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J., and Chen, J.: Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data, International Journal of Remote Sensing, 34, 2607-2654, https://doi.org/10.1080/01431161.2012.748992, 2013.

Gumbricht, T.: Hybrid mapping of pantropical wetlands from optical satellite images, hydrology, and geomorphology, Remote sensing of wetlands, 435-454, https://doi.org/10.1201/b18210, 2015.

Gumbricht, T., Roman‐Cuesta, R. M., Verchot, L., Herold, M., Wittmann, F., Householder, E., Herold, N., and Murdiyarso, D.: An expert system model for mapping tropical wetlands and peatlands reveals South America as the largest contributor, Global change biology, 23, 3581-3599, https://doi.org/10.1111/gcb.13689, 2017.

Guo, M., Li, J., Sheng, C., Xu, J., and Wu, L.: A Review of Wetland Remote Sensing, Sensors, 17, https://doi.org/10.3390/s17040777, 2017.

Hamilton, S. E. and Casey, D.: Creation of a high spatio-temporal resolution global database of continuous mangrove forest cover for the 21st century (CGMFC-21), Global Ecology and Biogeography, 25, 729-738, https://doi.org/10.1111/geb.12449, 2016.

Hansen, M. C., Egorov, A., Potapov, P. V., Stehman, S. V., Tyukavina, A., Turubanova, S. A., Roy, D. P., Goetz, S. J., Loveland, T. R., Ju, J., Kommareddy, A., Kovalskyy, V., Forsyth, C., and Bents, T.: Monitoring conterminous United States (CONUS) land cover change with Web-Enabled Landsat Data (WELD), Remote Sensing of Environment, 140, 466-484, https://doi.org/10.1016/j.rse.2013.08.014, 2014.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., and Townshend, J. R.: High-resolution global maps of 21st-century forest cover change, Science, 342, 850-853, https://doi.org/10.1126/science.1244693, 2013.

Hu, S., Niu, Z., and Chen, Y.: Global Wetland Datasets: a Review, Wetlands, 37, 807-817, https://doi.org/10.1007/s13157-017-0927-z, 2017a.

Hu, S., Niu, Z., Chen, Y., Li, L., and Zhang, H.: Global wetlands: Potential distribution, wetland loss, and status, Sci Total Environ, 586, 319-327, https://doi.org/10.1016/j.scitotenv.2017.02.001, 2017b.

Huang, X., Li, J., Yang, J., Zhang, Z., Li, D., and Liu, X.: 30 m global impervious surface area dynamics and urban expansion pattern observed by Landsat satellites: From 1972 to 2019, Science China Earth Sciences, https://doi.org/10.1007/s11430-020-9797-9, 2021.

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

Jia, M., Mao, D., Wang, Z., Ren, C., Zhu, Q., Li, X., and Zhang, Y.: Tracking long-term floodplain wetland changes: A case study in the China side of the Amur River Basin, International Journal of Applied Earth Observation and Geoinformation, 92, 102185, https://doi.org/10.1016/j.jag.2020.102185, 2020.

Jin, H., Stehman, S. V., and Mountrakis, G.: Assessing the impact of training sample selection on accuracy of an urban classification: a case study in Denver, Colorado, International Journal of Remote Sensing, 35, 2067-2081, https://doi.org/10.1080/01431161.2014.885152, 2014.

LaRocque, A., Phiri, C., Leblon, B., Pirotti, F., Connor, K., and Hanson, A.: Wetland Mapping with Landsat 8 OLI, Sentinel-1, ALOS-1 PALSAR, and LiDAR Data in Southern New Brunswick, Canada, Remote Sensing, 12, 2095, https://doi.org/10.3390/rs12132095, 2020.

Lehner, B. and Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands, Journal of hydrology, 296, 1-22, https://doi.org/10.1016/j.jhydrol.2004.03.028, 2004.

Li, Z., Chen, H., White, J. C., Wulder, M. A., and Hermosilla, T.: Discriminating treed and non-treed wetlands in boreal ecosystems using time series Sentinel-1 data, International Journal of Applied Earth Observation and Geoinformation, 85, 102007, https://doi.org/10.1016/j.jag.2019.102007, 2020.

