Reasonable agreements and mismatches between land-surface-water-area estimates based on a global river model and Landsat data

Xudong Zhou[1], Catherine Prigent[2], and Dai Yamazaki[3]

[1]The University of Tokyo
[2]Observatoire de Paris
[3]University of Tokyo

November 22, 2022

Abstract

Land surface water is a key component of the global water cycle. Compared to remote sensing by satellites, both temporal extension and spatial continuity is superior in modeling of water surface area. However, overall evaluation of models representing different kinds of surface waters at the global scale is lacking. We estimated land surface water area (LSWA) using the Catchment-based Macro-scale Floodplain model (CaMa-Flood), a global hydrodynamic model, and compared the estimates to Landsat with 3” spatial resolution at the global scale. Results show that the two methodologies show agreement in the general spatial patterns of LSWA (e.g., major rivers and lakes, open-to-sky floodplains), but globally consistent mismatches were found under several land surface conditions. CaMa-Flood underestimates LSWA in high northern latitudes (e.g., the Canadian Shield) and coastal areas, as the presence of isolated lakes in local depressions or small coastal rivers is not considered by the model’s physical assumptions. In contrast, model-estimated LSWA is larger than Landsat estimates in forest-covered areas (e.g., Amazon basin) due to the opacity of vegetation for optical satellite sensing, and in cropland areas due to the lack of dynamic water processes (e.g., re-infiltration, evaporation, water consumption) and constraints of water infrastructure (e.g., canals, levees). These globally consistent differences can be reasonably explained by the model’s physical assumptions or optical satellite sensing characteristics, and applying filters (e.g., floodplain topography mask, forest and cropland mask) to the two datasets allows the remaining local-scale discrepancies to be attributed to locally varying factors (e.g., channel parameters, atmospheric forcing).

Reasonable agreements and mismatches between land-surface-water-area estimates based on a global river model and Landsat data

Xudong Zhou[1], Catherine Prigent[2,3], and Dai Yamazaki[1]

[1] Institute of Industrial Science, The University of Tokyo, Tokyo, 153-8505, Japan

[2] Sorbonne Université, Observatoire de Paris, Université PSL, CNRS, LERMA, Paris, France

[3] Estellus, Paris, France

Corresponding author: Xudong Zhou (x.zhou@rainbow.iis.u-tokyo.ac.jp)

Key Points:

- Land surface water area is estimated globally at 3″ spatial resolution using CaMa-Flood and compared with Landsat.
- Agreement and mismatch between model and satellite data exhibit spatial features associated with topography and land cover conditions.
- Applying appropriate filtering masks to model and satellite results enables reasonable comparison of water surface areas.

Abstract

Land surface water is a key component of the global water cycle. Compared to remote sensing by satellites, both temporal extension and spatial continuity is superior in modeling of water surface area. However, overall evaluation of models representing different kinds of surface waters at the global scale is lacking. We estimated land surface water area (LSWA) using the Catchment-based Macro-scale Floodplain model (CaMa-Flood), a global hydrodynamic model, and compared the estimates to Landsat with 3″ spatial resolution at the global scale. Results show that the two methodologies show agreement in the general spatial patterns of LSWA (e.g., major rivers and lakes, open-to-sky floodplains), but globally consistent mismatches were found under several land surface conditions. CaMa-Flood underestimates LSWA in high northern latitudes (e.g., the Canadian Shield) and coastal areas, as the presence of isolated lakes in local depressions or small coastal rivers is not considered by the model's physical assumptions. In contrast, model-estimated LSWA is larger than Landsat estimates in forest-covered areas (e.g., Amazon basin) due to the opacity of vegetation for optical satellite sensing, and in cropland areas due to the lack of dynamic water processes (e.g., re-infiltration, evaporation, water consumption) and constraints of water infrastructure (e.g., canals, levees). These globally consistent differences can be reasonably explained by the model's physical assumptions or optical satellite sensing characteristics, and applying filters (e.g., floodplain topography mask, forest and cropland mask) to the two datasets allows the remaining local-scale discrepancies to be attributed to locally varying factors (e.g., channel parameters, atmospheric forcing).

1 Introduction

Land surface water area (hereafter LSWA) is of paramount importance to the survival of all life forms (Karpatne et al., 2016). Water not only provides habitat for aquatic organisms, but also affects various aspects of human life, such as for drinking and agricultural, domestic and industrial purposes (Vorosmarty and Sahagian, 2000). LSWA is highly dynamic and variations therein can be used as a direct indicator of climate change (Williamson et al., 2009) or human-induced changes (Pekel et al., 2016). LSWA is thus an essential variable in ecological, hydrological, climatic and economic studies (Hirabayashi et al., 2013; Raymond et al., 2013; Willner et al., 2018). For such applications, accurate water information at adequate spatiotemporal resolution is crucial.

Estimation of LSWA relies on three methods: ground surveys, remote sensing and models. Among these methods, ground surveys cannot fully describe the water dynamics due to their slow updating frequency (Carroll et al., 2009; Lehner and Döll, 2004) and the significant cost of covering a large spatial domain. Remote sensing using satellites is an outstanding method that can provide regular large-scale observations of water surfaces. Various satellites have been used to identify LSWA, including Landsat (Pekel et al., 2016; Qi et al., 2009), MODIS (Ji et al., 2018; Lai et al., 2014), and a combination of passive and active microwave satellites (Prigent et al., 2007; Schumann and Moller, 2015). Hydrodynamic models provide another method for determining water area and dynamics. Hydrodynamic models provide a powerful tool that can produce continuous water maps over time and space, regardless of weather (e.g., cloudy) or vegetation cover. Moreover, models are the only way to hindcast the water surface in the past before satellites were launched (Lewin and Hughes, 1980) and forecast the future changes when no observational results exist (Hirabayashi et al., 2013). Over the past two decades, several hydrodynamic models have been developed (e.g., LISFLOOD-FP, HEC-RAS, MIKE-Flood, DELFT3D, CaMa-Flood) and tested under various conditions (Bates and De Roo, 2000; Pappenberger et al., 2005; Patro et al., 2009; Dingle et al., 2020; Yamazaki et al., 2011). For a detailed review of flood inundation models, refer to Teng et al. (2017).

Modeling of river hydrodynamics builds on a process chain from climate forcing to runoff and then to routing. While simulation of river discharge is relatively straightforward, as it is explained mainly by the basin-integrated water budget, simulation of modeled LSWA is more difficult, as it is affected by local topography in addition to the basin-wide water budget. Therefore, estimates of LSWA contain multiple sources of uncertainties and require validation against observational results, which are generally satellite-derived inundation maps. Due to the need for high-quality large-scale topography data and model parameters as well as high computational capacity, most validations have been conducted for small catchments (e.g., the 10-km Alzette River (Schumann et al., 2007) and 60-km Severn River (Horritt, 2006)). These studies mainly compared inundation during specific flood events within a short period against inundation maps at a relatively high spatial resolution (Horritt, 2000, 2006; Khan et al., 2011; Revilla-Romero et al., 2015; Schumann et al., 2007; Try et al., 2018; Wilson et al., 2007), with a primary focus on evaluating whether the model could reasonably reproduce the flooding distribution in the region of interest.

However, those local studies are insufficient for determining the capacity of a model to represent the water surface extent under different conditions. For example, previous local studies have generally investigated the ability of a model to map inundation in the form of open-to-sky floodplains, and have not tested model performance on other water forms (e.g., normal rivers, thawing lakes, and man-made water areas such as dam reservoirs and irrigated fields), which account for a large portion of global water surface (Lehner and Döll, 2004). In addition, model validation at the local scale cannot attribute simulation errors to globally consistent issues related to the model assumptions or satellite characteristics or to locally varying error sources such as topography, channel parameters and input forcing data. Therefore, application and validation at large scales, from continental to global, are required to clarify the applicability of hydrodynamic models under various conditions and for different water forms.

By reducing the spatial resolution and improving the computation capacity, flood model applications have been expanded to the scales of large river basins (Try et al., 2018; Wilson et al., 2007), continents (Decharme et al., 2008; Schumann et al., 2016) and global (Decharme et al., 2012, Yamazaki et al., 2011). Global Inundation Extent from Multi-Satellites (GIEMS) data, which were first released in 2007 and have been occasionally updated (Prigent et al. 2007; Papa et al., 2010, Prigent et al., 2020), provide the most frequently used referenced satellite inundation maps for validation of model performance over large areas and long term. Wu et al. (2019) compared global modeling results with fractional water cover retrieved from enhanced brightness temperatures acquired by the Soil Moisture Active Passive (SMAP, Chaubell et al., 2018) mission. However, both of these applications fail to provide information on smaller water bodies due to their coarse spatial resolutions of 25 km and 9 km, respectively, and therefore they cannot answer the question of whether such water surfaces are well represented by global flood models. Moderate Resolution Imaging Spectroradiometer (MODIS, 500 m; Li et al., 2018) and Landsat (30 m; Pekel et al., 2016) products might be useful for answering this question, as they provide global water surface area data at a much finer spatial resolution, which can adequately represent individual water bodies. However, these two new products have not yet been utilized for large-scale model validation.

As noted above, various types of water bodies are formed and impacted by different external forcing factors and land surface conditions (Lehner and Döll, 2004; Pekel et al., 2016, Ji et al., 2018). Satellite-derived results from different sources will deviate in their estimates of water extent, depending on the location and size of the water bodies, as well as the weather and land surface conditions at the time of observation (Aires et al., 2018; Huang et al., 2018; Lamarche et al., 2017; Notti et al., 2018, Pham-Duc et al., 2017). Although multiple satellite-derived results have been used for validating hydrodynamic model performance, we have not sufficiently investigated when applying to validation how good the satellites themselves can identify different water types at a large scale. Meanwhile, hydrodynamic models have specific physics and limitations, and it is not possible to represent all types of water bodies accurately using existing model structures and physical assumptions. Therefore, when making comparisons between model and satellite results, pre-processing of the data is necessary. For example, Decharme et al. (2012) subtracted cropland area from GIEMS data to validate their model performance more reasonably, as their model did not include human processes. However, the importance of data processing is neglected in most current studies, making interpretation of the agreement or mismatch between model and satellite difficult and sometimes misleading.

In this study, we estimate global LSWA using the Catchment-based Macro-scale Floodplain model (CaMa-Flood), a global hydrodynamic model, at a much finer spatial resolution (3″) than previous global-scale model validation studies. The estimation focuses not only on floods but also includes other water forms under normal conditions. Evaluation is conducted against the Landsat water-occurrence product (Pekel et al., 2016). We discuss where the model and Landsat measurements agree and where globally consistent mismatches occur that can be reasonably explained by the limitations or characteristics of the model or satellite, rather than locally varying error sources. Then, we introduce various filtering masks and land cover conditions used to make reasonable and adequate comparisons of water surface areas between models and satellites. Finally, we provide instructions for making appropriate comparisons, including areas where the comparison of raw values from models and satellites are valid, the types of water surfaces that cannot be captured by model simulations, and the filters that should be applied to conduct appropriate comparisons.

2 Materials and Methods

2.1 Satellite products

2.1.1 Landsat

The historical water surface occurrence data used in this study were generated by Pekel et al. (2016) based on three million Landsat satellite images obtained between 1984 and 2015. The months in which water was present were recorded. Water occurrence was estimated as the ratio of months with water to the entire time period, excluding time points with invalid data (missing data, cloud or snow cover). This exclusion will affect the accuracy of estimates, especially in tropical regions where the cloud index is high and at high latitudes where snow cover is common. Due to the availability of its high-resolution and long-term data, the Landsat water-occurrence product has been used as a reference for water classification (Ji et al., 2018; Senyurek et al., 2020). The original Landsat water-occurrence product has spatial resolution of 1″ (~30 m at the equator), which is aggregated to 3″ (~90 m) to match the minimum spatial resolution of CaMa-Flood. Details of the processing of the Landsat water-occurrence product can be found in the original report (Pekel et al., 2016).

2.1.2 GIEMS

The GIEMS product is derived from a series of satellite sensors, primarily passive microwaves (Special Sensor Microwave/Imager, SSM/I), with additional data from visible and near-infrared observations and active microwave measurements. GIEMS is originally calculated on an equal-area grid of 0.25° at the Equator, and has been interpolated in this study to regular grids of 0.25°×0.25° for comparison with the other products. GIEMS is available monthly, and the latest version, GIEMS-2, extends the available period to 1992–2015 (Prigent et al., 2020). For details of data processing, see previous reports (Prigent et al., 2001, 2007, 2020). In this study, Landsat water-occurrence data were used as the primary reference, with GIEMS as a supplementary dataset to explain differences in water surface areas between the model and Landsat.

2.2 CaMa-Flood

CaMa-Flood is a global hydrodynamic model for continental-scale rivers. River networks are discretized into irregular unit catchments with sub-grid topographic parameters of river channels and floodplains. River discharge and other flow characteristics can be calculated using the local inertial equations along the river network map (MERIT Hydro, Yamazaki et al., 2019). Water storage in each catchment unit is the prognostic variable, and is determined using the water balance equation. The water level and flooded area are identified from the water storage in each unit catchment based on the sub-grid topographic information. Detailed descriptions of CaMa-Flood can be found in the original papers by Yamazaki et al.（2011, 2012, 2014).

2.2.1 Model settings

The overall workflow of this study is illustrated in Figure 1. We ran CaMa-Flood globally from 2001 to 2014. The unit catchment in CaMa-Flood was set to 0.1° spatial resolution (~10 km), meaning that only one unit catchment was assigned to each 0.1°×0.1° grid. This resolution is high for global studies, but still inadequate, especially in coastal regions and mountainous headwater catchments where multiple small rivers occur within a grid. As an input runoff for CaMa-Flood, we used eartH2Observe runoff data produced by the land surface hydrological model HTESSEL and forced with WFDEI weather boundary conditions (Balsamo et al., 2009). Runoff was provided at 0.25° resolution, and therefore was distributed to each unit catchment according to the areal proportion of the unit catchment in the corresponding grid.

