Remote Sensing of Environment 265 (2021) 112680

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Monitoring high spatiotemporal water dynamics by fusing MODIS, Landsat, water occurrence data and DEM

Xiaodong Li [a,*], Feng Ling [a], Giles M. Foody [b], Doreen S. Boyd [b], Lai Jiang [c], Yihang Zhang [a], Pu Zhou [a,d], Yalan Wang [a,d], Rui Chen [a,d], Yun Du [a]

[a] Key Laboratory for Environment and Disaster Monitoring and Evaluation, Hubei, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan 430071, China
[b] School of Geography, University of Nottingham, University Park, Nottingham NG7 2RD, UK
[c] Hubei Water Resources Research Institute, Hubei Water Resources and Hydropower Science and Technology Promotion Center, Wuhan 430070, China
[d] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Keywords:
Spatiotemporal surface water mapping
Surface water occurrence (SWO)
DEM
Abrupt water change
Superresolution

ABSTRACT

Monitoring the spatiotemporal dynamics of surface water from remote sensing imagery is essential for understanding water's impact on the global ecosystem and climate change. There is often a tradeoff between the spatial and temporal resolutions of imagery acquired from current satellite sensors and as such various spatiotemporal image fusion methods have been explored to circumvent the challenges this situation presents (e.g., STARFM). However, some challenges persist in mapping surface water at the desired fine spatial and temporal resolution. Principally, the spatiotemporal changes of water bodies are often abrupt and controlled by topographic conditions, which are usually unaddressed in current spatiotemporal image fusion methods. This paper proposes the SpatioTemporal Surface Water Mapping (STSWM) method, which aims to predict Landsat-like, 30 m, surface water maps at an 8-day time step (same as the MODIS 8-day composite product) by integrating topographic information into the analysis. In addition to MODIS imagery acquired on the date of map prediction and a pair of MODIS and Landsat images acquired temporally close to the date of prediction, STSWM also uses the surface water occurrence (SWO, which represents the frequency with which water is present in a pixel) and DEM data to provide, respectively, topographic information below and above the water surface. These data are used to translate the coarse spatial resolution water distribution representation observed by MODIS into a 30 m spatial resolution water distribution map. The STSWM was used to generate an 8-day time series surface water maps of 30 m resolution in six inundation regions globally, and was compared with several other state-of-the-art spatiotemporal methods. The stratified random sampling design was used, and unbiased estimators of the accuracies were provided. The results show that STSWM generated the most accurate surface water map in which the spatial details of surface water were well-represented.

1. Introduction

Water is a key land cover type on the Earth's surface, and its spatiotemporal dynamics have major interactions with environmental systems and processes (Holgerson and Raymond, 2016; Vorosmarty et al., 2000). Accurate monitoring of the spatiotemporal changes in the surface water is critical to activities such as the management of lakes, rivers, and coastal shorelines (Pekel et al., 2016). Remote sensing has a key role in monitoring the spatiotemporal dynamics of surface water at a range of scales. Data from systems such as the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) (Dietz et al., 2017), Moderate Resolution Imaging Spectro-radiometer (MODIS) (Pekel et al., 2014), and Landsat sensors (Mueller et al., 2016; Ogilvie et al., 2018; Pekel et al., 2016; Pickens et al., 2020; Tulbure et al., 2016) have been widely used to study surface water.

Although satellite remote sensing has greatly facilitated surface water monitoring in recent years, a major challenge is associated with the common tradeoff between the spatial and temporal resolutions of sensing systems (Belgiu and Stein, 2019; Zhu et al., 2018). Generally,

* Corresponding author.
E-mail address: lixiaodong@whigg.ac.cn (X. Li).

https://doi.org/10.1016/j.rse.2021.112680
Received 4 November 2020; Received in revised form 20 August 2021; Accepted 25 August 2021
Available online 3 September 2021

popular current remote sensing systems typically provide either imagery with a high-temporal and low-spatial resolution (i.e., coarse) or a low-temporal and a high-spatial resolution (unless commercial systems are employed). For example, MODIS provides imagery with a high-temporal (daily) and low-spatial resolution (typically 250–500 m). The latter thus tends to limit the ability to accurately map water bodies, especially in heterogeneous landscapes. In contrast, Landsat sensors can provide imagery with low-temporal (16 days) and high-spatial-resolution (30 m) images. The latter is suitable for capturing the fine detail of the surface water distribution. However, the 16-day repeat rate that is often observed may fail to capture rapid changes in surface water change. For example, the global mean duration of floods is less than ten days (Bates et al., 2014), limiting the utility of Landsat sensor data in studies of flood events. A further limitation of Landsat data is that the probability of obtaining a cloud-free image may be low. Cloud contamination results in series of missing data in the Landsat images, which becomes a big challenge for the Landsat data application in monitoring surface water (Ju and Roy, 2008; Zhu and Helmer, 2018; Zhu and Woodcock, 2014).

In order to map surface water at both fine spatial and temporal resolutions, it is often necessary to integrate the high-temporal low-spatial resolution data (such as MODIS) with low-temporal high-spatial-resolution data (such as Landsat). The Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) (Gao et al., 2006) is one of the first spatiotemporal image fusion methods that deal with the tradeoff in the spatial and temporal resolutions of imagery acquired by sensors such as MODIS and Landsat. STARFM and related algorithms predict a Landsat-like reflectance image at MODIS repetition rates (Li et al., 2020a; Wang and Atkinson, 2018; Wang et al., 2020; Zhu et al., 2010). Many advanced methods have been proposed recently; for instance, the spatiotemporal fusion based on virtual image pair-based is particularly advantageous when the observed image pairs are temporally far (Wang et al., 2020). Spatiotemporal image fusion methods have been explored for mapping surface water in recent years. Specifically, spatiotemporal image fusion was used to generate Landsat-like multi-spectral images, which were then processed using unsupervised, supervised, or object-based classification algorithms to derive a surface water map (Chen et al., 2018; Dao et al., 2019; Heimhuber et al., 2018; Tan et al., 2019; Zhang et al., 2014). Although the high-spatiotemporal-resolution water maps can be generated from the fused Landsat-like images, several limitations persist. In the temporal domain, the spatiotemporal image fusion is more suitable for predicting gradual changes in reflectance change caused by variables such as the bidirectional reflectance distribution function (BRDF) and phenological changes (Gao et al., 2006). The reflectance change caused by a flood is usually abrupt and nonlinear, making predictions from standard spatiotemporal fusion methods inaccurate (Zhu et al., 2018). In the spatial domain, spatiotemporal image fusion does not consider the surface water's spatial distribution in the fusion analysis.

An alternative method to spatiotemporal image fusion that addresses the tradeoff between spatial and temporal resolutions is spatiotemporal superresolution (land cover) mapping (He et al., 2020; Li et al., 2015b; Ling et al., 2011). Spatiotemporal superresolution mapping does not produce Landsat-like reflectance images like STARFM, but directly generates land cover maps at high-spatial-temporal resolutions. Spatiotemporal superresolution mapping has been developed to generate a series of 30 m land cover maps at MODIS temporal resolution (He et al., 2020; Li et al., 2017; Li et al., 2019b; Wang et al., 2016). Present spatiotemporal superresolution mapping methods are designed to map multiple land cover classes and use the spatial dependence principle, which assumes spatially close objects are likely to have the same class label. The spatial dependence principle is suitable for mapping land cover boundaries within the coarse spatial resolution pixel (Atkinson, 2009; Fisher, 1997; Ge et al., 2016; Ge et al., 2009; Zhang et al., 2008), but may be inappropriate for use with small objects such as rivers and roads which have a narrow linear shape and small patch size (Li et al., 2019b; Ling et al., 2019). For example, Lin et al. (2020) showed that the reach-level bankfall river width in the region of the study was smaller than 200 m. Allen and Pavelsky (2015) found 2.39×10^5 km of rivers with widths $\geq$30 m and 1.1×10^5 km of rivers wider than 100 m according to the North American River Width (NARWidth) data set. Yang et al. (2020a) found that the width of more than 80% of the rivers in China was narrower than 350 m. Since the river widths are mostly smaller than the MODIS pixel size, spatiotemporal superresolution mapping using the land cover spatial dependence principle will be inappropriate and this situation will occur most commonly in heterogeneous landscapes that are highly fragmented (Ling et al., 2019; Muad and Foody, 2012a, 2012b).

In a region with heterogeneous surface cover, using coarse-resolution data (e.g., MODIS) will give rise to the mixed pixel problem. Thus the integration of sub-pixel scale surface water distribution information into the image fusion analysis has considerable potential to enhance surface water mapping from coarse spatial resolution systems such as MODIS. Water distribution is strongly related to the prevailing topographic conditions reported to be one of the major contributors to ensuring high flood inundation mapping accuracy (Bates et al., 2003; Cook and Merwade, 2009). Widely used open-access digital elevation models (DEM) such as the Advanced Spaceborne Thermal Emission Radiometer-Global Digital Elevation Model (ASTER GDEM) and Shuttle Radar Topography Mission (SRTM) provide 30 m topographic information that can be integrated into surface water mapping. For instance, Li et al. (2013a) translated the 500 m water fractions produced from MODIS images to a sub-pixel surface water distribution using 30 m SRTM DEM data. The constraint is that the water level stays similar everywhere at the same surface level. Xiao et al. (2018) downscaled 500 m water maps from MODIS data to 30 m resolution by using a moving average filter method and high spatial resolution DEM data to smooth the errors of waterlines in shallow lakes. Optimization algorithms such as the neural networks were used to downscale surface water maps at the coarse spatial resolution to a finer spatial resolution scale using DEM (Huang et al., 2014; Li et al., 2015a; Ling et al., 2008).

Although a DEM can provide topographic information at a fine spatial scale to control the water distribution, the global open-access spaceborne-derived DEMs are limited for depicting water depth distribution and so do not represent bathymetry. Thus, the DEMs produced from radar altimetry do not give any information below the water surface (Berry et al., 2007). Despite the importance of bathymetry, there is often little information available for inland water bodies such as rivers, lakes, and reservoirs in open-source spaceborne-derived DEMs (Armon et al., 2020; Li et al., 2019c; Li et al., 2020b). Bathymetry is typically mapped through an echo sounder (sonar) mounted beneath or over the side of a boat. However, the acquisition of bathymetric information for large lakes, reservoirs, and river networks globally would be a time-consuming and non-trivial task (Getirana et al., 2018).

Since the generation of large-area three-dimensional bathymetry information to aid studies of surface water dynamics analysis is challenging, attempts have been made to explore alternative data sources. For example, the Landsat sensor data archive may be used to generate information on bathymetry. Historic Landsat sensor data have been combined to produce a water occurrence map that essentially extends the terrain downward into the water surface boundary (Armon et al., 2020; Li et al., 2019c). The basic idea is that the larger the water occurrence for a pixel (i.e., the more often it is covered by water) the deeper the water. Thus there is a strong relationship between the water occurrence and the bathymetry for lakes, rivers, and reservoirs. Specifically, Li et al. (2019c) obtained bathymetry by combining Lidar elevation data with time-series water maps classified from historical Landsat imagery to monitoring reservoir dynamics with the reservoir bathymetry contours derived from the water occurrence percentile.