Liu, L., Zhang, X., Gao, Y., Chen, X., Shuai, X., and Mi, J.: Finer-Resolution Mapping of Global Land Cover: Recent Developments, Consistency Analysis, and Prospects, Journal of Remote Sensing, 2021, 1-38, https://doi.org/10.34133/2021/5289697, 2021.

Liu, L., Zhang, X., Zhao, T., and Chen, X.,: GWL_FCS30: global 30 m wetland map with fine classification system in 2020. https://doi.org/10.5281/zenodo.6575731, 2022.

Lu, Y. and Wang, L.: How to automate timely large-scale mangrove mapping with remote sensing, Remote Sensing of Environment, 264, 112584, https://doi.org/10.1016/j.rse.2021.112584, 2021.

Ludwig, C., Walli, A., Schleicher, C., Weichselbaum, J., and Riffler, M.: A highly automated algorithm for wetland detection using multi-temporal optical satellite data, Remote Sensing of Environment, 224, 333-351, https://doi.org/10.1016/j.rse.2019.01.017, 2019.

Mahdianpari, M., Jafarzadeh, H., Granger, J. E., Mohammadimanesh, F., Brisco, B., Salehi, B., Homayouni, S., and Weng, Q.: A large-scale change monitoring of wetlands using time series Landsat imagery on Google Earth Engine: a case study in Newfoundland, GIScience & Remote Sensing, 57, 1102-1124, https://doi.org/10.1080/15481603.2020.1846948, 2020.

Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Homayouni, S., and Gill, E.: The First Wetland Inventory Map of Newfoundland at a Spatial Resolution of 10 m Using Sentinel-1 and Sentinel-2 Data on the Google Earth Engine Cloud Computing Platform, Remote Sensing, 11, 43, https://doi.org/10.3390/rs11010043, 2018.

Mao, D., Wang, Z., Du, B., Li, L., Tian, Y., Jia, M., Zeng, Y., Song, K., Jiang, M., and Wang, Y.: National wetland mapping in China: A new product resulting from object-based and hierarchical classification of Landsat 8 OLI images, ISPRS Journal of Photogrammetry and Remote Sensing, 164, 11-25, https://doi.org/10.1016/j.isprsjprs.2020.03.020, 2020.

Mao, D., Wang, Z., Wang, Y., Choi, C. Y., Jia, M., Jackson, M. V., and Fuller, R. A.: Remote Observations in China's Ramsar Sites: Wetland Dynamics, Anthropogenic Threats, and Implications for Sustainable Development Goals, Journal of Remote Sensing, 2021, 1-13, https://doi.org/10.34133/2021/9849343, 2021.

Matthews, E. and Fung, I.: Methane emission from natural wetlands: Global distribution, area, and environmental characteristics of sources, Global biogeochemical cycles, 1, 61-86, https://doi.org/10.1029/GB001i001p00061, 1987.

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

McCarthy, M. J., Radabaugh, K. R., Moyer, R. P., and Muller-Karger, F. E.: Enabling efficient, large-scale high-spatial resolution wetland mapping using satellites, Remote Sensing of Environment, 208, 189-201, https://doi.org/10.1016/j.rse.2018.02.021, 2018.

McOwen, C. J., Weatherdon, L. V., Bochove, J. V., Sullivan, E., Blyth, S., Zockler, C., Stanwell-Smith, D., Kingston, N., Martin, C. S., Spalding, M., and Fletcher, S.: A global map of saltmarshes, Biodivers Data J, https://doi.org/10.3897/BDJ.5.e11764, 2017.

Mellor, A., Boukir, S., Haywood, A., and Jones, S.: Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin, ISPRS Journal of Photogrammetry and Remote Sensing, 105, 155-168, https://doi.org/10.1016/j.isprsjprs.2015.03.014, 2015.

Murray, N. J., Phinn, S. R., DeWitt, M., Ferrari, R., Johnston, R., Lyons, M. B., Clinton, N., Thau, D., and Fuller, R. A.: The global distribution and trajectory of tidal flats, Nature, 565, 222-225, https://doi.org/10.1038/s41586-018-0805-8, 2019.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A.: Good practices for estimating area and assessing accuracy of land change, Remote Sensing of Environment, 148, 42-57, https://doi.org/10.1016/j.rse.2014.02.015, 2014.

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

Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brockmann, C., and Defourny, P.: Automated Training Sample Extraction for Global Land Cover Mapping, Remote Sensing, 6, 3965-3987, https://doi.org/10.3390/rs6053965, 2014.