2.2.2 Downscaling

Although 0.1° is a high spatial resolution for global modeling, it is insufficient for representing small water bodies and rapid changes in the water surface area (Fluet-Chouinard et al., 2015; Winsemius et al., 2013). Therefore, the CaMa-Flood outputs were downscaled to 3″ using high-resolution topography information (MERIT DEM), which is directly comparable to the high-resolution Landsat occurrence product. The downscaling process was based on the fundamental assumptions of CaMa-Flood that the movement of water within a unit catchment is instantaneous and that the water surface is flat within the unit catchment at each time step. The area of lowest elevation is inundated first, until the total water volume approximates the estimated water storage of the unit catchment. To reduce the computational cost, we first calculated the depth-duration curve (i.e., the cumulative number of days on which the water level was above different elevations at an interval of 0.1 m) for each unit catchment in each year. Downscaling was conducted by projecting the number of days when the water surface elevation of the flooded unit catchment exceeded the ground elevation of the corresponding 3″ DEM pixel. The downscaled inundation water extent was determined using the same flood duration for a given elevation. The final result is approximately equal to the result from direct downscaling (i.e., first downscaling the simulated flood depth for each day and later aggregating the inundation days), but this process saved significant time, as the number of repeats used for downscaling was efficiently reduced.

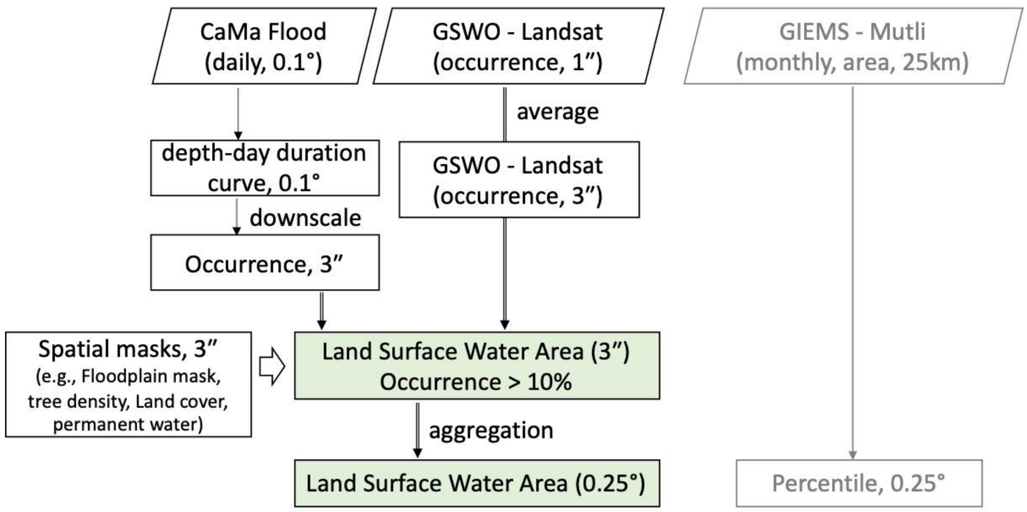

Figure 1. Flow chart of data preparation for water surface area from CaMa-Flood and two other data types derived from satellite remote sensing (Landsat and GIEMS).

2.3 Occurrence selection

Water occurrence ranges from 0% in non-water areas to 100% in permanent water areas. In this study, we aim to evaluate the abilities of the model and satellite data to capture different types of water bodies, and therefore we primarily examine areas with water occurrences greater than 10%. The water area above this threshold includes permanent water and most seasonal water. This threshold removes areas that are flooded only during very extreme flood events, to compare general trends between the model and the satellite data. This threshold also reduces the impact on water-area estimates of high sensitivity at the tail end of the low-occurrence criterion. Such a low threshold reduces the impact of cloud obstruction, as the affected areas are generally counted. The sum of the water areas with greater than 10% occurrence is the LSWA discussed in this study (Figure 1). The resolution is 3″, which is aggregated to 0.25° for better visualization and comparison with other products at 0.25° spatial resolution (e.g., GIEMS). Meanwhile, water surfaces with different occurrences can be mapped as needed for interpretation of the features present in model and satellite results for specific regions.

2.4 Spatial masking

Using the data-processing steps described above, occurrence data from CaMa-Flood and Landsat are both available at 3″ resolution. However, due to the properties of the model and satellites, discrepancies in water surface area occur with typical spatial patterns at the global scale, which are associated with various land surface conditions. To facilitate comparison, we applied different filtering masks (maps, 3″) to water-surface products from these two sources (Figure 1). Therefore, differences are grouped within the same land surface condition, allowing the source of the difference to be attributed to the limitations or characteristics of the model or satellite.

2.4.1 Land masking

A land mask excluding all seawater areas was applied to the Landsat occurrence product prior to comparison, as some marine areas along coastlines are included in the Landsat dataset. The land mask was prepared from a global hydrography dataset (MERIT Hydro) (Yamazaki et al., 2019), which is also used as the baseline map for CaMa-Flood. Applying the land mask to Landsat data ensures that the two water-surface products cover the same spatial extent of land.

2.4.2 CaMa-Flood floodplain masking

The water surface in CaMa-Flood is based on a few assumptions. First, all water from the input runoff data directly enters the river channel, and the water surface is formed by surface runoff routed along river networks. Water bodies that are recharged from other sources (e.g., melting snow and ice, shallow groundwater appearing at the surface, tides, or pluvial flooding due to local rainfall) and local depressions other than river channels are therefore not modeled. Second, CaMa-Flood assumes that the water surface is flat within each unit catchment (Figure 2-a); however, this assumption is invalid for rivers with high surface gradients, particularly mountainous springs. Third, because only one major river can be represented in each unit catchment, small coastal rivers are neglected in favor of major rivers. Underestimation of the water surface area is apparent at the local scale, especially where small water bodies (e.g., narrow rivers, small lakes, coastal rivers) are abundant. Although such water surfaces are relatively small, they can be captured by Landsat (Pekel et al., 2016).

Therefore, we prepared a floodplain mask that defines the potential maximum extent that can be simulated by CaMa-Flood (red line in the schematic diagrams in Figure 2-b,c) based on CaMa-Flood sub-grid topography. This mask is accomplished through inundation area downscaling from the historical maximum floodplain water elevation estimated by CaMa-Flood from 2001–2014. We increased these values by 1.5 times and set all values below 2.0 m (but above 0) to 2.0 m to consider the impact of uncertainties in runoff forcing on CaMa-Flood (Figure 2-b,c). The floodplain mask was then applied to the Landsat occurrence product to separate the results within and outside the potential maximum extent covered by CaMa-Flood. CaMa-Flood does not represent water outside this floodplain mask due to its modeling structure, and therefore water outside the floodplain mask in Landsat is excluded from comparisons when the floodplain mask is applied.

In addition, as CaMa-Flood calculates the hydrodynamics of only one major river within each unit catchment corresponding to a 0.1° grid box, only inundations of floodplains along the major river within each 0.1° grid box are simulated. Thus, inundations in small coastal river basins are not represented due to the assumptions of CaMa-Flood, and are excluded from the floodplain mask to allow for direct comparison against the Landsat water map.

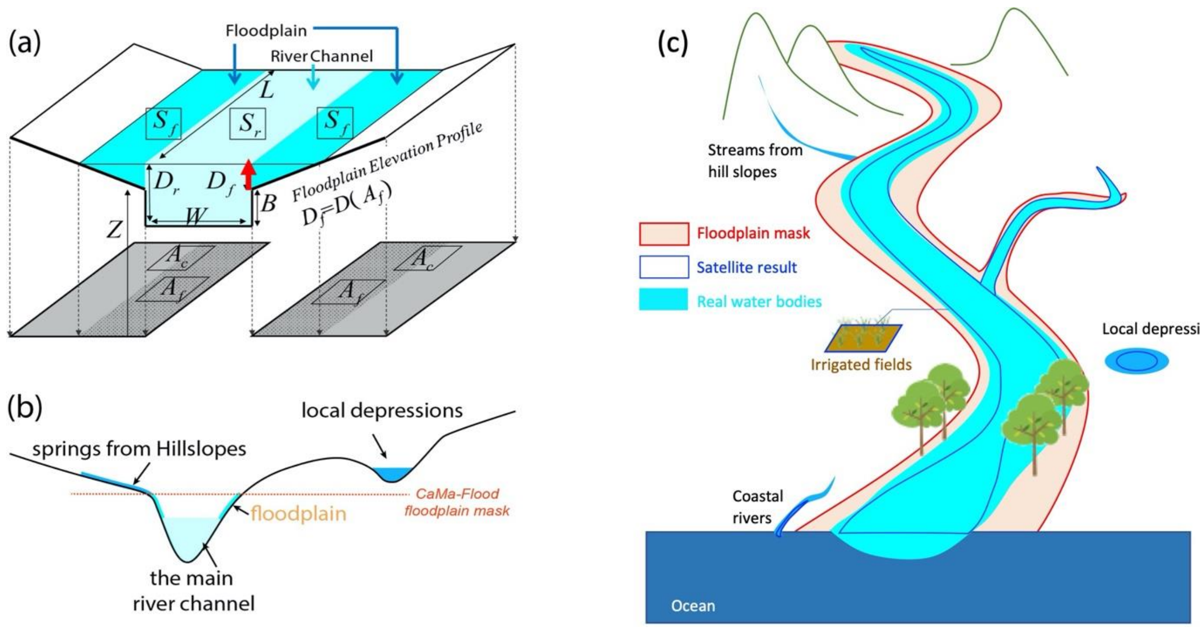

Figure 2. Schematic diagrams of (a) the river and floodplain representations in CaMa-Flood, (b) the realistic river profile and floodplain mask applied in this study, and (c) the different water bodies (e.g., rivers, local depressions, streams from hill slopes, coastal rivers, irrigated fields) and the floodplain mask as well as an illustration of the results from Landsat. The floodplain is approximated as a monotonically increasing function in CaMa-Flood, and therefore land water surfaces on hill slopes, local depressions, irrigated fields and coastal rivers are not well represented. The floodplain mask was introduced to exclude water areas that are not represented by CaMa-Flood from the analysis.

2.4.3 European Space Agency Climate Change Initiative (ESA CCI) land cover map

The ESA CCI land cover map was utilized in this study to determine land surface conditions. Water surface areas with different land cover types were grouped and compared between the model and satellite results to illustrate the relationship of water surface area with land cover type (e.g., forests, croplands, wetlands). The original CCI product was at 300-m spatial resolution, which was interpolated to 3″ using a simple nearest-neighbor interpolation method.

2.4.4 Tree density map

A limitation of optical satellites is that clouds and thick vegetation cannot be penetrated, thus, the water under clouds and thick vegetation is difficult to be detected. In the Landsat occurrence product, images with cloud cover are removed, but the impact of vegetation cannot be eliminated from the observations (Pekel et al., 2016). Although the CCI land cover map also contains information regarding trees, it does not provide tree density, and the performance of model or satellite data differs with the level of tree density. Therefore, we prepared a global tree density map (Hansen et al., 2013), which was originally at 3″ resolution. This high-resolution tree density was averaged to 0.25° for better visualization and comparison with other 0.25° maps (Figure S2). The density is a percentage value from 0 to 100, with higher values indicating denser vegetation. The maximum tree density is found in the Amazon River Basin, the Congo River Basin and the Indonesian Islands. Notably, the tree density value does not indicate the height of vegetation or the thickness of leaves, especially in high-latitude regions where needle-leaved or short vegetation dominates.

2.4.5 Static permanent water mask

Although channel bathymetry is considered in the model simulation using sub-grid parameters, underwater topography is not considered in the downscaling procedure because the high-resolution MERIT DEM represents mean water surface elevations over all water body pixels. Thus, the downscaled flooded water depth represents the water depth above the MERIT DEM, and not water stored below the MERIT DEM surface. This process leads to underestimation of CaMa-Flood water surface area during low-water seasons when water remains within a sub-grid river channel. Therefore, we extracted the permanent water surface where the occurrence is 100% (dark blue line in Figure 2) from the Landsat data. Permanent water is present with high confidence. If necessary, CaMa-Flood results within the statistical permanent water mask can be determined during post-processing by modifying the transitory water (occurrence < 100%) to permanent water (occurrence = 100%).

3 Results

3.1 Global land surface water from the model and Landsat

Figure 3 shows the global distribution of land surface water with occurrence greater than 10%. The original dataset has 3″ (~90 m) spatial resolution, which is aggregated to a 0.25° (~25 km) grid for better visualization. In total, the estimated water surface area is 3.98 million km^2 (hereafter Mkm^2) in CaMa-Flood and 5.53 Mkm^2 in Landsat. Except in Greenland, very little water surface is estimated by CaMa-Flood in mountainous (e.g., the Rocky Mountains and the Andes Mountains) and dry regions (e.g., Northern Africa, Central Asia and central Austria) (gray in Figure 3-a). The lack of water estimates in such areas is either due to insufficient surface runoff to form water bodies in dry regions or due to difficulty in representing rivers in mountainous areas by CaMa-Flood. In the Landsat water-occurrence product, the corresponding regions have larger values than CaMa-Flood, although the absolute values are still small (light yellow in Figure 3-b). As a result, the difference between the two datasets is very small in mountainous and dry regions (gray in Figure 3-c). Both CaMa-Flood and Landsat can delineate rivers and lakes. Large water surface areas (dark blue in Figure 3-a,b) are shown for lakes (e.g., the Caspian Sea, the Great Lakes, Lake Victoria), large rivers (e.g., the Amazon, the Ob, and Yangtze) and delta regions (e.g., the Mekong, Ganges, and Indus Deltas).