Among the various data sources for information on water occurrence at global (Pekel et al., 2016; Pickens et al., 2020) or national (Mueller et al., 2016; Tulbure et al., 2016; Zou et al., 2018) scale, the open-source Global Surface Water Mapping (GSW) Layers produced by the European

Commission's Joint Research Centre (JRC) was used to provide global bathymetry. The latter contains surface water maps generated from the Landsat 5, 7, and 8 imagery in which each pixel was classified into water/land from 1984 to 2020. The JRC GSW historical surface water maps represent the geographical distribution of surface water globally, and depicts the spatial extent of water from which temporal changes in surface water cover may be inferred. The JRC GSW contains a surface water occurrence (SWO) map of Pekel et al. (2016) produced by summing the times that the pixel is detected as water and dividing it by the number of total valid observations. The SWO map indicates the frequency with which water was present in each pixel. The JRC SWO map has been widely used in the monitoring of global and regional scale lakes (Luo et al., 2019; Yao et al., 2019a), reservoirs (Busker et al., 2019; Li et al., 2020b; Ling et al., 2020), sites of flood inundation (Li et al., 2020b) and surface water fractions (Li et al., 2019a). The potential of the JRC SWO data in providing bathymetry and sub-pixel surface water distribution information within the MODIS pixel is considerable.

In order to map surface water at both fine spatial and temporal resolutions, a new SpatioTemporal Surface Water Mapping (STSWM) method using a MODIS image acquired at the prediction time, a pair of prior/post-dated MODIS and Landsat imagery acquired temporally close to the prediction time, as well as the SWO and DEM data is proposed in this paper. The basic idea of using SWO and DEM is that topographic conditions control the geographical distribution of water, and SWO provides bathymetry information while a DEM provides above-water surface topographic information, respectively. Unlike traditional spatiotemporal image fusion algorithms and spatiotemporal superresolution mapping algorithms, STSWM integrates SWO with DEM to help to predict the surface water spatial distribution. The proposed STSWM is a combination of the MODIS-Landsat fusion method and the DEM and SWO maps. As a result, it is expected STSWM will be more accurate than standard MODIS-Landsat fusion methods in terms of surface water mapping. The STSWM was used to generate temporal continuous surface water maps (at an 8-day time step) of 30 m resolution in six sites globally and was compared with the GSW monthly water history maps and the state-of-the-art spatiotemporal fusion methods.

2. Methods

STSWM aims to generate a 30 m spatial resolution surface water map at time t_p, based on the MODIS image at time t_p, a temporally close pair of MODIS-Landsat imagery at time t_0, the SWO map of Pekel et al. (2016) and the DEM map (such as STRM and ASTER GDEM with a 30 m spatial resolution). Note that t_0 was selected to be temporally close to t_p, and can either pre-date or post-date t_p. The STSWM contains three main steps, as is shown in Fig. 1. The first step is water information extraction from the input data, including unmixing the MODIS images to coarse spatial resolution water fraction images, classifying the Landsat image at t_0 to a surface water map, and combining the SWO and DEM data to a surface water likelihood map. The second step is downscaling coarse spatial resolution water fraction change to 30 m resolution using the coarse fraction water fraction change image and other ancillary data. The third step is modifying the surface water map based on spatial filters.

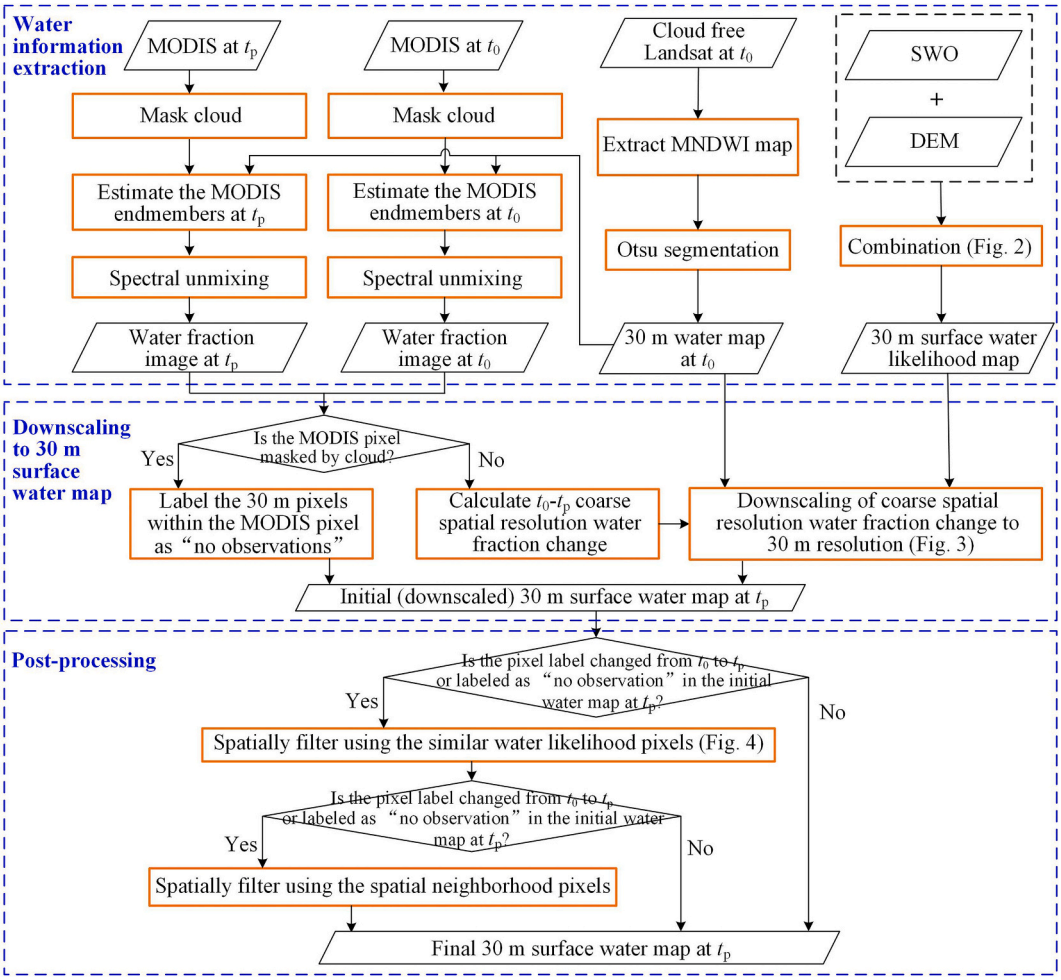

Fig. 1. Flowchart of the STSWM method.

2.1. Water information extraction

2.1.1. Classifying the cloud-free Landsat image at t_0 to produce a surface water map

The cloud-free Landsat image at t_0 is used to generate a surface water map at t_0 based on an automatic segmentation method. First, the Modified Normalized Difference Water Index (MNDWI) map is extracted from the Landsat data to distinguish water and land pixels on its excellent performance (Xu, 2006):

$$\mathrm{MNDWI} = (\rho_{GREEN} - \rho_{SWIR})/(\rho_{GREEN} + \rho_{SWIR}) \quad (1)$$

where ρ_{GREEN} and ρ_{SWIR} are the reflectance in the green and SWIR bands. Then the simple and widely used Otsu method (Otsu, 1979) that maximizes the inter-class variance is used to automatically determine a threshold to separate surface water bodies from the background (Du et al., 2012; Yang et al., 2020b; Yao et al., 2019b).

2.1.2. Unmixing the MODIS images to determine coarse spatial resolution water fractions t_0 and t_p

The MODIS images at t_0 and t_p are unmixed to yield water fraction images for each date. First, the water and land endmembers in each MODIS image are automatically extracted. Considering the presence of different water and land types and the intra-class variability in water endmembers due to different turbidities, a total of n_{water} water endmembers and n_{land} land endmembers are extracted in STSWM.

First, the Landsat pixels at t_0 that have the water labels in the Otsu map are clustered to n_{water} subclasses, and the Landsat pixels at t_0 that have the land labels in the Otsu map are clustered to n_{land} subclasses. In this study, the *k*-means algorithm is adopted as the clustering method that is directly applied to the Landsat multispectral image (the thermal and panchromatic bands were excluded). The number of clusters is set to n_{water} for water and n_{land} for land, respectively. In this way, the 30 m Landsat image at t_0 is clustered to a water map with n_{water} water classes and n_{water} land classes. In STSWM, the optimal number of water class (n_{water}) and land class (n_{land}) used to estimate MODIS endmembers are related to the water turbidity and the number of land cover classes in the study region. n_{water} and n_{land} are set to 2 in this paper for computing efficiency. Larger n_{water} and n_{land} values should be set if the region of analysis is heterogeneous in class composition. Finally, this 30 m water map is spatially degraded to the MODIS scale. Assume s is the scale factor between MODIS and Landsat pixel and s = 16 (namely, each MODIS pixel contains $16 \times 16 = 256$ 30 m Landsat pixels). The fractions of each subclass are calculated by dividing the total number of 30 m pixels that belong to that subclass by 256.

Then, the MODIS endmembers for each subclass at t_p are estimated. According to the linear spectral mixture model, the pixel spectral value in the b^{th} band ($b = 1,\cdots,B$, where B is the number of bands in the MODIS data used) of the m^{th} MODIS pixel at t_p, $r_{m,\,b,\,tp}$, is the weighted sum of all water and land endmembers at t_p within it:

$$r_{m,b,t_p} = \sum_{n_1=1}^{n_{water}} f_{m,t_p}^{water(n_1)} \times e_{b,t_p}^{water(n_1)} + \sum_{n_2=1}^{n_{land}} f_{m,t_p}^{land(n_2)} \times e_{b,t_p}^{land(n_2)} \quad (2)$$

where $f_{m,\,tp}^{water(n_1)}$ and $f_{m,\,tp}^{land(n_2)}$ are the $n_1{}^{\mathrm{th}}$ water and $n_2{}^{\mathrm{th}}$ land fractions in the m^{th} MODIS pixel at t_p, and $e_{b,\,tp}^{water(n_1)}$ and $e_{b,\,tp}^{land(n_2)}$ are the reflectance values on the b^{th} band for the $n_1{}^{\mathrm{th}}$ water and the $n_2{}^{\mathrm{th}}$ land classes at t_p, respectively. The solving of $e_{b,\,tp}^{water(n_1)}$ and $e_{b,\,tp}^{land(n_2)}$ are according to the inversion of the linear equations in (3) using the least-squares best fit solution (Li et al., 2020a; Zhu et al., 2016):

$$\begin{bmatrix} r_{1,b,t_p} \\ \vdots \\ r_{m,b,t_p} \\ \vdots \\ r_{M,b,t_p} \end{bmatrix} = \begin{bmatrix} f_{1,t_p}^{water(1)} & \cdots & f_{1,t_p}^{water(n_{water})} & f_{1,t_p}^{land(1)} & \cdots & f_{1,t_p}^{land(n_{land})} \\ \vdots & & \vdots & \vdots & & \vdots \\ f_{m,t_p}^{water(1)} & \cdots & f_{m,t_p}^{water(n_{water})} & f_{m,t_p}^{land(1)} & \cdots & f_{m,t_p}^{land(n_{land})} \\ \vdots & & \vdots & \vdots & & \vdots \\ f_{M,t_p}^{water(1)} & \cdots & f_{M,t_p}^{water(n_{water})} & f_{M,t_p}^{land(1)} & \cdots & f_{M,t_p}^{land(n_{land})} \end{bmatrix} \begin{bmatrix} e_{b,t_p}^{water(1)} \\ \vdots \\ e_{b,t_p}^{water(n_{water})} \\ e_{b,t_p}^{land(1)} \\ \vdots \\ e_{b,t_p}^{land(n_{land})} \end{bmatrix} \quad (3)$$

Both the class fractions are endmembers in Eq. (3) are unknown variables. In order to solve Eq. (3), M ($M > n_{\mathrm{water}} + n_{\mathrm{water}}$) MODIS pixels which are unchanged in class fractions are selected, so that the class fractions of the M MODIS pixels at t_p can be substituted with those from the degraded clustering images at t_0. Specifically, K pixels with the highest water fractions and land fractions at t_p are selected to avoid the collinearity problem. To reudce the impact of land cover change, the K pixels with Δr_b (i.e., the change in the MODIS pixels reflectance values on the b^{th} band from t_0 to t_p) outside of the range of 0.0–0.5 quantiles are excluded (Li et al., 2020a; Zhu et al., 2016).