Sexton, J., Feng, M., Channan, S., Song, X., Kim, D., Noojipady, P., Song, D., Huanga, C., Annand, A., and Collins, K.: Earth Science Data Records of Global Forest Cover and Change, User guide, available at: https://lpdaac.usgs.gov/documents/1370/GFCC_ATBD.pdf (last access: 25 May 2022), 2016.

Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K. M., Channan, S., DiMiceli, C., and Townshend, J. R.: Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error, International Journal of Digital Earth, 6, 427-448, https://doi.org/10.1080/17538947.2013.786146, 2013.

Slagter, B., Tsendbazar, N.-E., Vollrath, A., and Reiche, J.: Mapping wetland characteristics using temporally dense Sentinel-1 and Sentinel-2 data: A case study in the St. Lucia wetlands, South Africa, International Journal of Applied Earth Observation and Geoinformation, 86, 102009, https://doi.org/10.1016/j.jag.2019.102009, 2020.

Spalding, M., Kainuma, M., Collins, L.: World atlas of mangroves, Routledge, A collaborative project of ITTO, ISME, FAO, UNEP-WCMC, UNESCO-MAB, UNU-INWEH and TNC. London (UK): Earthscan, London, https://doi.org/10.34892/w2ew-m835, 2010.

Tachikawa, T., Hato, M., Kaku, M., and Iwasaki, A.: Characteristics of ASTER GDEM Version 2, Geoscience and Remote Sensing Symposium (IGARSS), 3657-3660, https://doi.org/10.1109/IGARSS.2011.6050017, 2011a.

Tachikawa, T., Kaku, M., Iwasaki, A., Gesch, D. B., Oimoen, M. J., Zhang, Z., Danielson, J., Krieger, T., Curtis, B., and Haase, J.: ASTER Global Digital Elevation Model Version 2 – Summary of validation results, Kim Fakultas Sastra Dan Budaya, https://doi.org/10.1093/oxfordjournals.pubmed.a024792, 2011b.

Thomas, N., Lucas, R., Bunting, P., Hardy, A., Rosenqvist, A., and Simard, M.: Distribution and drivers of global mangrove forest change, 1996–2010, PloS one, 12, https://doi.org/10.1371/journal.pone.0179302, 2017.

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

Tootchi, A., Jost, A., and Ducharne, A.: Multi-source global wetland maps combining surface water imagery and groundwater constraints, Earth Syst. Sci. Data, 11, 189-220, https://doi.org/10.5194/essd-11-189-2019, 2019.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., and Rostan, F.: GMES Sentinel-1 mission, Remote Sensing of Environment, 120, 9-24, https://doi.org/10.1016/j.rse.2011.05.028, 2012.

Townshend, J. R., Masek, J. G., Huang, C., Vermote, E. F., Gao, F., Channan, S., Sexton, J. O., Feng, M., Narasimhan, R., Kim, D., Song, K., Song, D., Song, X.-P., Noojipady, P., Tan, B., Hansen, M. C., Li, M., and Wolfe, R. E.: Global characterization and monitoring of forest cover using Landsat data: opportunities and challenges, International Journal of Digital Earth, 5, 373-397, https://doi.org/10.1080/17538947.2012.713190, 2012.

Veci, L., Prats-Iraola, P., Scheiber, R., Collard, F., Fomferra, N., and Engdahl, M.: The sentinel-1 toolbox, available at: https://sentinels.copernicus.eu/web/sentinel/toolboxes/sentinel-1 (last access: 25 May 2022), 2014.

Vermote, E., Justice, C., Claverie, M., and Franch, B.: Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product, Remote Sensing of Environment, 185, 46-56, https://doi.org/10.1016/j.rse.2016.04.008, 2016.

Wang, X., Xiao, X., Xu, X., Zou, Z., Chen, B., Qin, Y., Zhang, X., Dong, J., Liu, D., Pan, L., and Li, B.: Rebound in China's coastal wetlands following conservation and restoration, Nature Sustainability, https://doi.org/10.1038/s41893-021-00793-5, 2021.