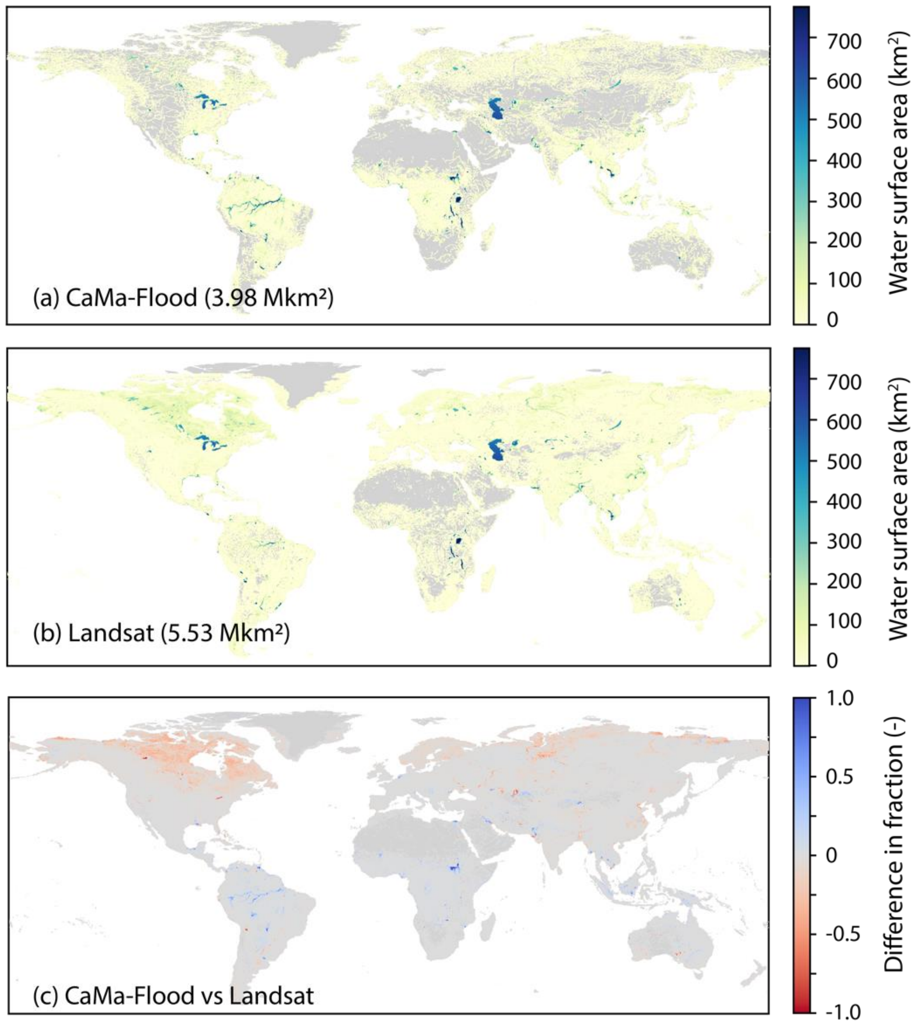

Figure 3. Global land water surface areas with water occurrences greater than 10%. (a) Results from CaMa-Flood, and (b) results from Landsat. (c) Differences between CaMa-Flood and Landsat in terms of the fraction of each grid (0.25°). Original results are at 3″, and are aggregated to 0.25° gridded values for visualization. Areas with no water surface (= 0) are masked out.

The two methodologies tested in this study showed good agreement in water surface area, especially over lakes (e.g., the Great Lakes, Caspian Sea, Lake Baikal, Lake Victoria, Tonle Sap Lake) (Figure 3-c). Strong agreement was also found along major rivers, aside from the Amazon and river deltas, for which differences were difficult to identify (Figure 3-c). The differences showed apparent spatial patterns (Figure 3-c). Lower estimates were obtained from CaMa-Flood than from satellite data for high-latitude regions, especially in the Canadian Shield and in the lower Ob River. Other differing regions included the Tibetan Plateau, the middle-lower reaches of the Yangtze and Ganges, and certain rivers in central Asia and northern Europe. In contrast, larger water surfaces were found along the Amazon and Indonesian rivers in CaMa-Flood. In addition, higher values were found in many river deltas such as those of the Nile, Mississippi, Congo, Tigris & Euphrates, Indus, and rivers in Southeast Asia. Two typical regions with high values were in South Sudan and the lower Tarim River in China. These discrepancies (overestimation and underestimation) are explained and discussed in the following sections, along with additional masks and the topographic maps used in this study.

3.2 Analysis with the CaMa-Flood floodplain mask

The extent of the CaMa-Flood floodplain mask is shown in Figure 4-a. The floodplain mask is the theoretical boundary where CaMa-Flood may simulate inundation. As the floodplain mask has been enlarged from that used in real simulations, applying the floodplain mask does not change the results of CaMa-Flood. However, only part of the water surface in the Landsat dataset falls within the floodplain mask (Figure 4-b), with a total of 4.20 Mkm^2 LSWA located within the floodplain mask.

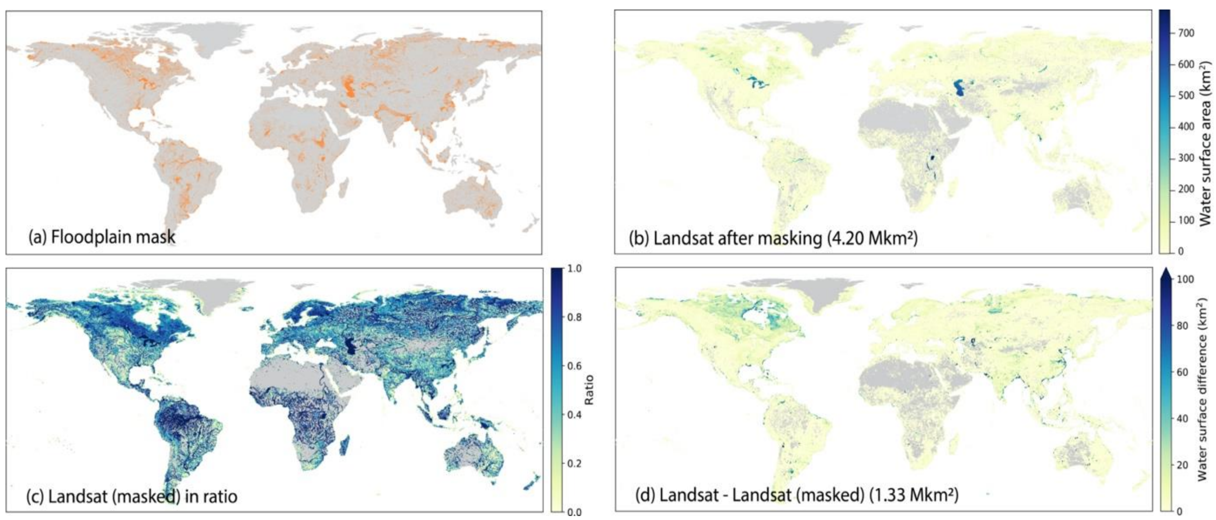

Figure 4. (a) Distribution of the CaMa-Flood floodplain mask over the globe (3″). (b) Landsat LSWA after applying the floodplain mask (0.25°). (c) The ratio of Landsat LSWA within the CaMa-Flood floodplain mask to total Landsat LSWA (0.25°). (d) The difference in the value of Landsat LSWA within the CaMa-Flood floodplain mask from the total Landsat LSWA (0.25°).

Figures 4-c and 4-d show the proportion and amount of water surface in Landsat that falls within the floodplain mask relative to the total LSWA, respectively. At 0.25° spatial resolution, the areal ratio is near 1.0 for large lakes and rivers (Figure 4-c), indicating that these large-scale water types are well covered by the CaMa-Flood floodplain mask. Low ratios of water surface were found on hill slopes, especially in mountainous areas. An example is given of the upstream Missouri River (Figure 5). The yellow color shows the extent of the CaMa-Flood floodplain mask, which covers the main river channel and most smaller tributaries. CaMa-Flood and Landsat tend to produce the same water surface in the main channel (blue in Figure 5-a). Because the floodplain mask was already extended from the historical maxima of modeled flood extent, some small tributaries lack water in both the simulation results and satellite observations. Outside the CaMa-Flood floodplain mask (Figure 5-b), Landsat is able to detect small water areas scattered across hill slopes, which cannot be modeled by CaMa-Flood due to its physical assumptions. As shown in the enlarged topographic map (Figure 5-c), the water surface is not continuous, and the distribution of the water surface is not consistent from lower elevations to higher elevations within each unit catchment. This inconsistency is caused by the unique kettle lake landform (Figure 5-d) of the Missouri Plateau, which was formed by retreating glaciers or draining floodwaters rather than surface river flows (Phillips and Gleckler, 2006). Another typical kettle lake landform is found in the Western Siberian Plain, near the Arctic Circle. However, for hill slopes other than this kettle lake landform, the main cause of water surface area underestimation in CaMa-Flood is the invalid assumption of a flat water surface used for downscaling.

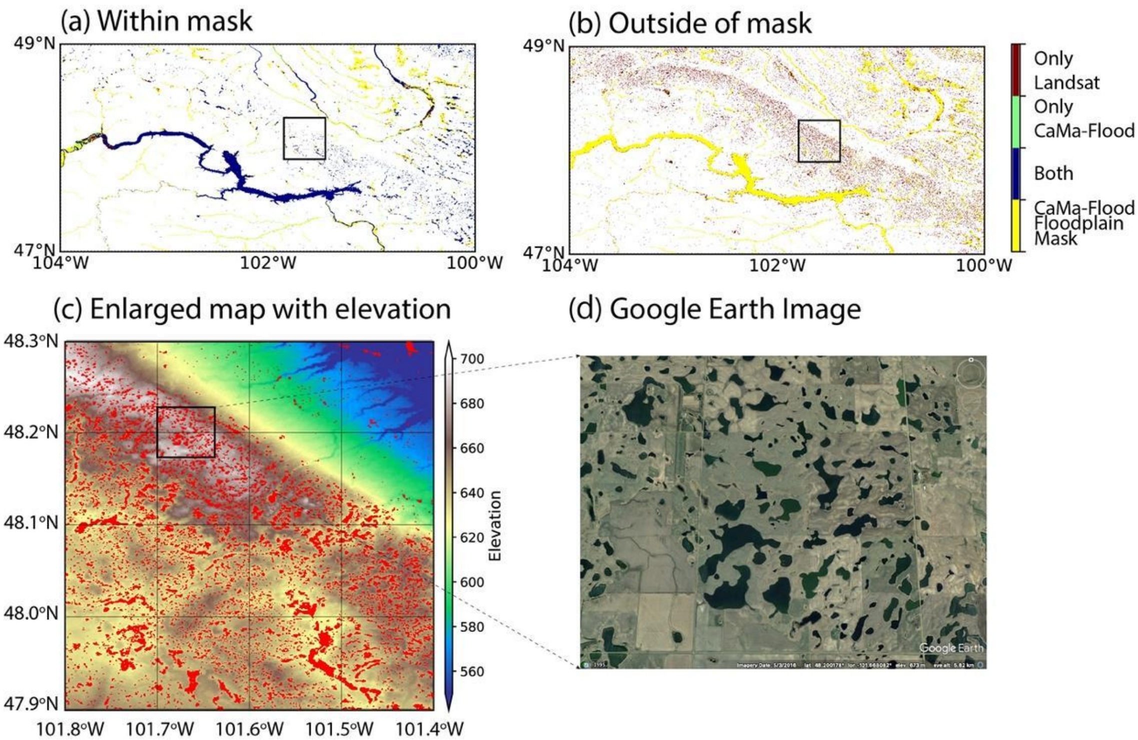

Figure 5. Water surface estimates from CaMa-Flood and Landsat for the source of the Missouri River: (a) comparison within the CaMa-Flood floodplain mask, and (b) comparison outside the floodplain mask. (c) Landsat water mask (occurrence > 10%) over the topography of the target region indicated in a and b. (d) Google Earth image of the target region marked in c.

Because the water surface area on hill slopes is relatively small and not widely distributed throughout the world, the cumulative area difference is not apparent (Figure 4-d) in these regions. Instead, large differences are found in the Canadian Shield, where the coverage ratio is high. An enlarged map (Figure 6) shows that within the floodplain water mask, CaMa-Flood tends to have good ability to delineate large water bodies (lakes) and long rivers. The two methods have consistent results for large lakes (blue in Figure 6-a). However, many local water depressions are not represented in the floodplain mask (red in Figure 6-b). These smaller water bodies are fed by melt water (from glaciers, snow or permafrost; Gilbert and Shaw, 1994; Shilts et al., 1987, Van Huissteden et al., 2011) and likely by shallow groundwater, which are not considered in the forcing of CaMa-Flood. Therefore, the model results are significant underestimations of water surface area compared to the Landsat product. Similar to the kettle lake landform, the distribution of wetlands in the Canadian Shield is scattered (Figure 6-c), and most local water depressions are not modeled in CaMa-Flood.