With the extracted water and land endmembers at t_p, the fully constrained linear mixing model is used to unmix the t_p time MODIS image for generating the water fraction image at t_p (Keshava and Mustard, 2002). Note that the evaluation of water fraction changes from t_0 to t_p by comparing the water fraction images from the degraded water maps at t_0 and the unmixed MODIS image at t_p is not proper, because they are acquired from different data sources and using different methods. Similar to the estimation of water and land endmembers at t_p, the water and land endmembers at t_0 are estimated using the same inversion process by substituting the reflectance image at t_p with that at t_0. The estimated water and land endmembers at t_0 are then used to unmixed the t_0 time MODIS image using the fully constrained linear mixing model. The unmixed water fractions of the n_{water} subclasses at t_0 and t_p were summed up to represent the final water fractions at t_0 and t_p, respectively. The flowchart of MODIS endmember extraction and unmixing approaches are referred to as Fig. S1 in the supplementary data.

2.1.3. 30 m resolution surface water likelihood map

The SWO and DEM maps are used to provide bathymetry and above-water topographic information in STSWM. The SWO map indicates the surface water occurrence (SWO) or frequency, as a percentile value, with which water was present in the historical Landsat sensor data. Different SWO values indicate permanent water pixels (SWO = 100), permanent land pixels (SWO = 0), and seasonal water pixels (SWO > 0 and SWO < 100) (Pekel et al., 2016); note the expression seasonal does

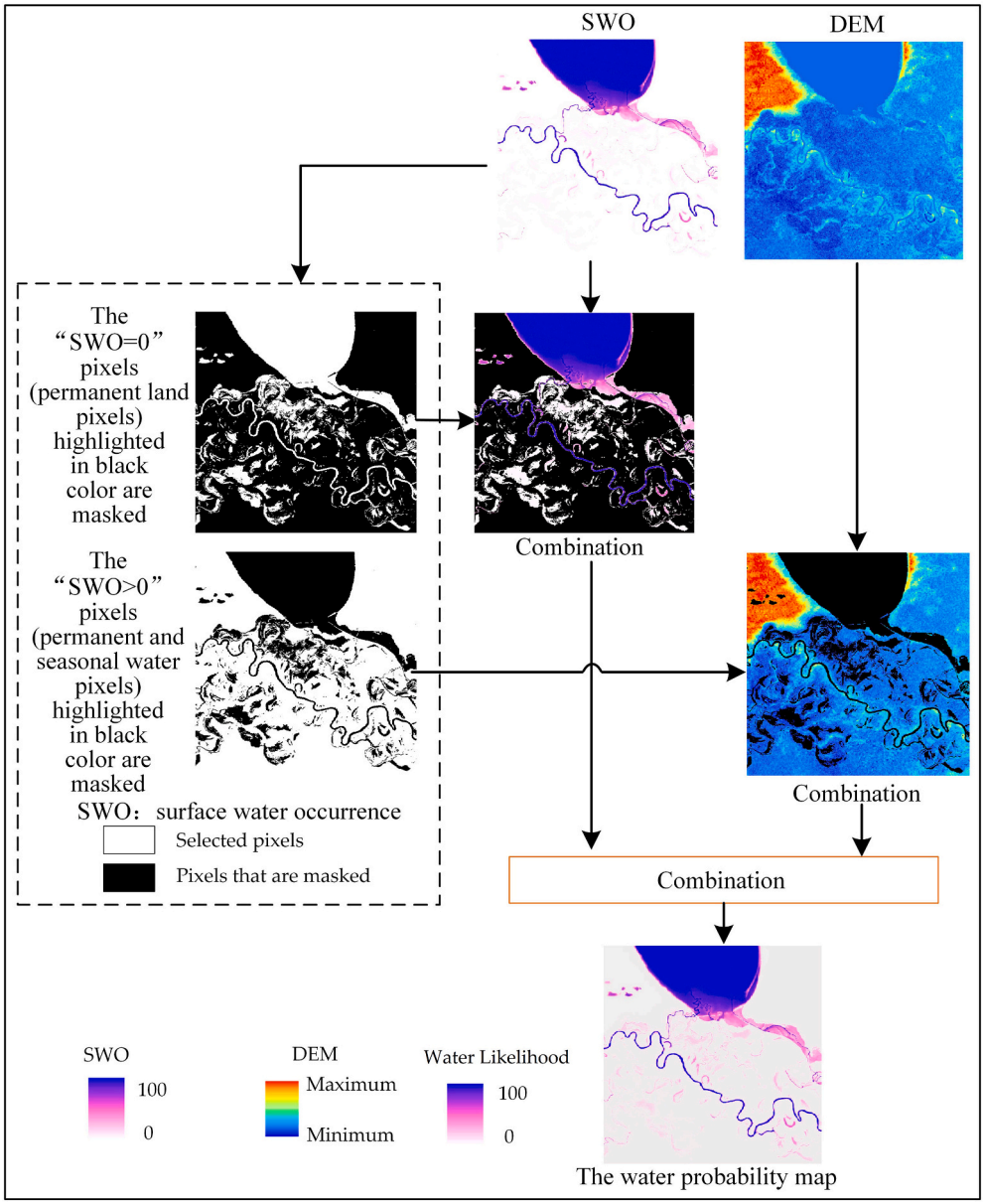

Fig. 2. The flowchart of producing the surface water likelihood map.

not relate to a climatic season but simply indicates that water cover is not constant. The DEM map indicates the elevation in each pixel. Since water flows downwards due to gravity, the probability of a pixel being water-covered is greatest the higher the SWO value and the lower the elevation (Armon et al., 2020; Li et al., 2019c).

The SWO and DEM maps in a target inland/coastal area region are combined to produce the 30 m water likelihood map in Fig. 2. For the permanent and seasonal water pixels in the SWO map (the SWO > 0 pixels are masked in Fig. 2), the water likelihood index is set equal to the SWO percentile value in Eq. (4). For permanent land pixels in the SWO map (the SWO = 0 pixels are masked in Fig. 2), the water likelihood index is set inversely proportional to the elevation value in the DEM map. The water likelihood index is normalized in Eq. (4) so that the maximal water likelihood index derived from the DEM is not larger than the smallest water likelihood index of permanent and seasonal water pixels in the SWO map in the region of a target area as:

$$L_{water} = \begin{cases} SWO & if\ SWO > 0 \\ (E_{max} - E)/(E_{max} - E_{min}) & if\ SWO = 0 \end{cases} \quad (4)$$

where L_{water} is the final water likelihood map, and E is the elevation value in the DEM map in the region of a target area, E_{max} and E_{min} are the maximal and minimal DEM values of the permanent land pixels in the SWO map in the region of the area to be analyzed, and the value of $(E_{max} - E)/(E_{max} - E_{min})$ ranges from 0 to 1.

2.2. Downscaling of the surface water data to an initial 30 m resolution surface water map at t_p

In this section, the surface water fraction data is downscaled to generate an initial 30 m surface water map. Before the downscaling, the cloud and shadow pixels in the 8-day composite MOD09A1 images are first masked according to the MODIS quality assessment data. Suppose cloud or shadow is detected in a MODIS pixel. In that case, the 30 m pixels within the MODIS pixel are labels as 'no observations' in the initial 30 m surface water map. The downscaling process is only used for the MODIS pixels that are not masked by cloud and shadows at t_0 and t_p.

Suppose a MODIS pixel is not masked by cloud at either t_0 to t_p according to the MODIS quality assessment data. In that case, the coarse

spatial resolution water fraction change in the MODIS pixel from t_0 to t_p is calculated by subtracting the coarse spatial resolution water fraction map at t_0 from the coarse spatial resolution water fraction map at t_p, which is then downscaled using the ancillary water likelihood map and the water map at t_0. The details of the downscaling approaches in the m^{th} MODIS pixel are introduced below.

Case1: water fraction increases in the m^{th} MODIS pixel. STSWM assumes, in the m^{th} MODIS pixel, the 30 m pixels labeled as belonging to the water class at t_0 are unchanged in the class label at t_p, and the 30 m pixels labeled as belonging to the land class at t_0 could have changed membership to the water class. STSWM uses a step-wise allocation approach to predict the pixel labels at t_p. First, all the pixels labeled as belonging to the water class at t_0 in the m^{th} MODIS pixel are determined as the 'unchanged water pixels'. Then, the number of land-to-water-conversion pixels is calculated. Assume Δf is the increase of water fraction in this MODIS pixel. Each MODIS pixel is divided into $16 \times 16 = 256$ 30 m Landsat pixels (16 is the scale factor between MODIS and Landsat images). Then there are $l = \text{round}(256 \times \Delta f)$ land-to-water-conversion pixels in the m^{th} MODIS pixel, where round(·) means rounding to the nearest integer number. The locations of the land-to-water-conversion pixels are determined by selecting l pixels of the land class at t_0 with the highest water likelihood index (such as those highlighted with red circles in Fig. 3), assuming the higher the value in the water likelihood map, the more likely the pixel of the land class at t_0 is changed to water class at t_p (i.e., the lower the elevation, the more likely the pixel is inundated (Armon et al., 2020; Li et al., 2019c; Pekel et al., 2016)). In selecting the l land-to-water-conversion pixels in the m^{th} MODIS pixel, the water likelihood indexes of the pixels of the land class at t_0 are first sorted in descending order. Then the l^{th} largest pixel likelihood index is used as the threshold. Finally, the pixels of the land class at t_0 which have the water likelihood that is not lower than the threshold are selected as the land-to-water-conversion pixels. Note that if l is larger than the number of land pixels in the m^{th} MODIS pixel, all the land pixels are determined as the land-to-water-conversion pixels in this m^{th} MODIS pixel. The other pixels of the land class at t_0 are then determined as the 'unchanged land pixels' from t_0 to t_p. Finally, the surface water map at t_p in the m^{th} MODIS pixel is produced by combining the 'unchanged water pixels', 'land-to-water-conversion pixels', and the 'unchanged land pixels'.