Wang, X., Xiao, X., Zou, Z., Hou, L., Qin, Y., Dong, J., Doughty, R. B., Chen, B., Zhang, X., Chen, Y., Ma, J., Zhao, B., and Li, B.: Mapping coastal wetlands of China using time series Landsat images in 2018 and Google Earth Engine, ISPRS J Photogramm Remote Sens, 163, 312-326, https://doi.org/10.1016/j.isprsjprs.2020.03.014, 2020.

Worthington, T. A., Zu Ermgassen, P. S., Friess, D. A., Krauss, K. W., Lovelock, C. E., Thorley, J., Tingey, R., Woodroffe, C. D., Bunting, P., and Cormier, N.: A global biophysical typology of mangroves and its relevance for ecosystem structure and deforestation, Scientific reports, 10, 1-11, https://doi.org/10.1038/s41598-020-71194-5, 2020.

Xi, Y., Peng, S., Ciais, P., and Chen, Y.: Future impacts of climate change on inland Ramsar wetlands, Nature Climate Change, 11, 45-51, https://doi.org/10.1038/s41558-020-00942-2, 2020.

Zhang, H., Wang, T., Liu, M., Jia, M., Lin, H., Chu, L. M., and Devlin, A.: Potential of Combining Optical and Dual Polarimetric SAR Data for Improving Mangrove Species Discrimination Using Rotation Forest, Remote Sensing, 10, 467, https://doi.org/10.3390/rs10030467, 2018.

Zhang, H. K. and Roy, D. P.: Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification, Remote Sensing of Environment, 197, 15-34, https://doi.org/10.1016/j.rse.2017.05.024, 2017.

Zhang, X., Liu, L., Chen, X., Gao, Y., and Jiang, M.: Automatically Monitoring Impervious Surfaces Using Spectral Generalization and Time Series Landsat Imagery from 1985 to 2020 in the Yangtze River Delta, Journal of Remote Sensing, 2021, 1-16, https://doi.org/10.34133/2021/9873816, 2021a.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., and Mi, J.: GLC_FCS30: global land-cover product with fine classification system at 30 m using time-series Landsat imagery, Earth Syst. Sci. Data, 13, 2753-2776, https://doi.org/10.5194/essd-13-2753-2021, 2021b.

https://doi.org/10.5194/essd-2022-180
Preprint. Discussion started: 19 August 2022
© Author(s) 2022. CC BY 4.0 License.

935 Zhang, X., Liu, L., Wu, C., Chen, X., Gao, Y., Xie, S., and Zhang, B.: Development of a global 30 m impervious surface map using multisource and multitemporal remote sensing datasets with the Google Earth Engine platform, Earth Syst. Sci. Data, 12, 1625-1648, https://doi.org/10.5194/essd-12-1625-2020, 2020.

Zhang, X., Liu, L., Zhao, T., Gao, Y., Chen, X., and Mi, J.: GISD30: global 30 m impervious-surface dynamic dataset from 1985 to 2020 using time-series Landsat imagery on the Google Earth Engine platform, Earth Syst. 940 Sci. Data, 14, 1831-1856, https://doi.org/10.5194/essd-14-1831-2022, 2022a.

Zhang, Z., Xu, N., Li, Y., and Li, Y.: Sub-continental-scale mapping of tidal wetland composition for East Asia: A novel algorithm integrating satellite tide-level and phenological features, Remote Sensing of Environment, 269, 112799, https://doi.org/10.1016/j.rse.2021.112799, 2022b.

Zhu, P. and Gong, P.: Suitability mapping of global wetland areas and validation with remotely sensed data, 945 Science China Earth Sciences, 57, 2283-2292, https://doi.org/10.1007/s11430-014-4925-1, 2014.

Zhu, Z., Gallant, A. L., Woodcock, C. E., Pengra, B., Olofsson, P., Loveland, T. R., Jin, S., Dahal, D., Yang, L., and Auch, R. F.: Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative, ISPRS Journal of Photogrammetry and Remote Sensing, 122, 206-221, https://doi.org/10.1016/j.isprsjprs.2016.11.004, 2016.

950 Zhu, Z., Wang, S. X., and Woodcock, C. E.: Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images, Remote Sensing of Environment, 159, 269-277, https://doi.org/10.1016/j.rse.2014.12.014, 2015.

Zhu, Z. and Woodcock, C. E.: Object-based cloud and cloud shadow detection in Landsat imagery, Remote Sensing of Environment, 118, 83-94, https://doi.org/10.1016/j.rse.2011.10.028, 2012.

955