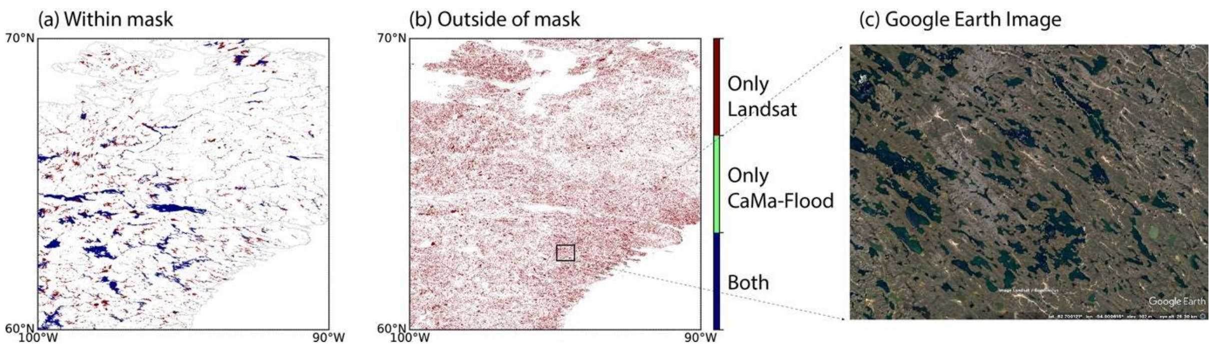

Figure 6. Maps showing the consistency of the water surface prediction between CaMa-Flood and Landsat for the Canadian Shield. (a) Comparison within the CaMa-Flood floodplain mask, and (b) comparison outside the floodplain mask. (c) Google Earth image of the target region marked in c.

Other typical regions where the CaMa-Flood floodplain mask cannot cover the water surface identified by the satellite include coastlines, especially those around mainland China and the Bay of Bengal (see example of the Indus Delta in Figure S3). On the one hand, the spatial resolution of CaMa-Flood is 0.1°, which insufficient to represent the large number of small rivers along the coast. On the other hand, water surface area is caused not only by land-origin water, but also tidal inundation of lowlands, which is not considered with the current settings of CaMa-Flood. These small coastal rivers and lowlands do not belong to CaMa-Flood catchments.

By applying the floodplain mask, the total global water surface area for Landsat decreases from 5.53 Mkm^2 to 4.20 Mkm^2. The underestimation of water surface area is reduced from -1.55 Mkm^2 (-28.1%) to -0.22 Mkm^2 (-5.2%). However, applying the floodplain mask does not alter the spatial patterns of differences between the two results (Figure 7-a) relative to Figure 3-c. Underestimation by CaMa-Flood occurs mainly at high latitudes, while overestimation is found mainly in low-latitude areas around the Equator (Figure 7-c). Although the masking effect is also stronger at high latitudes, the pattern is unaffected, likely because we used a floodplain mask with a relatively modest threshold (see Section 2.4.2) to account for potential errors in runoff forcing and avoid overestimating the predictive ability of CaMa-Flood. We can expect some water outside the CaMa-Flood simulation ability range to be included in the floodplain mask. The following sections will explain the remaining differences between the model and satellite results (e.g., underestimation at high latitudes, overestimation at low latitudes).

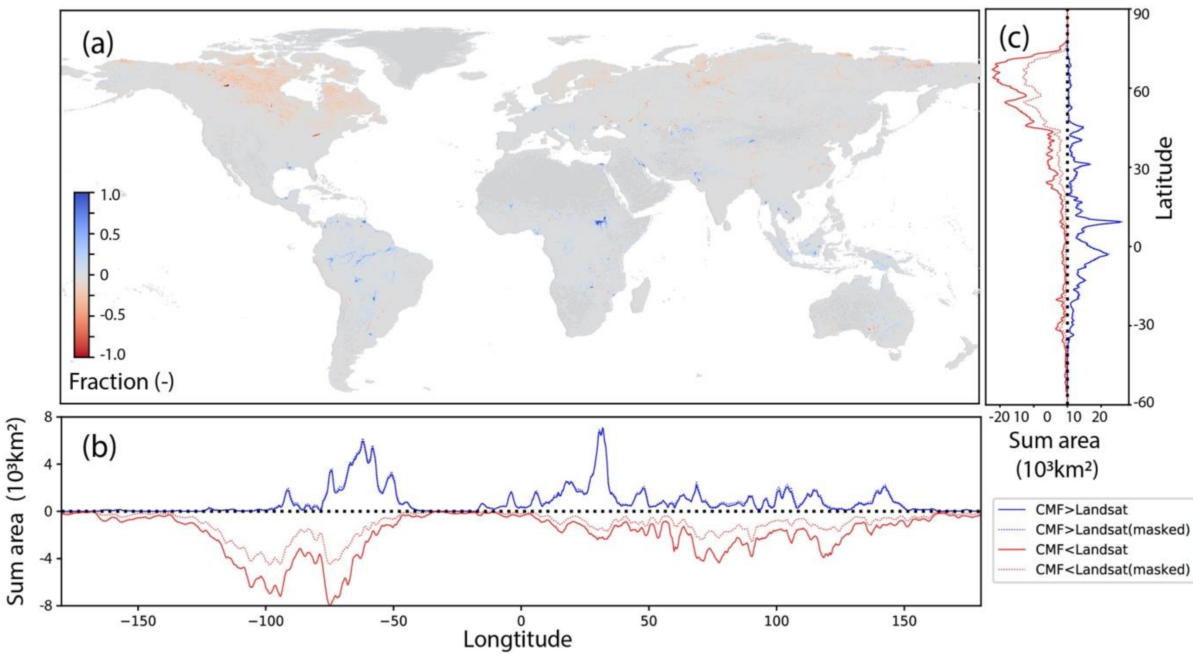

Figure 7. Map of differences between CaMa-Flood and Landsat data after masking (in terms of the fraction of grid area); (b) and (c) show the longitudinal and latitudinal summaries of the differences in area. Overestimation and underestimation are displayed in blue and red, respectively. The solid line represents the results before masking, while the dashed line represents the results after masking.

3.3 Analysis of land cover types

In this section, we discuss the types of water surfaces that can be captured by CaMa-Flood or Landsat by investigating the relationship of water surface with land cover type. The total water surface area corresponding to different land cover types is illustrated in Figure 8, and related statistics are presented in Table 1. The land surface classes were applied to Landsat both before (gray) and after (blue) applying the floodplain mask. Application of the CaMa-Flood floodplain mask reduced the water surface extent obtained from Landsat by 0.36 Mkm^2 (-10.7%) of water bodies (cci_code = 210), and by 0.96 Mkm^2 (-44.8%) of other areas. The impact varied among land cover types (Figure 8-b, Table 1). The impact of masking was significantly greater in areas covered with permanent snow and ice (cci_code = 220, -90.8%). Water bodies present in areas with snow or ice land cover are generally located in local depressions or on high mountains, and therefore are considered to occur outside the floodplain mask. The impact of masking was also strong in areas covered with saline water (cci_code = 170, -65.1%) along the coastline, mainly due to the limitation of CaMa-Flood in representing small coastal rivers.

Many small water bodies are found at high latitudes with needle-leaved tree cover (cci_code = 70, 0.25 Mkm^2), sparse vegetation (cci_code = 150, 0.17 Mkm^2), or lichens and mosses (cci_code = 140, 0.15 Mkm^2) (see distributions in Figure S4). Such water bodies are difficult to simulate with CaMa-Flood, as 48.5%, 69.7%, and 59.2%, respectively, of the water surface area from Landsat is removed with the CaMa-Flood floodplain mask. The reason for this discrepancy was explained using the example of the Canadian Shield in the previous section, as many water bodies within local depressions are excluded from CaMa-Flood. For other land cover types, the effect of masking is less significant. The main reason for this difference could be that small water bodies fed by local runoff are not represented in the model. The CaMa-Flood floodplain mask may also miss areas that are seldom flooded, causing further differences associated with masking. On the other hand, such small water bodies might not be precisely represented at the original resolution of the CCI (300 m).

In terms of the differences between CaMa-Flood and Landsat after masking, excluding water bodies (cci_code = 210) and the aforementioned land cover types (cci_code = 70, 140 and 150) at high latitudes, CaMa-Flood results were higher than Landsat results. The regions with the largest differences included forest-related regions (cci_code = 50), with an overestimation of 0.25 Mkm^2 (441.6%) in CaMa-Flood, and cropland-related regions (cci_code = 10 and 20, +0.09/0.08 Mkm^2), with an overestimation ratio greater than 66%. For regions with short vegetation or wetlands, the modeled water surface in CaMa-Flood was generally larger than that from Landsat, except in regions concentrated at high latitudes. However, the reasons underlying the differences between CaMa-Flood and Landsat differed among land cover types. These reasons will be discussed in the following sections.

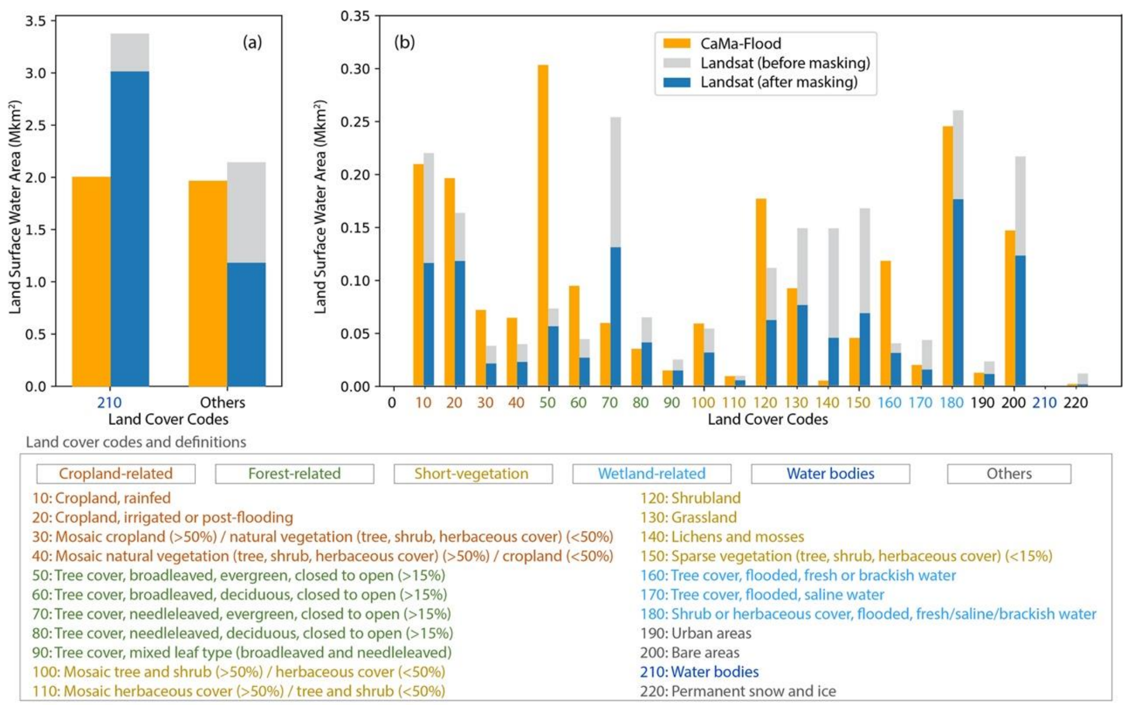

Figure 8. Comparison of LSWA among groups of land cover types. All land cover types other than water bodies (cci_code = 210) are grouped in the type "others" in (a). The orange bars represent the results of CaMa-Flood. The gray bars represent Landsat observations before application of the CaMa-Flood floodplain mask, and the blue section represents the results of Landsat after applying the floodplain mask. A list of cci_code values and definitions of the land cover types is attached (ESA, 2017). All land cover types can be categorized as cropland-related land cover types, forest-related land cover types, short vegetation, wetland-related land cover types and others including water bodies (shown in different colors in the list and figure).

Table 1. Comparisons between LSWA estimates based on CaMa-Flood and those derived from Landsat data (unit: Mkm2). The areas for Landsat before and after application of the floodplain mask are both shown. Areas where the water area is larger than 0.1 Mkm2 are shown in bold. The colors represent different land cover type categories, as defined in Figure 8. C0: CCI land cover map codes; C1: water surface area in CaMa-Flood; C2 (C3): water surface area in Landsat before (after) applying floodplain mask; C4: the change ratio of the Landsat water surface with application of the floodplain mask (C4 = (C3-C2)/C2*100); C5: water surface area difference between CaMa-Flood and Landsat after application of the floodplain mask (C5 = C1-C3); C6: difference ratio between CaMa-Flood and Landsat (C6 = C5/C3*100).