Case 2: water fraction unchanged in the m^{th} MODIS pixel. In this case, the class labels of all the 30 m pixels in the m^{th} MODIS pixel are assumed to be unchanged from t_0 to t_p.

Case 3: water fraction decreases in the m^{th} MODIS pixel. STSWM assumes, in the m^{th} MODIS pixel, the 30 m pixels labeled as belonging to the land class at t_0 are unchanged in the class label at t_p, and the 30 m pixels labeled as belonging to the water class at t_0 could have changed membership to the land class. In STSWM, first, all the pixels labeled as belonging to the land class at t_0 in the m^{th} MODIS pixel are determined as the 'unchanged land pixels'. Then, the number of water-to-land-conversion pixels is calculated. Assume Δf is the increase of land fraction in this MODIS pixel, and then there are $l = \text{round}(256 \times \Delta f)$ water-to-land-conversion pixels in the m^{th} MODIS pixel. The locations of the water-to-land-conversion pixels are determined by selecting l pixels of the water class at t_0 with the lowest water likelihood index (such as those highlighted with green circles in Fig. 3), assuming the lower the value in the water likelihood map, the more likely the pixel of the water class at t_0 is changed to land class at t_p (i.e., the higher the elevation, the more likely the pixel is not inundated and exposed to land). In selecting the l water-to-land-conversion pixels in the m^{th} MODIS pixel, the water likelihood indexes of the pixels of the water class at t_0 are first sorted in ascending order. Then the l^{th} lowest pixel likelihood index is used as the threshold. Finally, the pixels of the water class at t_0 which have the water likelihood that is not higher than the threshold are selected as the water-to-land-conversion pixels. Note that if l is larger than the number of water pixels in the m^{th} MODIS pixel, then all the water pixels are determined as the water-to-land-conversion pixels in this m^{th} MODIS pixel. The other pixels of the water class at t_0 are then determined as the 'unchanged water pixels'. Finally, the surface water map at t_p in the m^{th} MODIS pixel is produced by combining the 'unchanged land pixels', 'water-to-land-conversion pixels' and the 'unchanged water pixels' from t_0 to t_p.

2.3. Post-processing of the downscaled 30 m surface water map based on spatial filters

A post-processing step that uses spatial filters applied to the land-to-water-conversion pixels and water-to-land-conversion pixels is performed to incorporate local topographic information in producing the final surface water map (Fig. 1). The basic idea is that the water level should be relatively uniform for inundated surface water areas in a local region. Thus, in STSWM, if a pixel is inundated at t_p, then the spatially neighboring pixels with similar elevation should be inundated at t_p. The water likelihood representing the topographic information is used to select local pixels with similar topographic (elevation) conditions to post-process the initial downscaled 30 m resolution surface water map.

The local similar water likelihood pixels, which indicate pixels with similar elevation information, are selected. First, for a target pixel, the surrounding neighborhood pixels inside a square window whose center is the target pixel are defined as candidate pixels. Then a set of neighboring pixels from the candidate pixels, which have the most similar water likelihood to the target pixel, are selected as the similar water likelihood pixels to the target pixel.

A schematic diagram of selecting similar water likelihood pixels is shown in Fig. 4. First, the size of the square window to define the candidate pixels is determined (Step (1) in Fig. 4). The size is set to 9 MODIS pixels if the target pixel is the permanent and seasonal water pixels (SWO > 0) and 5 MODIS pixels if the target pixel is the permanent land pixels (SWO = 0) in the SWO map, respectively, after a process of trial and error (Step (1) in Fig. 4). Note that the presence of cloud cover in the MODIS pixel results in the 'no observations' pixels in the initial surface water map at t_p such as those highlighted in black circles in Fig. 4 (d), thereby the number of candidate pixels may be insufficient within the square window in the selection of similar pixels. In STSWM, the size of the moving window is enlarged until the square window contains a certain number of cloud-free MODIS pixels.

Once the size of the square window is determined, the water likelihood values in the square window in the surface water likelihood map are incorporated (Step (1) in Fig. 4), and a set of n neighboring pixels from the candidate pixels are selected using a threshold (Step (2) in Fig. 4). In this step, the absolute differences in water likelihood between the target pixel and its neighboring candidate pixels within the square window are calculated and sorted in descending order. The threshold is set to the n^{th} (n is set to 30 through many trails) smallest absolute difference value from the candidate pixels. The candidate pixels that have the water likelihood that is not lower than the threshold are defined as the local similar water likelihood pixels.

Based on the set of selected similar water likelihood pixels, the labels of the similar pixels using the initial surface water map at t_p are used to determine the label of the target pixel (Step (3) in Fig. 4). The label of the i^{th} target pixel p_i is determined according to the labels of the similar pixels in the initial surface water map at t_p based on a weighted function in Eqs. (5–8):

$$U_i^{water} = \sum_{k=1}^{n} w_k \times f_{water}(p_k) \tag{5}$$

$$U_i^{land} = \sum_{k=1}^{n} w_k \times f_{land}(p_k) \tag{6}$$

where p_k is the k^{th} similar water likelihood pixel of the target pixel p_i. $f_{water}(p_k)$ equals to 1 if p_k is labeled as 'water' in the initial surface water map and 0 otherwise. $f_{land}(p_k)$ equals to 1 if p_k is labeled as 'land' in the

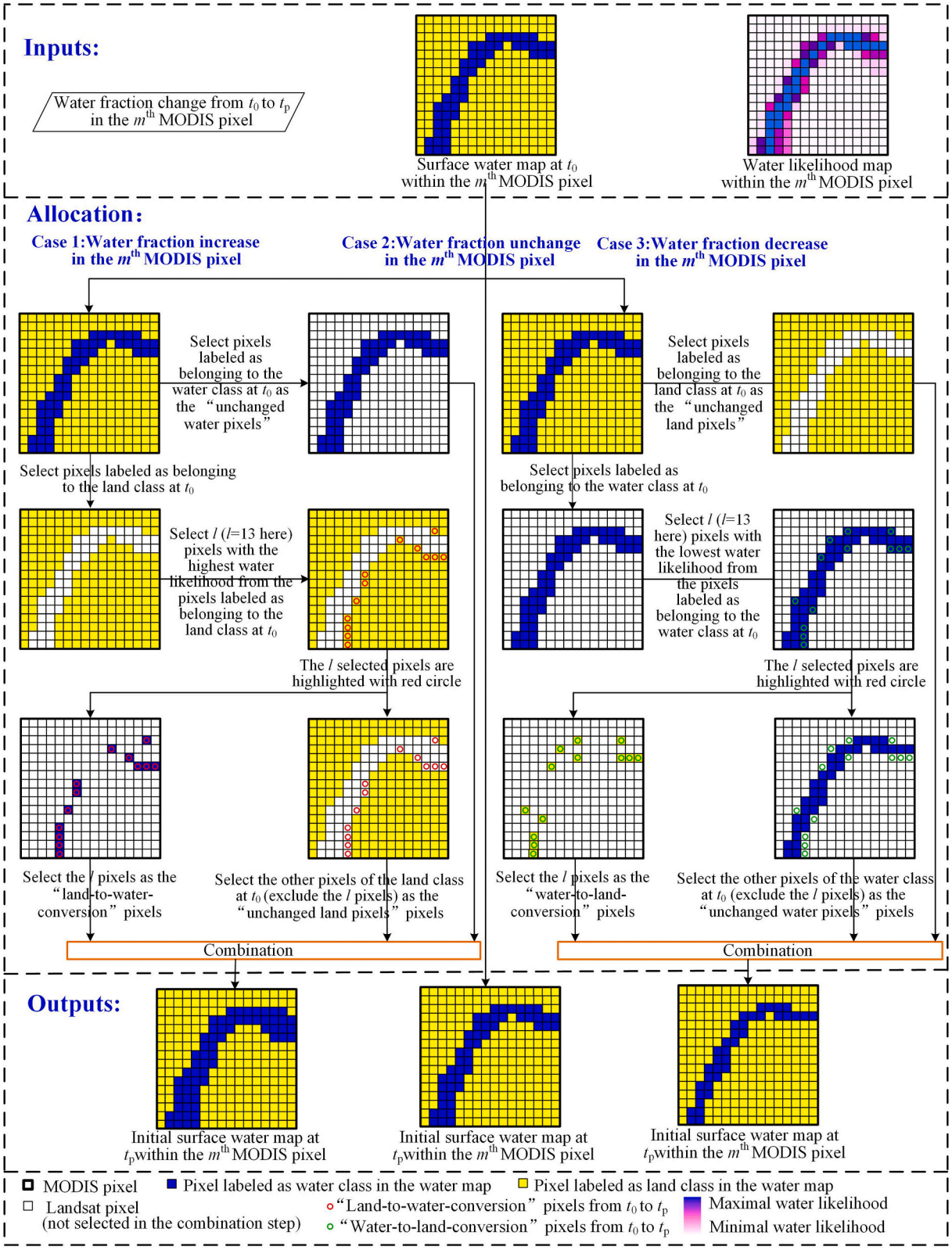

Fig. 3. Summary of the downscaling of coarse spatial resolution water fraction change to 30 m resolution in the m^{th} MODIS pixel. The figure shows the downscaling using the m^{th} MODIS pixel which contains s^2 Landsat pixels (s is the scale factor between MODIS and Landsat pixel, and is set to 16 in this figure). The water fraction increase (case 1) and decrease (case 3) are assumed to be 5% ($\Delta f = 0.05$) in the m^{th} MODIS pixel, and then there are $l = 16^2 \times 0.05 = 13$ high-resolution land-to-water-conversion pixels (highlighted with red circles) and water-to-land-conversion pixels (highlighted with green circles) in STSWM. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