CCI code	CaMa-Flood	Landsat (before masking)	Landsat (after masking)	Diff. due to masking (%)	Diff. (Mkm^2)	Diff. (%)
C0	C1	C2	C3	C4	C5	C6
10	**0.21**	**0.22**	0.12	-47.3	0.09	80.7
20	**0.20**	**0.16**	0.12	-27.9	0.08	66.6
30	0.07	0.04	0.02	-44.7	0.05	244.4
40	0.06	0.04	0.02	-43.1	0.04	187.7
50	**0.30**	0.07	0.06	-23.1	0.25	441.6
60	0.09	0.04	0.03	-40.2	0.07	259.2
70	0.06	**0.25**	0.13	-48.5	-0.07	-54.8
80	0.03	0.06	0.04	-36.9	-0.01	-14.6
90	0.01	0.02	0.01	-42.0	0.00	0.5
100	0.06	0.05	0.03	-42.2	0.03	88.4
110	0.01	0.01	0.00	-46.1	0.00	81.9
120	**0.18**	**0.11**	0.06	-44.4	0.11	186.0
130	0.09	**0.15**	0.08	-48.9	0.02	20.9
140	0.00	**0.15**	0.04	-69.7	-0.04	-89.6
150	0.04	**0.17**	0.07	-59.2	-0.02	-34.1
160	**0.12**	0.04	0.03	-22.9	0.09	283.3
170	0.02	0.04	0.01	-65.1	0.00	29.5
180	**0.24**	**0.26**	0.18	-32.3	0.07	39.2
190	0.01	0.02	0.01	-52.6	0.00	11.6
200	**0.15**	**0.22**	0.12	-43.3	0.02	19.4
210	**2.01**	**3.38**	3.02	-10.7	-1.01	-33.4
220	0.00	0.01	0.00	-90.8	0.00	30.9
Summary	3.98	5.53	4.20	-24.0	-0.22	-5.3

3.3.1 Forest-related regions

Optical sensors have difficulty detecting surface water when clouds or vegetation are present. Invalid data collected on cloudy days were removed when calculating the water occurrence based on Landsat. However, the impact of vegetation was not eliminated (Pekel et al., 2016). A typical region affected by thick vegetation is the Amazon River Basin, where the tree density is unusually high (Figure 9-a). In this case, Landsat is able to detect water only along the main channels where tree density is relatively low (Figure 9-c). In contrast, CaMa-Flood can simulate floodplain water along the main channel (Figure 9-b), even in regions with thick vegetation. This improvement is related to the tree bias removal in the MERIT DEM, upon which CaMa-Flood is built (Yamazaki et al., 2017). In terms of the spatial pattern, at tree densities lower than 30%, CaMa-Flood and Landsat have high consistency for water surface estimation (blue in Figure 9-d), while at tree densities greater than 30%, only CaMa-Flood can model the water surface effectively (green in Figure 9-e). An histogram of the summed area (Figure 9-f) shows that when tree density is greater than 60%, the difference in water surface area between Landsat and CaMa-Flood will increase significantly. As noted above, GIEMS is based mainly on microwave observations, and thus can detect water covered with thick vegetation. Because GIEMS has relatively low spatial resolution (0.25°), the water surfaces from CaMa-Flood and Landsat were also aggregated to 0.25° (Figure 9-g,h). Notably, CaMa-Flood values were similar to those in GIEMS, especially along the mainstream channel, whereas Landsat had low values in that map tile. The histogram plot (Figure 9-f) also shows similar values for CaMa-Flood and GIEMS, especially when the tree density is less than 90%. However, as GIEMS cannot detect small water surfaces easily due to its coarse spatial resolution, the water surfaces in smaller tributaries are not well captured when the water surface is less than 10% of the fractional coverage of equal-area grid cells (Figure 9-i, Papa et al., 2010). Such differences are mainly distributed in areas where the river density is very low and vegetation is dense (Figure 9-b). As a result, the total water surface area obtained from GIEMS for tree densities above 95% is only half of the corresponding value from CaMa-Flood (Figure 9-f).

Similar situations occur in Indonesia (Figure S5) and the Congo River Basin (see Figure S4, cci_code = 50 and 60), as CaMa-Flood has higher values than Landsat where the tree density is high. However, the results from CaMa-Flood are much closer to the GIEMS values, which are not affected by vegetation, indicating superior performance of CaMa-Flood compared to Landsat in these areas. CaMa-Flood results were higher than those of GIEMS over regions of very high tree density (>90%), where numerous small, narrow rivers may flow through forests.

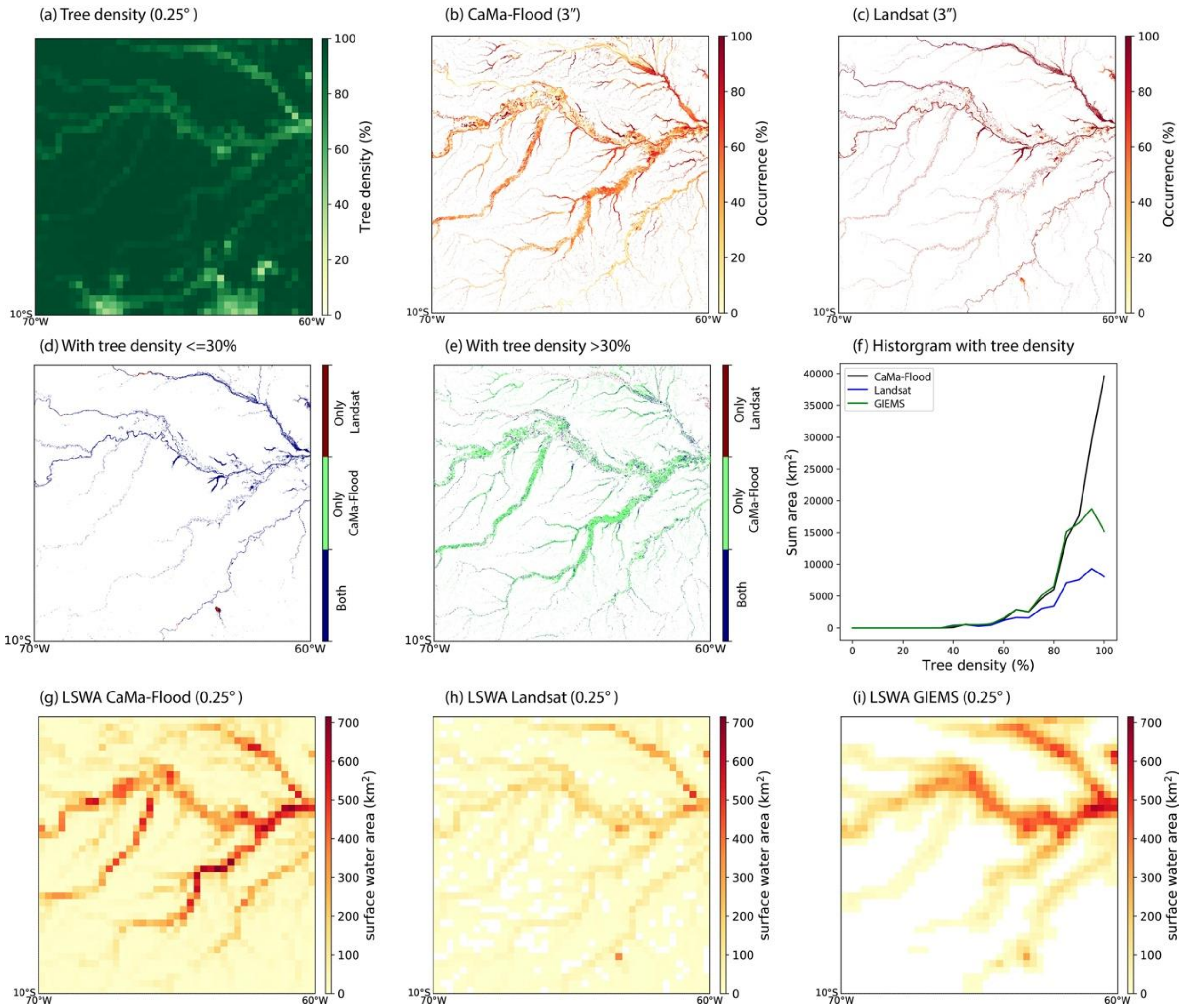

Figure 9. Comparisons of surface water area in CaMa-Flood, Landsat, and GIEMS for the central Amazon River Basin. (a) Tree density; (b) and (c) surface water occurrences in CaMa-Flood and Landsat at 3″; (d) and (e) indicate the consistency of water surface results from CaMa-Flood and Landsat using categories of tree density lower and higher than 30%. (f) Histogram of the water surface areas in CaMa-Flood, Landsat and GIEMS in terms of tree density. (g-i) Spatial maps of water surface area (occurrence>10%) at 0.25° from CaMa-Flood, Landsat and GIEMS, respectively.

3.3.2 Cropland-related regions

In cropland-related areas (cci_code = 10 and 20), CaMa-Flood tends to estimate larger water surface areas than does Landsat (Figure 7-c). Such regions are mainly distributed around river deltas including those of the Nile, Indus, Mekong, Chao Phraya in Thailand, Irrawaddy in Myanmar (shown in Figure 10-a), and lower Mississippi. This difference is likely caused by man-made infrastructure that regulates river flows for human purposes.

Slight differences were found for agricultural and flood defense structures, as canals built for irrigation will alter natural topography and flow paths. In contrast, the construction of flood defenses (e.g., levees) only increases the height of the riverbank, while maintaining the natural river flow path. In low plains where agriculture is dense and developed, irrigation water is transferred by pumping water from rivers, which then flows through canals. These canals, especially the smallest ones, are not represented in the model, and therefore flowing water is assumed to spread over a large area rather than flowing through canals. On the other hand, due to the presence of canals, the flow path is no longer natural. Thus, the flow directions assumed from natural topography are invalid. The continuity of flow is also affected by numerous floodgates. These differences cause inaccuracy in the downscaling of flooding to the high-resolution inundation map. The effect of canals is especially apparent for river deltas in dry climates (e.g., the Nile River, the Tigris & Euphrates Rivers and the Lower Indus River). Levees are built to protect residences and farms from the effects of river floods or tides. In CaMa-Flood, the height of riverbanks is estimated through empirical regression, which does not represent the real conditions of the rivers (see Data and Methods). The presence and height of levees is also neglected, which increases the possibility of flooding in CaMa-Flood estimates.

The Nile Delta has one of the highest population densities in the world. This region includes large urban areas (red color in CCI map, Figure 10-c), with major cities located along the main river channel. Although observational evidence is lacking, there must be levees along the river channel, resulting in the water surface estimated by Landsat aligning perfectly with the main channel (and canals) and not covering the riverbanks (Figure 10-b). In contrast, a high occurrence of water surface is estimated by CaMa-Flood along riverbanks and in flat plains used for agriculture (Figure 10-a, green color in Figure 10-d). As a result, the CaMa-Flood results show larger water areas compared to Landsat and GIEMS, which represent reality better. The constraint of levees is also found in the lower Mississippi River, where houses are built along tributaries (Figure S6), as well as in Baghdad, the capital city of Iraq, where the Tigris River flows through an urban area (Figure S7).

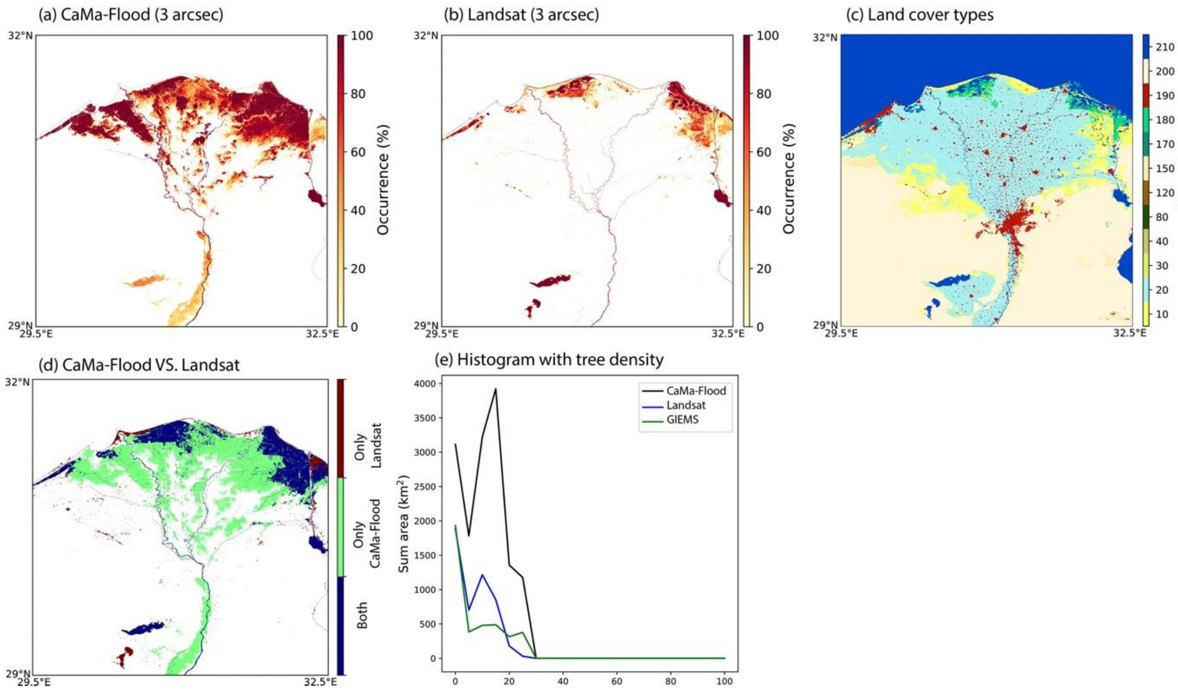

Figure 10. Comparisons of surface water areas based on CaMa-Flood and Landsat, as well as CCI land cover types, for the lower Nile Delta region: (a) and (b) show the water occurrences obtained from CaMa-Flood and Landsat, respectively; (c) is a land cover map; (d) shows the differences in water coverage between CaMa-Flood and Landsat (occurrence >10%); and (e) is a histogram of water surface area against tree density.

Another possible reason for the overestimation of water surface area when using CaMa-Flood relative to Landsat is the lack of water losses (e.g., re-infiltration, evaporation, water consumption) in the routing processes. This impact is stronger for rivers in dry regions (e.g., the Tarim, Tigris and Euphrates Rivers). In the example of the Tarim Basin (Figure 11), water surface areas are found with high occurrence at the foot of the Tian Shan Mountains and around a small tributary to the north of the main Tarim River stem. However, no large water surface is detected in the Landsat data. In this area, a large proportion of water is extracted for irrigation. Due to the local soil properties and high rate of potential evaporation, the amount of water remaining in some rivers will be much less than that calculated by CaMa-Flood. Therefore, only seasonal rivers (occurrence less than 90%) are identified using Landsat data (Figure 11-f). Discontinuous river flow in the lower Tarim has been reported in the media and documented in the literature (Xu et al., 2008). A similar situation can occur in the Tigris and Euphrates Rivers, as no supplemental water enters the lower river section before it reaches the lower delta. The loss of water to soil or evaporation to the atmosphere leads to a lower occurrence of small inundation areas in reality (Landsat) relative to the results of CaMa-Flood. Neglecting water consumption and evaporation also enhances the overestimation of CaMa-Flood results in the Nile Delta.