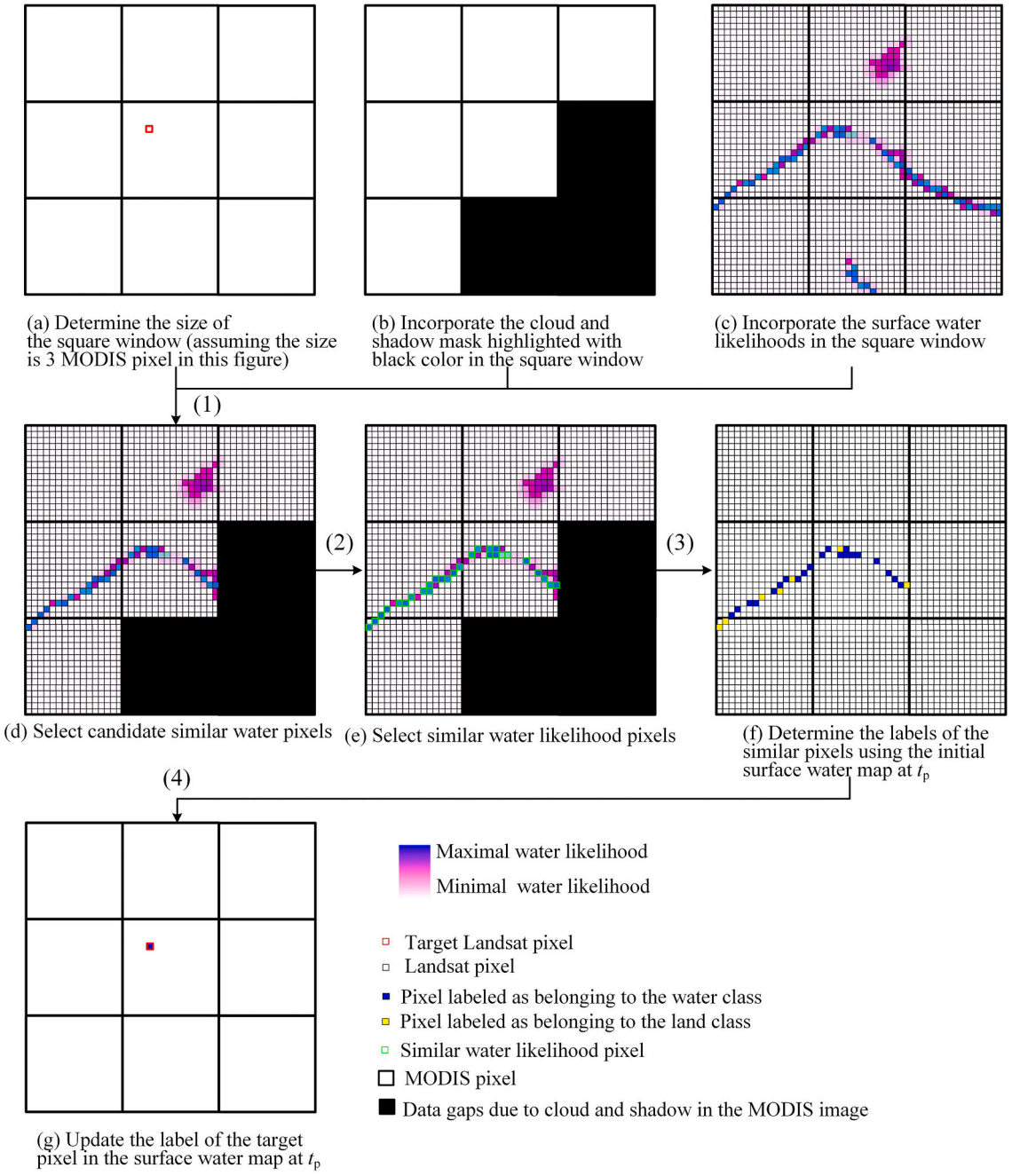

Fig. 4. The use of similar water likelihood pixels in updating the class label for a target pixel. The square window size is set to 5 MODIS pixels or larger in real applications. The surface water likelihood map is used to select a set of similar pixels. The initial surface water map at t_p is used to calculate the contributions of water and land pixels from similar pixels to the target pixel. Step (1) incorporates the cloud and shadow mask and the surface water likelihood values in the square window. Step (2) selects a set of similar water likelihood pixels (highlighted with green rectangles) for the target pixel (highlighted with a red rectangle in Fig. 4(a)). Step (3) determines the labels of the similar pixels using the initial surface water map at t_p. Step (4) updates the label of the target pixel in the surface water map at t_p according to the contributions of similar water likelihood pixels based on Eqs. (5–8). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

initial surface water map and 0 otherwise. w_k is the weight of similar water likelihood pixel p_k that is calculated as:

$$w_k = (1/D_k) \Big/ \sum_{k=1}^{n} (1/D_k) \tag{7}$$

$$D_k = 1 + d(p_i, p_k)/(W/2) \tag{8}$$

where $d(p_i, p_k)$ is the geometric distance between the p_i and p_k, and W is the size of the local square window (Gao et al., 2006; Zhu et al., 2016).

Once U_i^{water} and U_i^{land} are calculated, the target pixel p_i label is updated as water class at t_p if U_i^{water} is larger than U_i^{land} and land class otherwise (step (4) in Fig. 4). Note that only the land-to-water-conversion pixels highlighted with red circles in Fig. 3 and water-to-land-conversion pixels highlighted with green circles in Fig. 3 are predicted using the aforementioned downscaling approach.

At the end of STSWM, a spatial filter that has a smoothing effect is applied to modify the surface water map to eliminate speckle-like artifacts that may have occurred. The spatial filter is applied only to pixels that have changed labels by comparing the surface water map at t_0 and

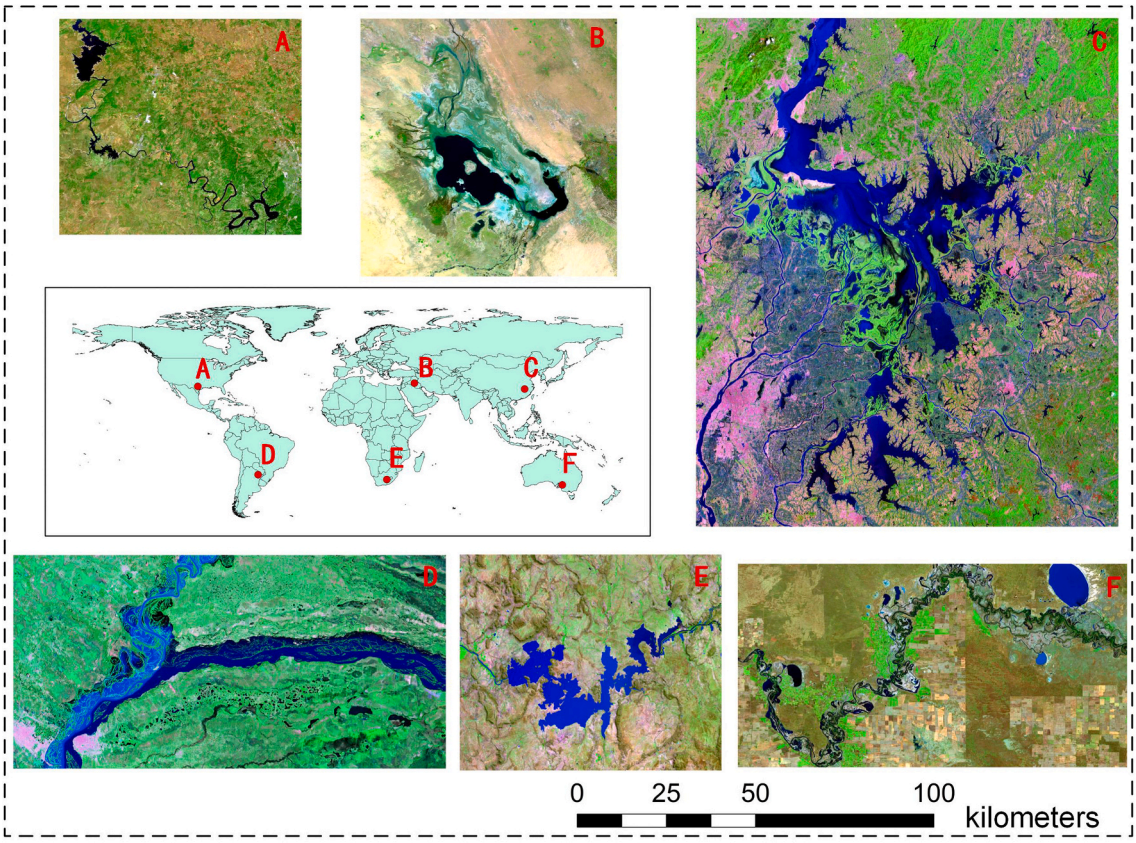

Fig. 5. The six study sites in the experiment.

the initial surface water map at t_p and the 'no observations' pixels labeled in the initial surface water map at t_p. For these pixels, a majority filter using the target pixel as the central pixel is applied to the initial surface water map at t_p. The size of the local square window of the majority filter is set, after a trial and error process, to 3 Landsat pixels. After the majority filter, the final 30 m surface water map at t_p is generated.

3. Experiments

3.1. Study area and data

The potential of STSWM was evaluated using experiments in six floodplain regions. These regions are located on five continents in the world in Fig. 5. Specifically, the regions include a site near the Colorado River, USA (Fig. 5A), a site near the Razazza Lake, Iraq (Fig. 5B), a site near the Poyang Lake, China (Fig. 5C), a site near the Parana River, Argentina and Paraguay (Fig. 5D), a site near the Oviston Nature Reserve, South Africa (Fig. 5E) and a site near the Murray River, Australia (Fig. 5F). The STSWM was used to predict 30 m surface water maps at an 8-day time step based on 8-day composite surface reflectance MODIS (MOD09A1, 500 m) and cloud-free Landsat data downloaded from the United States Geological Survey (USGS) thorough a selected year. The selected years and the Landsat data used in each region in STSWM are in Table 1. These years were selected if the region experienced inundation in that year and more than four cloud-free Landsat data were available throughout the year. The locations of the regions are in Figs. S2-S7 in the supplementary data.

The input of STSWM includes a MODIS image at the prediction time, a MODIS-Landsat image pair at the ancillary date, the SWO map, and a DEM map. The STSWM used the 8-day composite MOD09A1 data instead of the daily MODIS image because the daily MODIS is affected not only by a high percentage of cloud cover, but also by the BRDF effect due to the very high viewing angles near the edge of the satellite swaths (Heimhuber et al., 2018; Walker et al., 2012). The percentages of cloud cover and shadows in the MOD09A1 data that were determined based on the MODIS quality assessment data are shown in Fig. 6. The SWO data used are from the SWO map of Pekel et al. (2016) created using the Joint Research Centre's (JRC) Global Surface Water Mapping Layers (v1.3) downloaded from the Google Earth Engine, which provides the frequency at which the water was present (from 0 to 100%). The 30 m DEM data used in the experiment are 30 m SRTM DEMs (Farr et al., 2007).

In each site, STSWM was used to predict forty-six 30 m resolution surface water maps at the dates of the forty-six scenes of MOD09A1 imagery thorough the selected year. For each prediction date, a pair of Landsat-MODIS images was selected as the STSWM ancillary data. Note that more than one cloud-free MODIS-Landsat pairs were available as the candidate ancillary data for STSWM. The selection criterion of the ancillary cloud-free MODIS-Landsat pairs is introduced in section S3 in the supplementary data.

3.2. Model comparison

The STSWM predictions were compared with the GSW monthly water history maps downloaded from the Google Earth Engine. The GSW monthly water history maps were derived from Landsat 5, 7, and 8 images, in which each pixel was classified as 'water detected', 'not water' or 'no observations' in each month using an expert system. The 'no observations' pixels were due to cloud cover, striping, and no Landsat imagery available. For instance, regions such as Siberia that weren't imaged until 2000 (Pekel et al., 2016; Pickens et al., 2020). The monthly water history maps provide information about all of the water detections at a monthly interval, i.e., whether or not water was detected in a particular month (Pekel et al., 2016).