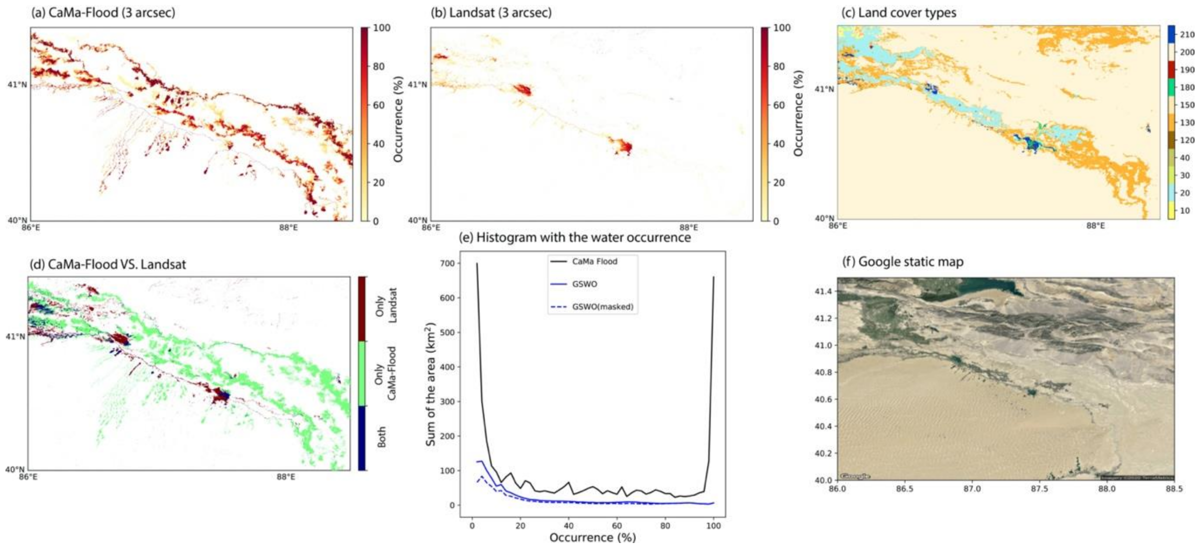

Figure 11. Comparisons of surface water areas based on CaMa-Flood and Landsat, as well as CCI land cover types for the lower Tarim River: (a) and (b) show water occurrences from CaMa-Flood and Landsat, respectively; (c) is the land cover map; (d) shows the differences between water coverage from CaMa-Flood and Landsat (occurrence >10%); and (e) is a histogram of water surface area against occurrence level. (f) Google Earth image of the study region.

Exceptions to this trend, where CaMa-Flood underestimates the irrigated water surface area, are around the lower Ganges River in Bangladesh and the lower Yangtze River in China (see Figure S4, cci_code = 20). These two regions have very high densities of rice paddy fields (Dong and Xiao, 2016). Standing water in the rice-growing season is not represented in CaMa-Flood, while it is highly likely to be detected by Landsat. In these regions, soil moisture is much higher than elsewhere and the water surface area identified in GIEMS is higher than those of both CaMa-Flood and Landsat, as GIEMS may misclassify saturated soil as a water surface (Aires et al., 2018). Water surfaces in other areas of Northeastern China, the Lower Mekong Delta, and the Lower Irrawaddy River Delta are also underestimated by CaMa-Flood due to paddy fields (see Figure S4).

3.3.3 Short vegetation and wetlands

In regions with short vegetation or wetlands, extraction of the real topography or river bathymetry becomes more difficult. Biases in the topography will have strong impacts on water surface estimation in such areas. In particular, CaMa-Flood overestimates the water surface area in Sudd Swamp in the Nile Basin, which is one of the world's largest wetlands (Figure 12-a, shrub or grass, cci_code = 120 and 180). The Sudd Swamp region is extremely flat (blue in Figure 12-c) and the land surface gradient of the floodplain is very difficult to discern in the MERIT DEM, even after error removal. Downscaling to a high resolution (3″) results in overestimation of the inundation extent due to inaccurately flat topography (Figure 12-b). This effect is especially strong on estimates of water extent with occurrences greater than 95% (Figure 12-d). For comparison, the water extent extracted from GIEMS is significantly smaller than that from CaMa-Flood, providing further evidence of overestimation by CaMa-Flood (Figure 12-e). On the other hand, the GIEMS result is larger than that of Landsat, indicating the shortcoming of optical sensors for detecting water surfaces with vegetation cover. The re-infiltration of flooded water into the ground and evaporation are secondary reasons for the overestimation of CaMa-Flood, as these natural processes are not considered in the model. Similar regions can be observed in Figure S4 with cci_code values of 120 and 180 in the Pantanal in Brazil, Niger Inland Delta, and other areas.

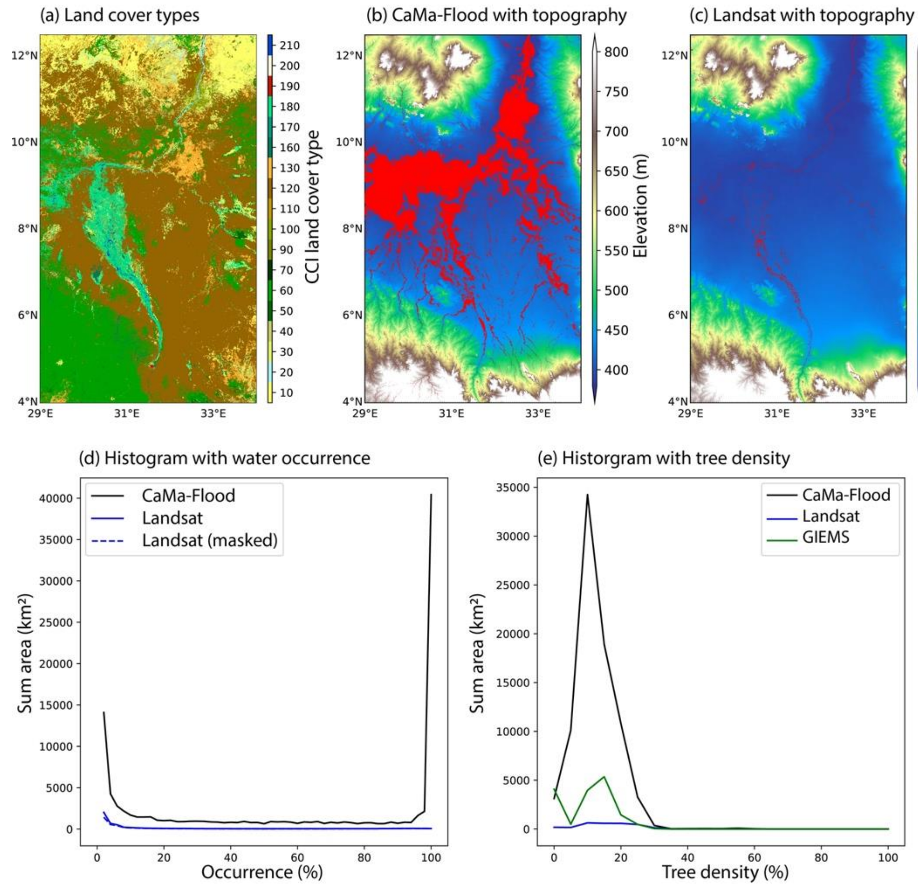

Figure 12. Comparisons of surface water area estimates from CaMa-Flood, Landsat and GIEMS for the Sudd Swamp. (a) Land cover map; (b) and (c) show inundation maps (occurrence > 10%) from CaMa-Flood and Landsat at 3″ overlaid with the topographic map; (d) and (e) are histograms of water surface area against water occurrence and tree density, respectively.

3.4 Statistical permanent water mask

In the analysis described above, underestimation of water surface area by CaMa-Flood was as high as 1.01 Mkm^2 (Figure 8, Table 1) in areas covered with water bodies (cci_code = 210), which made the largest contribution to the difference between model and satellite results. However, because the CCI land cover map was originally at 300-m resolution and was interpolated to 3″ and because processing of CCI land cover classifications was based on a combination of observations, surveys and mathematical programs, the locations of water bodies were not precise (ESA, 2017). Therefore, in this section, the permanent water mask derived from the Landsat occurrence product was applied to the CaMa-Flood results to investigate the ability of CaMa-Flood to estimate water surface areas within and outside the water mask.

Within the permanent water mask, CaMa-Flood underestimates the water surface area by -1.03 Mkm^2 in total for the globe (Figure 13-a, Table 2). The underestimates are mainly concentrated at high latitudes (e.g., the Canadian Shield, Lake Erie, the Lower Ob River Basin and two rivers in eastern Europe). Ignorance of water inputs to local depressions, which are treated as floodwaters and routed along rivers, may be the reason for this underestimation. For regions outside the permanent water mask (Figure 13-b), the spatial pattern of regions with overestimated CaMa-Flood values does not change after application of the permanent water mask. The underestimation by CaMa-Flood almost disappears with this mask, especially at high latitudes, indicating that underestimation by CaMa-Flood is primarily occurring within the permanent water mask. Underestimates obtained outside the permanent water mask are caused by rice paddy fields, which are identified as seasonal water areas in the Landsat product.

As the permanent water extent is obtained from Landsat, we can modify CaMa-Flood in post-processing to compensate for the limitation of CaMa-Flood in estimating permanent water surfaces under certain conditions. If all places previously identified as permanent water in the Landsat data are marked as water surfaces in CaMa-Flood, the total water surface from CaMa-Flood increases to 5.57 Mkm^2, which has very little deviation from the Landsat result obtained without applying the floodplain mask (5.53 Mkm^2, Table 2; because permanent water is sometimes outside the floodplain mask, we used the Landsat water extent without the floodplain mask for this comparison). Furthermore, the difference between the model and satellite results decreases to only 0.04 Mkm^2.

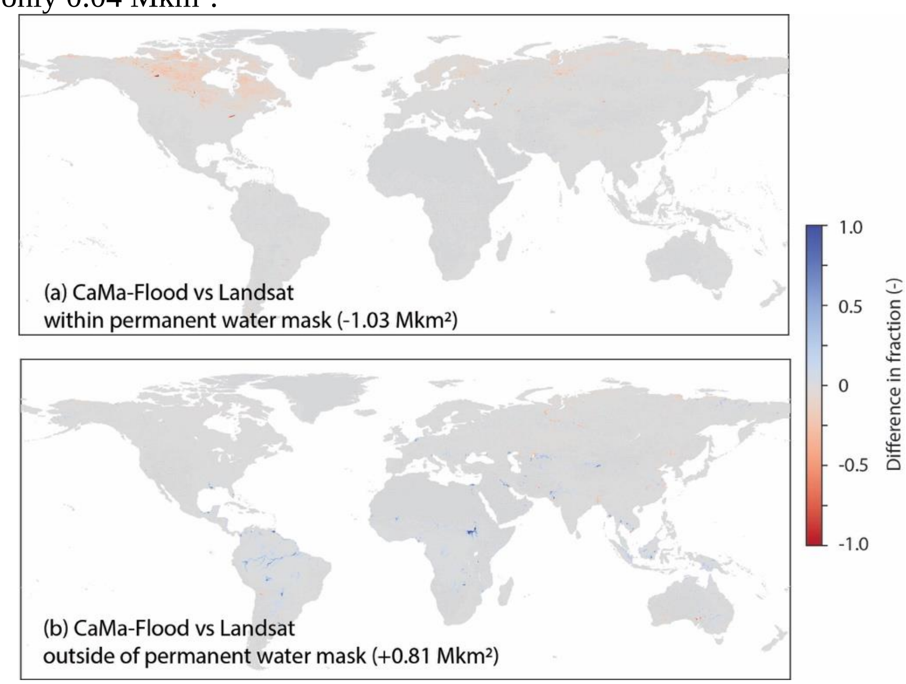

Figure 13. Difference in LSWA based on CaMa-Flood and Landsat within and outside the permanent water mask defined from Landsat data. The unit is area as a fraction of the grid size.

Table 2. Total water surface areas under different conditions (unit: Mkm2).

	CaMa-Flood	Landsat	Var
Original result	3.98	5.53	-1.55
With CaMa-Flood floodplain mask	3.98	4.20	-0.22
Within permanent water mask	2.21	3.24	-1.03
Outside permanent water mask	1.77	0.96	0.81
With permanent water in CaMa-Flood	5.57	5.53	0.04

In summary, we obtained the total water surface area based on water occurrence greater than 10% under various conditions. The original water extents were 3.98 and 5.53 Mkm^2 based on CaMa-Flood and Landsat, respectively, a difference of -1.55 Mkm^2. Within the CaMa-Flood floodplain mask, Landsat identified a water extent of 4.20 Mkm^2. The underestimation by CaMa-Flood compared to Landsat was mainly within the permanent water mask at high northern latitudes (2.21 vs. 3.24 Mkm^2), while overestimation by CaMa-Flood was distributed in tropical regions and croplands within river deltas. Applying the Landsat permanent water mask to the CaMa-Flood increased the CaMa-Flood result from 3.98 Mkm^2 to 5.57 Mkm^2, which reduced the difference between model and satellite results to 0.04 Mkm^2.

4 Discussion

In this study, we investigated land water surface areas extracted from model simulations (CaMa-Flood) and compared the results with satellite-derived results, primarily from Landsat data. Due to the limitations of the model processes and assumptions for downscaling of low-resolution model results in high-resolution inundation areas, CaMa-Flood is not able to represent all types of water bodies that exist in the real world. At the same time, the satellite-derived results also have limitations related to the properties of the sensors and land surface conditions. Therefore, when comparing the two types of results, we applied filters to allow for the most reasonable comparison. The agreements and mismatches between the model and satellite were discussed with example regions. The reliability of CaMa-Flood results and adaptions to ensure appropriate comparison of CaMa-Flood with other methods are discussed in this section.

Only LSWA with occurrence estimates greater than 10% were selected for comparisons between the model and satellite methods in this study. This limitation markedly reduces the impact of clouds on the Landsat data, and also focuses the discussion on the types of water surfaces that can be captured by the model and satellite. Investigation of this broad occurrence range helps to control uncertainty due to model inputs and parameters. Meanwhile, 10% is not too close to 0%, where the modeled water extent is more sensitive to the threshold (see Figure 12-d). One limitation of this study is that we did not investigate the water surface at a specific time, as the values in Landsat for each month are not measured simultaneously around the world.

We also did not investigate the temporal variability in water surface as conducted previously (Wu et al., 2019), as the variability in our results is more closely related to the runoff series than to the accuracy of the inundation model. However, now that the long-term water surface area has been evaluated, we have the confidence to investigate temporal variations further and make reliable comparisons using the filtering methods proposed in this study; such analysis will support more detailed discussion of local-scale or time-variant differences between the performance of the model and satellite.

The modeled water extent is based on a few fundamental assumptions and therefore its applicability is limited to certain conditions. The floodplain mask generated using CaMa-Flood results shows the full extent of the area that CaMa-Flood is able to model. Overall, 24% of the water surface identified by Landsat (1.33 Mkm^2) was excluded when the floodplain mask was applied. The excluded area is mainly distributed at high latitudes and in coastal regions, where numerous local depressions and small rivers occur. Ignorance of local runoff into local depressions rather than routed river flow (e.g., glacial meltwater and shallow groundwater in the Canadian Shield, tidal effects along the coastlines) also reduces the coverage of CaMa-Flood. Springs on hill slopes are not well represented in CaMa-Flood due to its limited spatial resolution and the invalid assumption of a flat-water surface used for downscaling. To overcome the shortcomings of CaMa-Flood in modeling those small water bodies, the model's spatial resolution must be upgraded to represent more rivers. Currently, CaMa-Flood has a resolution of up to 1′ for routing, but this requires a dramatic increase in computational resources, as increasing the number of unit catchments requires shortening the optimal time step for the Courant–Friedrichs–Lewy (CFL) condition. Due to its computational expense, such an improvement can be applied only to specific regions, rather than globally.

CaMa-Flood provides larger water extent estimates compared to Landsat data in forest-related regions (e.g., the Amazon River), approximating the estimates of GIEMS. This difference indicates the advantage of this model compared to Landsat in areas with obstructions caused by vegetation or clouds. CaMa-Flood overestimates the water surface in cropland-related areas, as human water infrastructure (e.g., levees, canals, dikes) is not yet represented in the model. Moreover, because water consumption from the systems for various uses (especially agriculture) is not considered, river discharge can be overestimated in CaMa-Flood compared to reality, which leads to a larger modeled water extent. This impact is cumulative until the end of the river (delta). Ignorance of these natural water dynamics in the routing process also leads to overestimation of river flow in the model in lower delta areas (Dadson et al., 2010; Zhan et al., 2019). To prove this assumption, validation against discharge observations can be employed. In addition to overestimations, ignorance of human interventions can also lead to underestimations by CaMa-Flood in areas comprised of rice paddy fields where standing water can be detected by satellites but is not modeled. The extent of water in reservoirs formed by dams is not considered and will also lead to underestimation, although the impact is negligible at the global scale. Therefore, new modules for natural processes (e.g., re-infiltration, evaporation), water infrastructure (e.g., levees, dikes) and human activities (e.g., irrigation, dam regulation) should be developed to better represent inundation processes.

Based on our analysis, most of the LSWA differences between the model and satellite show distinct spatial features and can be readily explained based on globally consistent reasons. However, some locally varying conditions also affected the results. First, we estimated only the water surface driven by a single runoff input (see Section 2.2.1). The biases in atmospheric forcing (i.e., WFDEI) and the Land Surface Model (i.e., HTESSEL) differ spatially among river basins (Pappenberger et al., 2010). Such biases are propagated to the LSWA estimates. Second, the channel parameters in CaMa-Flood (e.g., river width, river depth) are estimated through global regression with the estimated mean discharge with globally uniform roughness (Yamazaki et al., 2011, 2013). Thus, the bias from runoff generation is again propagated to estimates of the river channel parameters. On the other hand, the river channel parameters are affected by the type of material comprising the riverbed and riverbank, which varies significantly among river sectors (Dunne and Jerolmack, 2020). As these locally varying conditions are difficult to measure or correct for, we suggest the use of ensembles with multiple runoff inputs or parameter settings to evaluate the sensitivity of LSWA results and to possibly identify further globally consistent features.

Given the limitations of the downscaling process and model modules, the estimated water extent shows deviations from the Landsat results, especially at high latitudes in the Canadian Shield region. Although Landsat has its own limitations, especially related to water, valuable information can be obtained from the Landsat-derived results (e.g., permanent water areas). Within the permanent water mask identified based on Landsat data (occurrence = 100%), the CaMa-Flood estimate is 1.03 Mkm^2 smaller than the value from Landsat (3.24 Mkm^2). However, because Landsat determines permanent water with high confidence, we can add a post-processing step to CaMa-Flood whereby the water occurrence is modified to 100% for pixels identified as permanent water based on Landsat. In this case, the total water surface area estimated by CaMa-Flood increases to 5.57 Mkm^2, showing very little deviation (0.04 Mkm^2) from Landsat results (5.53 Mkm^2). Although post-processing does not change the core structures or parameters of CaMa-Flood, this solution is an efficient way to obtain a reasonable result for total water surface area. Additional validation using available river discharge data and model calibration against observations is recommended for regional studies. Data assimilation using *in situ* or satellite-derived observations of water surface area would also be useful for improving the ability of CaMa-Flood to estimate water surface extent (Bates, 2012; Ogilvie et al., 2018; Schumann et al., 2009).

Although the water surface area estimated using CaMa-Flood deviates from that of Landsat, CaMa-Flood offers great advantages over satellite results related to the following aspects. CaMa-Flood is flexible in its temporal scale and can provide hourly estimates if hourly forcing input data are available. This high temporal resolution is vitally important for evaluating rapid changes in water level or flood extent during flood events. However, due to its long revisit time (16 days), Landsat has difficulty capturing rapid changes. MODIS can provide daily results, but its spatial resolution is limited to 500 m, which is too large for flood estimates in normal rivers. MODIS is also significantly limited during floods with continuous rainfall due to widespread cloud cover. Moreover, when driven by runoff inputs corresponding to different scenarios, CaMa-Flood can be used to evaluate the impacts of various factors on water surface area or flood extent. For example, the effects of water consumption (e.g., agricultural usage) on the water surface or the individual contributions of climate variables (e.g., temperature or precipitation) to changes in the water surface could be explored in future studies. Models enable the projection of future water surfaces, which will be useful for evaluating future changes in flood exposure under various climate change scenarios (Hirabayashi et al., 2013). Such studies will be immensely helpful for evaluating the sustainability of water resources against the background of global warming.

5 Conclusions

In this study, we estimated global land surface water area using a global hydrodynamic model (CaMa-Flood). The estimates of water extent exhibited good agreement in spatial patterns with Landsat-derived results. However, due to the limitation of the model's original spatial resolution (0.1°), small depressions away from main river channels and small coastal rivers within a unit catchment are not represented due to the CaMa-Flood model's physical assumptions. This results in underestimation of water surface area in CaMa-Flood compared to Landsat, especially at high latitudes (e.g., Canadian Shield) and for kettle landforms (e.g., the Missouri Plateau) where a cold climate dominates and in coastal areas where many small rivers are present. Water surfaces in irrigated areas (e.g., delta regions and irrigated districts) are generally overestimated due to ignorance of some natural processes (e.g., re-infiltration, evaporation) and human water regulation (e.g., canals, levees, water consumption) in CaMa-Flood. Ignoring irrigation processes in paddy fields leads to underestimation by CaMa-Flood, as these seasonal water bodies are captured by Landsat. Water bodies covered with thick vegetation (e.g., the Amazon Basin, Indonesia) are better represented in the model, as these water bodies cannot easily be detected using optical satellite sensors due to the opacity of clouds and vegetation.

Our analysis suggests that these globally consistent mismatches between CaMa-Flood and Landsat can be reasonably explained based on the model's physical assumptions (e.g., unit catchment concept, downscaling) or limitations of satellite sensing (e.g., weak ability to detect water under vegetation). Applying additional filtering masks (e.g., CaMa-Flood floodplain mask, land cover map, and permanent water mask) to the two datasets helps to constrain the comparison to an appropriate extent, making it much easier to attribute their differences to specific causes. Uncertainties in the runoff forcing, model parameters and baseline topography are potential reasons for the remaining local-scale differences. In this global study, we show that a global hydrodynamic model can represent the areas of different water types and that appropriate comparisons can be made between models and satellite-derived results. By utilizing the findings of this study (e.g., suggested masks for appropriate comparison), more advanced analyses of global river model simulations (e.g., uncertainty attribution using land water surface extent data) will be possible.

Acknowledgments

This study is supported by KAKENHI 20H02251 and 20K22428 by JSPS, Inoue Research Award by Inoue Foundation for Science, and "LaRC-Flood project" by MS&AD Holdings. We gratefully acknowledge the European Commision Joint Research Centre for providing us with the Landsat data. CaMa-Flood and MERIT DEM can be downloaded from public website http://hydro.iis.u-tokyo.ac.jp/~yamadai/. The GIEMS and outputs from CaMa-Flood are available from the authors upon request.

References

Aires, F., Prigent, C., Fluet-Chouinard, E., Yamazaki, D., Papa, F., & Lehner, B. (2018). Comparison of visible and multi-satellite global inundation datasets at high-spatial resolution. *Remote Sensing of Environment*, *216*(June), 427–441. https://doi.org/10.1016/j.rse.2018.06.015

Balsamo, G., Viterbo, P., Beijaars, A., van den Hurk, B., Hirschi, M., Betts, A. K., & Scipal, K. (2009). A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *Journal of Hydrometeorology*, *10*(3), 623–643. https://doi.org/10.1175/2008JHM1068.1

Bates, P. D., & De Roo, A. P. J. (2000). A simple raster-based model for flood inundation simulation. *Journal of Hydrology*, *236*(1–2), 54–77. https://doi.org/10.1016/S0022-1694(00)00278-X

Bates, P. D. (2012). Integrating remote sensing data with flood inundation models: How far have we got? *Hydrological Processes*, *26*(16), 2515–2521. https://doi.org/10.1002/hyp.9374

Carroll, M. L., Townshend, J. R., DiMiceli, C. M., Noojipady, P., & Sohlberg, R. A. (2009). A new global raster water mask at 250 m resolution. *International Journal of Digital Earth*, *2*(4), 291–308. https://doi.org/10.1080/17538940902951401

Chaubell, M.J., Chan, S.K., Dunbar, R.S., Peng, J., Yueh, S., (2018). SMAP Enhanced L1C Radiometer Half-Orbit 9 km EASE-Grid Brightness Temperatures, Version 2. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/0DGGEWUC6MLY

Dadson, S. J., Ashpole, I., Harris, P., Davies, H. N., Clark, D. B., Blyth, E., & Taylor, C. M. (2010). Wetland inundation dynamics in a model of land surface climate: Evaluation in the Niger inland delta region. *Journal of Geophysical Research Atmospheres*, *115*(23), 1–7. https://doi.org/10.1029/2010JD014474

Decharme, B., Alkama, R., Papa, F., Faroux, S., Douville, H., & Prigent, C. (2012). Global off-line evaluation of the ISBA-TRIP flood model. *Climate Dynamics*, *38*(7–8), 1389–1412. https://doi.org/10.1007/s00382-011-1054-9

Decharme, Bertrand, Douville, H., Prigent, C., Papa, F., & Aires, F. (2008). A new river flooding scheme for global climate applications : Off-line evaluation over South America. *Journal of Geophysical Research Atmospheres*, *113*(11), 1–11. https://doi.org/10.1029/2007JD009376

Dong, J., & Xiao, X. (2016). Evolution of regional to global paddy rice mapping methods: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *119*, 214–227. https://doi.org/10.1016/j.isprsjprs.2016.05.010

Dingle, E. H., Creed, M. J., Sinclair, H. D., Gautam, D., Gourmelen, N., Borthwick, A. G. L., & Attal, M. (2020). Dynamic flood topographies in the Terai region of Nepal. *Earth Surface Processes and Landforms*. https://doi.org/10.1002/esp.4953

Dunne, K. B. J., & Jerolmack, D. J. (2020). What sets river width? *Science Advances*, *6*(41), eabc1505. https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.abc1505

ESA. (2017). *Land Cover CCI: Product User Guide*. 1–91.