Table 1
The information about the Landsat data used in STSWM for each region.

	Selected year for prediction	Landsat type	Dates of Landsat used in STSWM	Region size in Landsat pixels
Colorado River, USA	2011	Landsat 5	2011/03/03, 2011/06/07, 2011/07/25, 2011/08/26, 2011/09/11, 2011/10/13, 2011/10/29	2000 × 2240
Razazza Lake, Iraq	2014	Landsat 8	2014/01/25, 2014/02/10, 2014/02/26, 2014/05/17, 2014/06/02, 2014/06/18, 2014/07/04, 2014/07/20, 2014/08/05, 2014/08/21, 2014/09/06, 2014/09/22, 2014/10/08, 2014/10/24, 2014/11/09	2400 × 2400
Poyang Lake, China	2014	Landsat 8	2014/01/25, 2014/05/01, 2014/08/05, 2014/10/08, 2014/10/24	4720 × 3920
Parana River, Argentina and Paraguay	2010	Landsat 5	2010/01/29, 2010/10/28, 2010/11/13, 2010/12/15	2000 × 4000
Oviston Nature Reserve, South Africa	2014	Landsat 8	2014/03/28, 2014/08/03, 2014/09/04, 2014/09/20, 2014/10/06, 2014/10/22, 2014/11/07, 2014/12/25	1600 × 2400
Murray River, Australia	2011	Landsat 5	2011/01/02, 2011/01/18, 2011/03/07, 2011/04/08, 2011/04/24, 2011/08/30, 2011/09/15,2011/10/17	1920 × 3600

The proposed STSWM was also compared to two spatiotemporal methods. The first spatiotemporal method was the spatiotemporal subpixel mapping (STPSA) method (Xu and Huang, 2014). In the STPSA, first, the MODIS image at t_p was unmixed into the water and land fraction images, which were then downscaled to 30 m surface water maps using an optimization algorithm and using the 30 m surface water map t_0 as ancillary data. The other spatiotemporal method is a combination of STARFM and Otsu's segmentation method. STARFM was used to fuse the MODIS image at t_p with the MODIS and Landsat image pair at t_0 to generate a Landsat-like image at t_p, which was then used to calculate the 30 m MNDWI map; finally, Otsu segmentation was applied to the MNDWI map to automatically generate the 30 m water-land map at t_p. In addition, the STSWM was also compared to a Hard Classification (HC) method, which was directly applied to the Landsat image at t_p based on the same method described in section 2.1.1. The proposed STSWM was also compared to the GSW monthly water history maps. The STSPA and STARMF were not used to generate time-series surface water maps because they are unsuitable for processing the cloud-cover and shadows in the MODIS images. The selection criterion of the ancillary MODIS-Landsat image pairs for STPSA and STARFM is the same as that for STSWM, as is introduced in section S3 in the supplementary data.

3.3. Accuracy assessment of the time series maps

The stratified random sampling was used to provide unbiased estimators of the accuracies of the STSWM surface water maps (Olofsson et al., 2014; Olofsson et al., 2013; Pickens et al., 2020; Stehman, 2009; Tulbure et al., 2016). First, a 'surface water stratum' and a 'land stratum' were created from the STSWM results for each region. The 'surface water stratum' was created based on the maximum water extend map derived from the forty-six STSWM prediction maps (Tulbure et al., 2016). The 'land stratum' was created based on pixels that were predicted as 'land' in all the forty-six STSWM maps. Then, 200 sample points, including 100 samples from the 'surface water stratum' and 100 from the 'land stratum', were selected for each of the six study sites. The individual sample pixels were manually interpreted using all available Landsat scenes (including Landsat 5, 7, and 8) in the study year. Each Landsat observation for a sample pixel was visually interpreted as 'water pixel', 'land pixel', or 'bad data' using the Google Earth Engine cloud-computing platform. The 'bad data' includes pixels covered by cloud, shadow, or striping at a Landsat scene. All the 'water pixels' and 'land pixels' constituted the reference data for validation.

For each STSWM map, a subset of the reference data was used for validation. Since each STSWM map was generated from an 8-day composite MODIS data, only the reference data from the Landsat scenes observed within the period represented by the MODIS data is used to assess the accuracy of that STSWM map. Note that there may be more than on Landsat scenes available within the period represented by the MODIS data. In this case, if the sample pixels at the same location have the same visual interpretation label within the 8-days period represented by the MODIS, then only one sample was maintained for validation. Otherwise, if the sample pixels at the same location have different visual interpretation labels, these samples were all excluded in the accuracy assessment.

The accuracy and variances of the estimators of each STSWM map were estimated based on the corresponding reference sample data. The estimation of accuracies for stratified random sampling includes overall accuracy, producer's accuracy, and user's accuracy (Olofsson et al., 2014; Olofsson et al., 2013). First, the error matrix in terms of sample counts was first computed for each STSWM map. The sample counts in the error matrix were then converted to estimated area proportions of the error matrix. The stratified estimator of the overall accuracy of the map, the producer's and user's accuracies of water, and the producer's and user's accuracies of land were calculated. The standard error of the accuracy estimates were also computed to report the sampling variability associated with the accuracy estimates and provide the uncertainty of the estimates. The details of the accuracy estimation calculations can be found in Section S4 in the supplementary data.

The original 200 samples may be covered by cloud, shadow, or striping in the Landsat image as 'bad data' that were excluded in validating the STSWM map. Thus the actual valid samples may vary for different STSWM maps. In order to contain a sufficient number of samples to get reliable accuracy estimations in the validation, the accuracies of an STSWM map was estimated only when no less than 60% of the samples (i.e., 120 samples) were valid and not 'bad data' in the Landsat image within the period of that MODIS image. In addition, the comparison methods, including STPSA and STARFM, are unsuitable for mapping cloud-contaminated MODIS data; thus, only the STSWM predictions were assessed.

4. Results

4.1. Comparison of the STSWM predicted surface water maps with the GSW monthly water history maps

4.1.1. Comparison of the surface water frequency

The yearly surface water frequency maps from the GSW layers and STSWM in Fig. 7 were used to reveal the surface water spatiotemporal dynamics. First, the yearly surface water count maps according to the GSW and STSWM maps were produced, respectively. For each site, the GSW yearly surface water count map is produced according to the twelve GSW monthly water history maps in that year, and the STSWM yearly surface water count map is produced according to the forty-six prediction maps in that year. The surface water count is defined as the number of pixels labeled as water in the GSW monthly water history maps or the STSWM prediction maps. Then, the surface water frequency

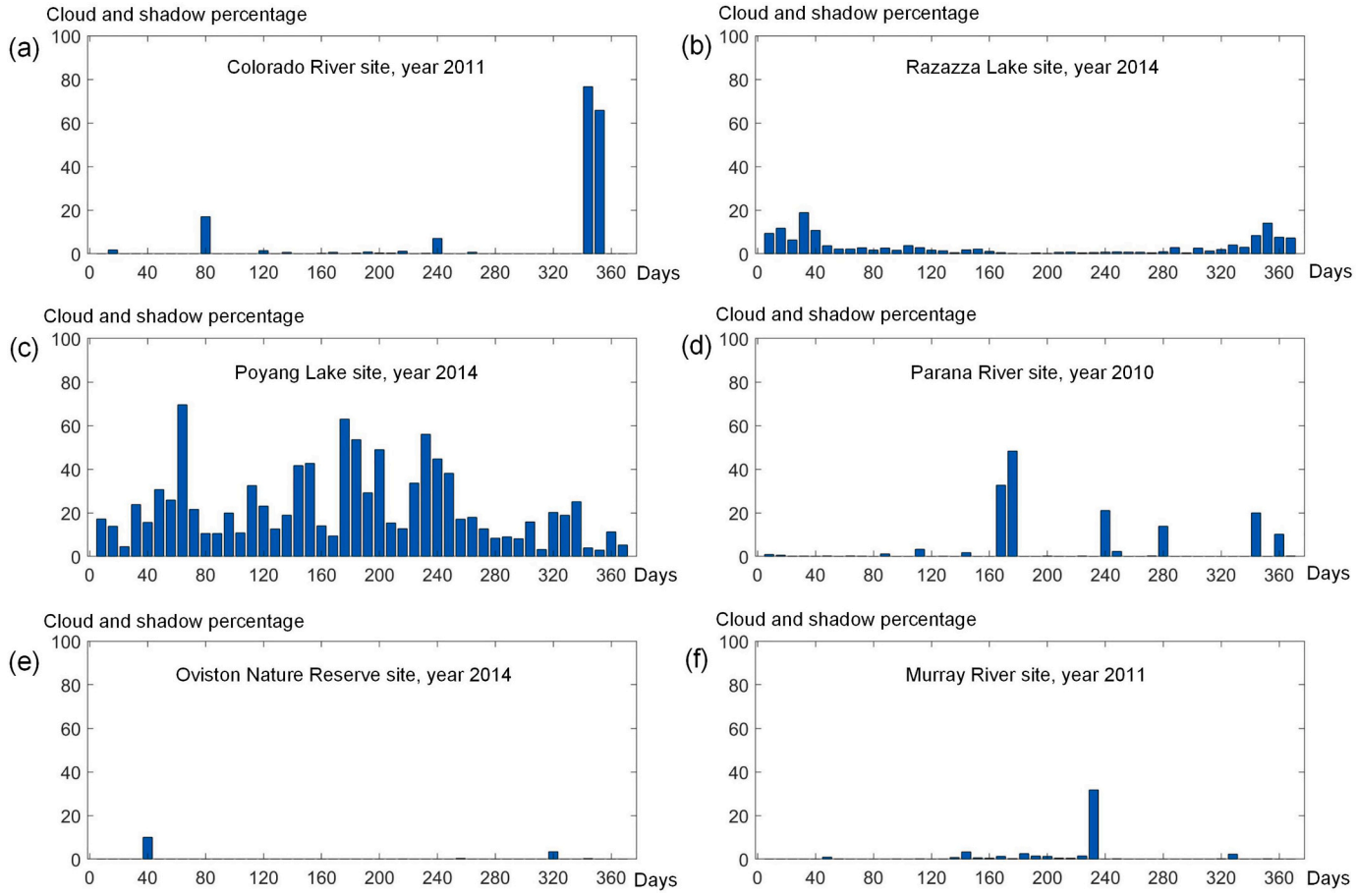

Fig. 6. The percentages of cloud cover and shadows in the MOD09A1 data in the six study sites.

is defined by dividing the surface water count by the total number of valid observations.

Differences are found in the GSW and STSWM surface water frequency maps in Fig. 7. In particular, in the Razazza River site in Fig. 8, the surface water frequency highlighted in the red rectangle in the GSW frequency map (Fig. 8(a)) is higher than that highlighted in the red rectangle in the STSWM frequency map (Fig. 8(b)). This difference resulted in the fact that many pixels are 'no observations' pixels in Jan., Feb., Mar., and Dec., such as those in black color in Fig. 8 (c), (d), (e), and (n). In the GSW monthly water history maps, the 'no observations' pixels were mainly due to cloud in Feb., and due to striping in Mar., and due to the fact that Pekel et al. (2016) do not map >32°N in Jan. and do not map >30°N in Dec. in the JRC Monthly Water History (v1.3) dataset. The 'no observations' pixels in the GSW monthly water history maps were not counted in calculating the water frequency. Therefore, the water frequency produced from the GSW monthly water history maps would be biased by the temporal distributions of the valid Landsat observation. In contrast, the STSWM prediction maps did not contain 'no observations' pixels (such as the maps in Fig. 10). Thus the valid observation in the time series STSWM prediction maps equals forty-six for all pixels in all sites. The impact of the spatiotemporal distribution of the 'no observations' pixels could be reduced in the STSWM frequency map.

4.1.2. Comparison of the spatiotemporal dynamics of surface water maps in the GSW and STSWM maps

The spatiotemporal dynamics of surface water maps in the GSW maps (Fig. 9) and STSWM maps (Fig. 10) in the Parana River site were compared.

In the GSW monthly water history maps in Fig. 9, the water map in June 2010 is covered mainly by the 'no observations' pixels because there are no available Landsat 5 and 7 images for the center of the area. The monthly water history maps for Feb., July, and Aug. contain 'no observations' pixels due to the striping of the Landsat 7 images. Therefore, the GSW monthly water history maps neither provide surface water changes within a month nor provide water maps when there is no Landsat image available for the month.

In contrast to the GSW monthly water history maps, the STSWM could predict 30 m surface water maps at an 8-day time step (Fig. 10) due to MODIS images with a finer temporal resolution than the Landsat images. The STSWM maps show that the water area decreased from days 177–184 (June 26 to July 3) to days 185–192 (July 4 to July 11) in 2010, such as those indicated by the red circles in Fig. 10. This information is not provided in the GSW monthly water history maps because there were mostly no observations in the Landsat pixels for June 2010 in Fig. 9. Similarly, the STSWM maps indicate an increase in the water area from days 337–344 (Oct. 3 to Oct. 10) to days 345–352 (Oct. 11 to Oct. 18) in 2010 such as those indicated by the red circles in Fig. 10. This increase in water area is not seen on the GSW monthly water history map for Dec. 2010, which was generated with a monthly frequency. Therefore, the GSW monthly water history maps may fail to detect the surface water changes due to the presence of cloud cover and the availability of Landsat data, while the STSWM can represent the surface water dynamics more comprehensively. It should also be noticed that artifacts could be found in the STSWM maps. For instance, the extents of water highlighted with black circles in row 6, column 2 in Fig. 10 (days 169–176 in 2010, the percentage of cloud and shadow cover in the predicted MODIS image is 48.35%) were predicted that are the result of shadows in the MODIS image according to the MODIS quality assessment data.

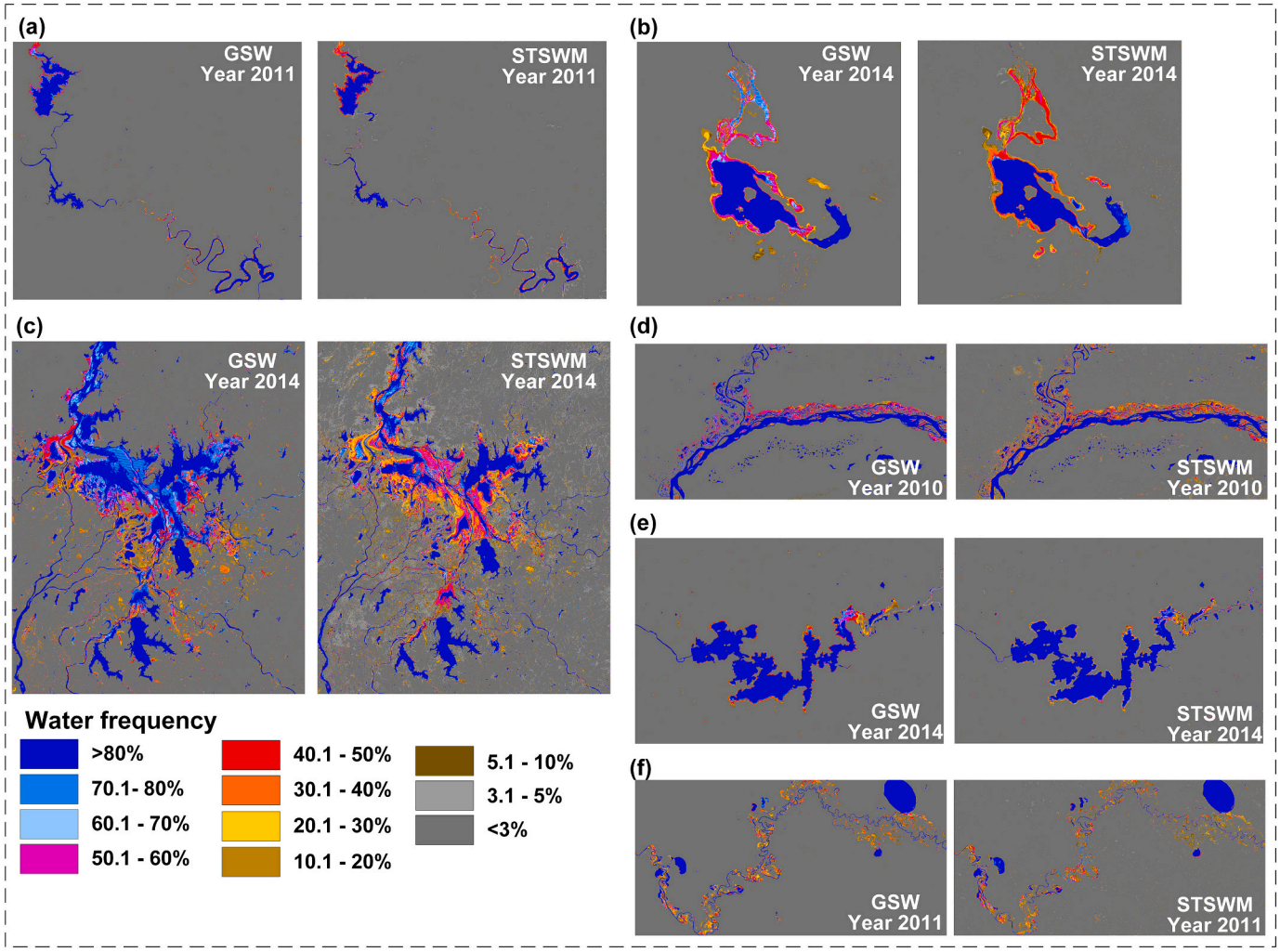

Fig. 7. The yearly surface water frequency maps from the GSW monthly water history maps and the STSWM predictions. (a) Colorado River site, USA; (b) Razazza Lake site, Iraq; (c) Poyang Lake site, China; (d) Parana River site, Argentina and Paraguay; (e) Oviston Nature Reserve site, South Africa; (f) Murray River site, Australia.

Fig. 11 (b)-(y) shows the comparison between the GSW monthly water history maps and the STSWM maps in the zoomed-in area in the red square in Fig. 11(a) (the STSWM time series maps at an 8-day time step in the zoomed-in area is shown in Fig. S8 in the supplementary data). The GSW monthly water history map for June (Fig. 11(g) is fully filled with 'no observations' data, and the maps for Feb. (Fig. 11(c)), July (Fig. 11(h)), Aug. (Fig. 11(i)), Sept. (Fig. 11(j)) are affected by the striping of the Landsat 7 image. The GSW monthly water history map for Sept. (Fig. 11(j)) also contained 'no observations' pixels due to cloud cover, as indicated by the red circle. In contrast, the STSWM reduced the impacts of the cloud cover and striping in the Landsat 7 images on the surface water maps, and the predicted maps shown in Fig. 11 are free of 'no observations' pixels. In addition, the change in the water areas is evident in the STSWM maps, such as those indicated by the red circles in Fig. 11(r–s), which are not observed on the GSW monthly water history maps in Fig. 11(f-g). Thus, the generation of temporally dense water maps generated using the STSWM can better reflect the water dynamics than those produced for relatively long temporal frequencies.

4.2. *Comparison of the predicted surface water maps obtained using different methods*

Fig. 12 shows the comparison of predicting maps in the Parana River site at a specific time t_p (the MODIS 8-day composite data at t_p represents the time from 2010/12/11 to 2010/12/18) from different spatiotemporal fusion methods and a classification method. The studied region experienced inundation from t_0 to t_p, which is shown in the Landsat images in Fig. 12(c–d). The differences in the predicted surface water maps obtained using the compared methods were obvious. The HC map shown in Fig. 12(g) was produced using the Landsat images shown in Fig. 12(d), in which the inundated areas are evident. The STPSA map (Fig. 12(h)) did not predict the shape of the rivers, indicating that the spatiotemporal correlation model in the STPSA is not suitable for use with rivers that have linear spatial distributions and are usually smaller than the coarse pixel resolution (Atkinson, 2009). In addition, the STPSA map contained many small water class patches, which were represented as speckle-like artifacts such as those indicated by the black circle in Fig. 12(h). These incorrectly predicted water patches are due to the spectral unmixing error. For instance, suppose a MODIS pixel contains 100% land class, but the unmixing analysis's fraction image contains 5% water class. In this case, $16 \times 16 \times 5\% = 13$ (16 is the scale factor between the MODIS and Landsat pixels) Landsat resolution pixels would be labeled as water class in this MODIS pixel.

Both the STARFM map (Fig. 12(i)) and the final STSWM map (Fig. 12 (j)) eliminated the speckle-like artifacts. The STARFM failed to predict some of the water areas, such as those indicated by the black circle in the zoomed-in area in Fig. 12(i), and the small water bodies were erroneously clustered such as those highlighted in the red circle in the zoomed-