Fluet-Chouinard, E., Lehner, B., Rebelo, L. M., Papa, F., & Hamilton, S. K. (2015). Development of a global inundation map at high spatial resolution from topographic downscaling of coarse-scale remote sensing data. *Remote Sensing of Environment*, *158*(January 2020), 348–361. https://doi.org/10.1016/j.rse.2014.10.015

Gilbert, R., & Shaw, J. (1994). Inferred subglacial meltwater origin of lakes on the southern border of the Canadian Shield. *Canadian Journal of Earth Sciences*, *31*(11), 1630–1637. https://doi.org/10.1139/e94-144

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., & Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., & Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change*, *3*(9), 816–821. https://doi.org/10.1038/nclimate1911

Horritt, M. S. (2000). Calibration of a two-dimensional finite element flood flow model using satellite radar imagery. *Water Resources Research*, *36*(11), 3279–3291. https://doi.org/10.1029/2000WR900206

Horritt, M. S. (2006). A methodology for the validation of uncertain flood inundation models. *Journal of Hydrology*, *326*(1–4), 153–165. https://doi.org/10.1016/j.jhydrol.2005.10.027

Huang, C., Chen, Y., Zhang, S., & Wu, J. (2018). Detecting, Extracting, and Monitoring Surface Water From Space Using Optical Sensors: A Review. *Reviews of Geophysics*, *56*(2), 333–360. https://doi.org/10.1029/2018RG000598

Ji, L., Gong, P., Wang, J., Shi, J., & Zhu, Z. (2018). Construction of the 500-m Resolution Daily Global Surface Water Change Database (2001–2016). *Water Resources Research*, *54*(12), 10,270-10,292. https://doi.org/10.1029/2018WR023060

Karpatne, A., Khandelwal, A., Chen, X., Mithal, V., Faghmous, J., & Kumar, V. (2016). *Global monitoring of inland water dynamics: State-of-the-art, chal- lenges, and opportunities.* Computational sustainability.

Khan, S. I., Hong, Y., Wang, J., Yilmaz, K. K., Gourley, J. J., Adler, R. F., Brakenridge, G. R., Policelli, F., Habib, S., & Irwin, D. (2011). Satellite remote sensing and hydrologic modeling for flood inundation mapping in lake

Victoria basin: Implications for hydrologic prediction in ungauged basins. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(1 PART 1), 85–95. https://doi.org/10.1109/TGRS.2010.2057513

Lai, X., Liang, Q., Yesou, H., & Daillet, S. (2014). Variational assimilation of remotely sensed flood extents using a 2-D flood model. *Hydrology and Earth System Sciences*, *18*(11), 4325–4339. https://doi.org/10.5194/hess-18-4325-2014

Lewin, J., & Hughes, D. (1980). Welsh floodplain studies. II. Application of a qualitative inundation model. *Journal of Hydrology*, *46*(1–2), 35–49. https://doi.org/10.1016/0022-1694(80)90034-7

Lamarche, C., Santoro, M., Bontemps, S., d'Andrimont, R., Radoux, J., Giustarini, L., Brockmann, C., Wevers, J., Defourny, P., & Arino, O. (2017). Compilation and validation of SAR and optical data products for a complete and global map of inland/ocean water tailored to the climate modeling community. *Remote Sensing*, *9*(1). https://doi.org/10.3390/rs9010036

Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, *296*(1–4), 1–22. https://doi.org/10.1016/j.jhydrol.2004.03.028

Notti, D., Giordan, D., Caló, F., Pepe, A., Zucca, F., & Galve, J. P. (2018). Potential and limitations of open satellite data for flood mapping. *Remote Sensing*, *10*(11). https://doi.org/10.3390/rs10111673

Ogilvie, A., Belaud, G., Massuel, S., Mulligan, M., Le Goulven, P., Malaterre, P. O., & Calvez, R. (2018). Combining Landsat observations with hydrological modelling for improved surface water monitoring of small lakes. *Journal of Hydrology*, *566*(March), 109–121. https://doi.org/10.1016/j.jhydrol.2018.08.076

Papa, F., Prigent, C., Aires, F., Jimenez, C., Rossow, W. B., & Matthews, E. (2010). Interannual variability of surface water extent at the global scale, 1993-2004. *Journal of Geophysical Research Atmospheres*, *115*(12), 1–17. https://doi.org/10.1029/2009JD012674

Pappenberger, F., Beven, K., Horritt, M., Blazkova, S. (2005). Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations. *Journal of Hydrology*. 302 (1e4), 46e69. https://doi.org/10.1016/j.jhydrol.2004.06.036

Pappenberger, F., Cloke, H. L., Balsamo, G., Ngo-Duc, T., & Oki, T. (2010). Global runoff routing with the hydrological component of the ECMWF NWP system. *International Journal of Climatology*, *30*(14), 2155–2174. https://doi.org/10.1002/joc.2028

Patro, S., Chatterjee, C., Mohanty, S., Singh, R., & Raghuwanshi, N. S. (2009). Flood inundation modeling using MIKE FLOOD and remote sensing data. *Journal of the Indian Society of Remote Sensing*, 37(1), 107-118. https://doi.org/10.1007/s12524-009-0002-1

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

Pham-Duc, B., Prigent, C., Aires, F., & Papa, F. (2017). Comparisons of global terrestrial surface water datasets over 15 years. *Journal of Hydrometeorology*, *18*(4), 993–1007. https://doi.org/10.1175/JHM-D-16-0206.1

Phillips, T. J., & Gleckler, P. J. (2006). Evaluation of continental precipitation in 20th century climate simulations: The utility of multimodel statistics. *Water Resources Research*, *42*(3), 1–10. https://doi.org/10.1029/2005WR004313

Prigent, C., Aires, F., Rossow, W., & Matthews, E. (2001). Joint characterization of vegetation by satellite observations from visible to microwave wavelengths: A sensitivity analysis. *Journal of Geophysical Research Atmospheres*, *106*(D18), 20665–20685. https://doi.org/10.1029/2000JD900801

Prigent, C., Jimenez, C., & Bousquet, P. (2020). Satellite-Derived Global Surface Water Extent and Dynamics Over the Last 25 Years (GIEMS-2). *Journal of Geophysical Research: Atmospheres*, *125*(3), 1–18. https://doi.org/10.1029/2019JD030711

Prigent, C., Papa, F., Aires, F., Rossow, W. B., & Matthews, E. (2007). Global inundation dynamics inferred from multiple satellite observations, 1993-2000. *Journal of Geophysical Research Atmospheres*, *112*(12), 1993–2000. https://doi.org/10.1029/2006JD007847

Qi, S., Brown, D. G., Tian, Q., Jiang, L., Zhao, T., & Bergen, K. M. (2009). Inundation extent and flood frequency mapping using LANDSAT imagery and digital elevation models. *GIScience and Remote Sensing*, *46*(1), 101–127. https://doi.org/10.2747/1548-1603.46.1.101

Raymond, P. A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M., Butman, D., Striegl, R., Mayorga, E., Humborg, C., Kortelainen, P., Dürr, H., Meybeck, M., Ciais, P., & Guth, P. (2013). Global carbon dioxide emissions from inland waters. *Nature*, *503*(7476), 355–359. https://doi.org/10.1038/nature12760

Revilla-Romero, B., Hirpa, F. A., Pozo, J. T. del, Salamon, P., Brakenridge, R., Pappenberger, F., & de Groeve, T. (2015). On the use of global flood forecasts and satellite-derived inundation maps for flood monitoring in data-sparse regions. *Remote Sensing*, *7*(11), 15702–15728. https://doi.org/10.3390/rs71115702

Schumann, G., Bates, P. D., Horritt, M. S., Matgen, P., & Pappenberger, F. (2009). Progress in integration of remote sensing-derived flood extent and stage data and hydraulic models. *Reviews of Geophysics*, *47*(3), 1–5. https://doi.org/10.1029/2008RG000274

Schumann, G., Matgen, P., Hoffmann, L., Hostache, R., Pappenberger, F., & Pfister, L. (2007). Deriving distributed roughness values from satellite radar data for flood inundation modelling. *Journal of Hydrology*, *344*(1–2), 96–111. https://doi.org/10.1016/j.jhydrol.2007.06.024

Schumann, G., & Moller, D. K. (2015). Microwave remote sensing of flood inundation. *Physics and Chemistry of the Earth*, *83–84*, 84–95. https://doi.org/10.1016/j.pce.2015.05.002

Schumann, G., Stampoulis, D., Smith, A. M., Sampson, C. C., Andreadis, K. M., Neal, J. C., & Bates, P. D. (2016). Rethinking flood hazard at the global scale. *Geophysical Research Letters*, *43*(19), 10,249-10,256. https://doi.org/10.1002/2016GL070260

Senyurek, V., Lei, F., Boyd, D., Kurum, M., Gurbuz, A. C., & Moorhead, R. (2020). Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS. *Remote Sensing*, *12*(7). https://doi.org/10.3390/rs12071168

Shilts, W. W., Aylsworth, J. M., Kaszycki, C. A., & Klassen, R. A. (1987). Canadian Shield. *Geomorphic Systems in North America*, *January 1987*, 119–161. https://doi.org/10.1130/dnag-cent-v2.119

Stacke, T., & Hagemann, S. (2012). Development and evaluation of a global dynamical wetlands extent scheme. *Hydrology and Earth System Sciences*, *16*(8), 2915–2933. https://doi.org/10.5194/hess-16-2915-2012

Try, S., Lee, G., Yu, W., Oeurng, C., & Jang, C. (2018). Large-scale flood-inundation modeling in the Mekong River Basin. *Journal of Hydrologic Engineering*, *23*(7), 285–299. https://doi.org/10.1061/(ASCE) HE.1943-5584.0001664

Van Huissteden, J., Berrittella, C., Parmentier, F. J. W., Mi, Y., Maximov, T. C., & Dolman, A. J. (2011). Methane emissions from permafrost thaw lakes limited by lake drainage. *Nature Climate Change*, *1*(2), 119–123. https://doi.org/10.1038/nclimate1101

Vorosmarty, C. J., & Sahagian, D. (2000). Anthropogenic Disturbance of the Terrestrial Water Cycle. *Bioscience*, *50*(9), 753–765. https://doi.org/10.1641/0006-3568(2000)050[0753:ADOTTW]2.0.CO

Williamson, C. E., Saros, J. E., Vincent, W. F., & Smol, J. P. (2009). Lakes and reservoirs as sentinels, integrators, and regulators of climate change. *Limnology and Oceanography*, *54*(6 PART 2), 2273–2282. https://doi.org/10.4319/lo.2009.54.6_part_2.2273

Willner, S. N., Otto, C., & Levermann, A. (2018). Global economic response to river floods. *Nature Climate Change*, *8*(7), 594–598. https://doi.org/10.1038/s41558-018-0173-2

Wilson, M. D., Bates, P., Alsdorf, D., Forsberg, B., Horritt, M., Melack, J., Frappart, F., & Famiglietti, J. (2007). Modeling large-scale inundation of Amazonian seasonally flooded wetlands. *Geophysical Research Letters*, *34*(15). https://doi.org/10.1029/2007GL030156

Winsemius, H. C., Van Beek, L. P. H., Jongman, B., Ward, P. J., & Bouwman, A. (2013). A framework for global river flood risk assessments. *Hydrology and Earth System Sciences*, *17*(5), 1871–1892. https://doi.org/10.5194/hess-17-1871-2013

Wu, H., Kimball, J. S., Zhou, N., Alfieri, L., Luo, L., Du, J., & Huang, Z. (2019). Evaluation of real-time global flood modeling with satellite surface inundation observations from SMAP. *Remote Sensing of Environment*, *233*(August), 111360. https://doi.org/10.1016/j.rse.2019.111360

Xu, H., Ye, M., & Li, J. (2008). The water transfer effects on agricultural development in the lower Tarim River, Xinjiang of China. *Agricultural Water Management*, *95*(1), 59–68. https://doi.org/10.1016/j.agwat.2007.09.004

Yamazaki, D., De Almeida, G. A. M., & Bates, P. D. (2013). Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector-based river network map. *Water Resources Research*, 49(11), 7221–7235. https://doi.org/10.1002/wrcr.20552

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset. *Water Resources Research*, *55*(6), 5053–5073. https://doi.org/10.1029/2019WR024873

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., & Bates, P. D. (2017). A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, *44*(11), 5844–5853. https://doi.org/10.1002/2017GL072874

Yamazaki, D., Kanae, S., Kim, H., & Oki, T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*, *47*(4), 1–21. https://doi.org/10.1029/2010WR009726

Yamazaki, D., Lee, H., Alsdorf, D. E., Dutra, E., Kim, H., Kanae, S., & Oki, T. (2012). Analysis of the water level dynamics simulated by a global river model: A case study in the Amazon River. *Water Resources Research*, *48*(9), 1–15. https://doi.org/10.1029/2012WR011869

Yamazaki, D., Sato, T., Kanae, S., Hirabayashi, Y., & Bates, P. D. (2014). Regional flood dynamics in a bifurcating mega delta simulated in a global river model. *Geophysical Research Letters*, *41*(9), 3127–3135. https://doi.org/10.1002/2014GL059744

Yin, Z., Wang, X. H., Otté, C., Zhou, F., Guimberteau, M., Polcher, J., Peng, S. S., Piao, S. L., Li, L., Bo, Y., Chen, X. L., Zhou, X. D., Kim, H., & Ciais, P. (2020). Improvement of the irrigation scheme in the ORCHIDEE land surface model and impacts of irrigation on regional water budgets over China. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001770. https://doi.org/10.1029/2019ms001770

Zhan, S., Song, C., Wang, J., Sheng, Y., & Quan, J. (2019). A Global Assessment of Terrestrial Evapotranspiration Increase Due to Surface Water Area Change. *Earth's Future*, *7*(3), 266–282. https://doi.org/10.1029/2018EF001066