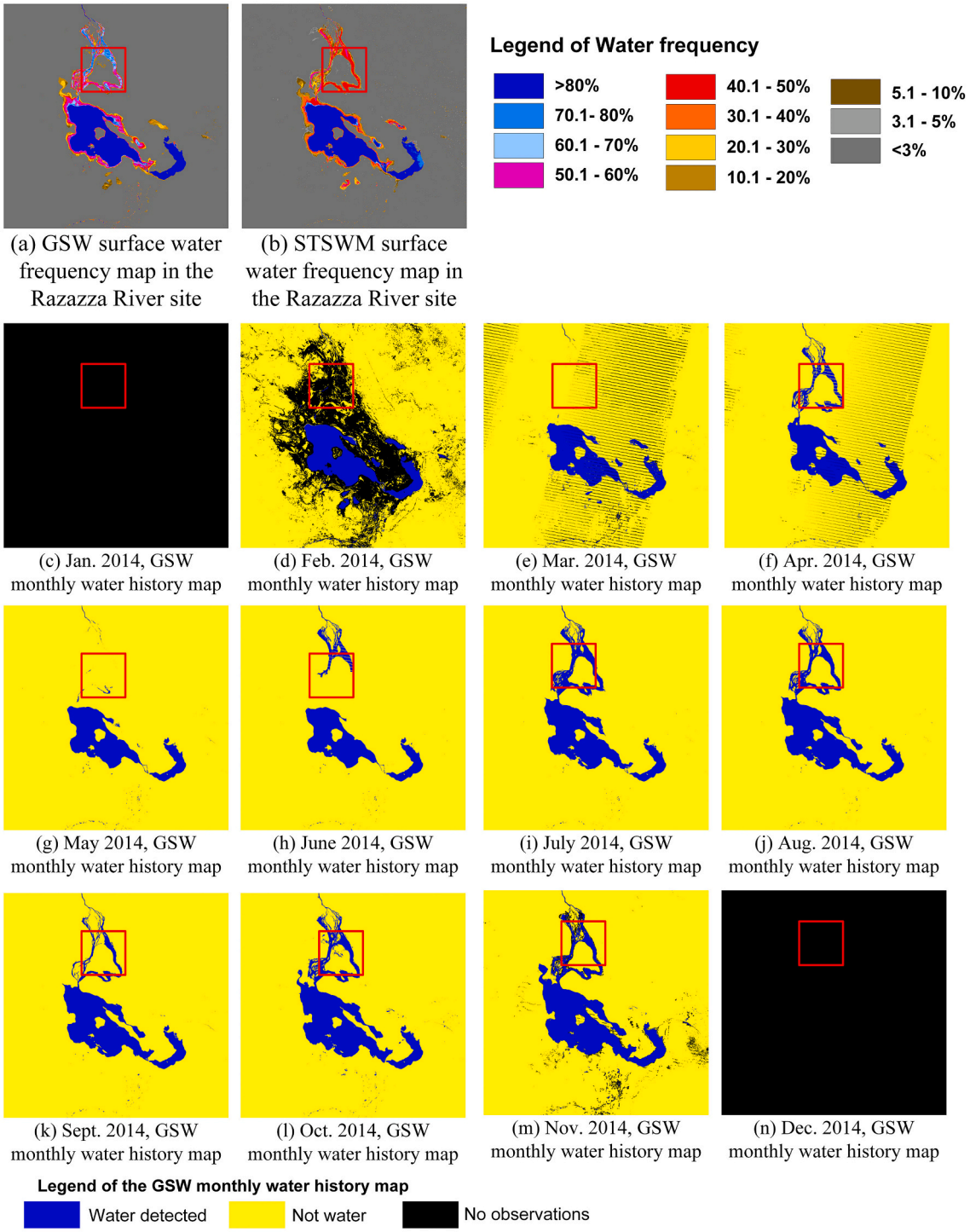

Fig. 8. GSW and STSWM yearly surface water frequency maps and the GSW monthly water history maps in 2014 in the Razzaza River site.

in area in Fig. 12(i). In contrast, the STSWM correctly predicted the water areas, such as those indicated by the black circle in the zoomed-in area in Fig. 12(j), and it reconstructed the shapes of the small water bodies that are more close to that predicted from Landsat at t_p from HC in Fig. 12(g).

4.3. Quantitative assessment of STSWM

The stratified estimators of different accuracies in the STSWM maps are shown in Figs. 13-15 and in Figs. S9-S10 in the supplementary data. The overall accuracies were about higher than 0.98 in the sites except for the Poyang Lake site in Fig. 13. The producer's accuracy of water was about higher than 0.75 (Fig. 14), and the producer's accuracy of land was about higher than 0.97 (Fig. S9 in the supplementary data) in all the sites. The stratified estimator of producer's accuracies of a class is related to the weight ($W_i \div n_{i.}$) of that stratum in calculating the estimated area proportion (Eq. S1 and S3), where W_i is the proportion of the area mapped as class i, and $n_{i.}$ is the total sample counts that are mapped as class i (Olofsson et al., 2013). If the weight ($W_i \div n_{i.}$) of a stratum of class i is small, the producer's accuracy will be more sensitive to the omission of a few samples of class i (Olofsson et al., 2013). In all the six sites, the weight ($W_i \div n_{i.}$) for the land stratum was more than twice that

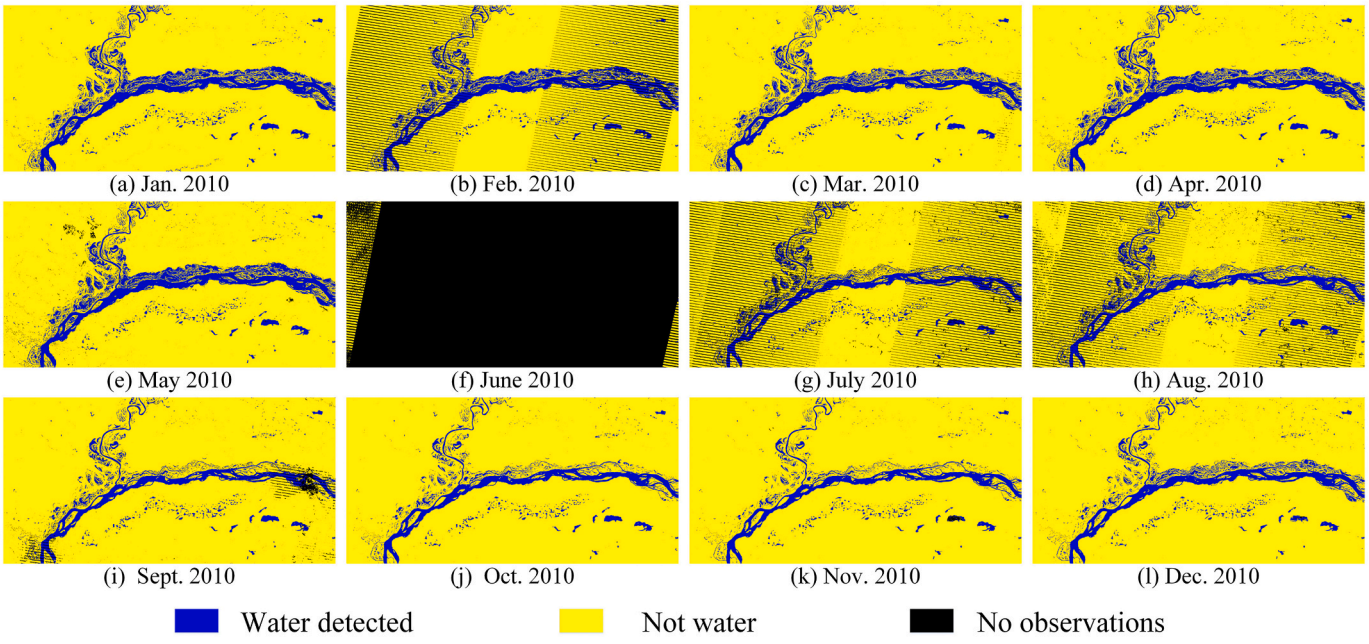

Fig. 9. GSW monthly water history maps in 2010 for the Parana River experiment.

of the water stratum, and thus the producer's accuracy of water is sensitive to the omission error of a few water samples. The user's accuracy of water was about higher than 0.88 (Fig. 15), and the user's accuracy of land was about higher than 0.95 (Fig. S10 in the supplementary data) in all the sites.

The producer's and user's accuracies of water represent variances in different sites. Fig. 14 shows the producer's accuracy of water in the six sites. The STSWM maps were generated with relatively lower producer's accuracies of water and large standard errors in Fig. 14(a). The main reason is that, in the computing of the producer's accuracy, the weight ($W_i \div n_{i.}$) for the land stratum was higher than 19 times that of the water stratum in the Colorado River site in Fig. 14(a), and the weight for the land stratum was less than seven times that of the water stratum in the other five sites. The relatively small weight of the water stratum resulted in a high sensitivity to the omission of a few water samples in Fig. 14(a) (Olofsson et al., 2013). The producer's accuracy of water in the Murray River site in Fig. 14(f) was relatively lower than those in Fig. 14(b-e). The Murray River site contains many narrow rivers (narrower than 3–5 Landsat pixels) and small water bodies in Fig. 5(f) and Fig. S7 in the supplementary data, while the other sites contain large lakes and wide rivers in Fig. 5. It shows a warning of possible omission error of water in the mixed water-land MODIS pixels for STSWM. In Fig. 15, the user's accuracies of water were relatively lower in the Poyang Lake site in Fig. 15(c) than in other sites. The main reason for the high commission error of water is that the Poyang Lake site contained a massive cloud coverage through the chosen year in Fig. 6(c). The average percentage of cloud and shadow in the study site accounts for about 22.7% in the Poyang Lake site and about less than 4% in the other five sites. Since water and shadow have similar spectral values, STSWM may probably generate commission error of water for shadow MODIS pixels that are not masked in the MODIS spectral unmixing stage.

The producer's and user's accuracies of water vary with different dates in Figs. 14-15. The main reason is that different MODIS images may contain different percentages of cloud and shadow cover. The presence of cloud and shadow cover in the MODIS image could result in a water commission error if the shadow pixels are not masked and unmixed as surface water. Besides, the selection of ancillary Landsat-MODIS image pair also has an impact on the STSWM accuracy. STSWM compares the ancillary and predicted MODIS images to estimate the change in surface water area in each MODIS pixel, and then downscales the water change to a 30 m scale. The STSWM downscaling approach will contain a large uncertainty if the surface water change is drastic. The experiments show that the producer's and user's accuracy of water decreased when the surface water change was large between the ancillary and predicted MODIS images, such as those highlighted with red ellipses in Fig. 14(d) and Fig. 15(f). In Fig. 14(d), STSWM generated a relatively low producer's accuracy (high omission error) of water for the STSWM map highlighted with a red ellipse. For this prediction map, STSWM compared the predicted MODIS (2010/01/25–2010/02/01, representing time t_p) in Fig. S11(b) with the ancillary MODIS (2010/11/09–2010/11/16, representing time t_0) in Fig. S11(a) in the supplementary data to estimate surface water change. During this t_0-t_p period, a large surface water increase was found according to the Landsat images in Fig. S11(c-d) in the supplementary data. STSWM has a relatively large uncertainty in predicting water area expansion, resulting in a high risk of water omission error. Similarly, in Fig. 15(f), STSWM generated a relatively low user's accuracy (high commission error) of water for the prediction map highlighted with a red ellipse. For this prediction map, STSWM compared the predicted MODIS (2011/07/20–2011/07/27, representing time t_p) in Fig. S12(b) with the ancillary MODIS (2011/04/23–2011/04/30, representing time t_0) in Fig. S12(a) in the supplementary data to estimate surface water change. During this t_0-t_p period, a large surface water decrease was found according to the Landsat images in Fig. S12(c-d) in the supplementary data. STSWM has a relatively large uncertainty in predicting water area contraction, resulting in a high risk of water commission error.

5. Discussion

5.1. Comparison between different spatiotemporal methods for surface water mapping

The characteristics of surface water dynamics are distinctive in both the spatial and temporal domains relative to that for other land covers such as forest (Assuncao et al., 2017) and urban (Deng and Zhu, 2020). For surface water, in the spatial domain, the spatial distribution of water is related to topography. In the temporal domain, the water change is often abrupt. The STSWM method presented here is different to other spatiotemporal methods in predicting surface water dynamics in both the spatial and temporal domains.

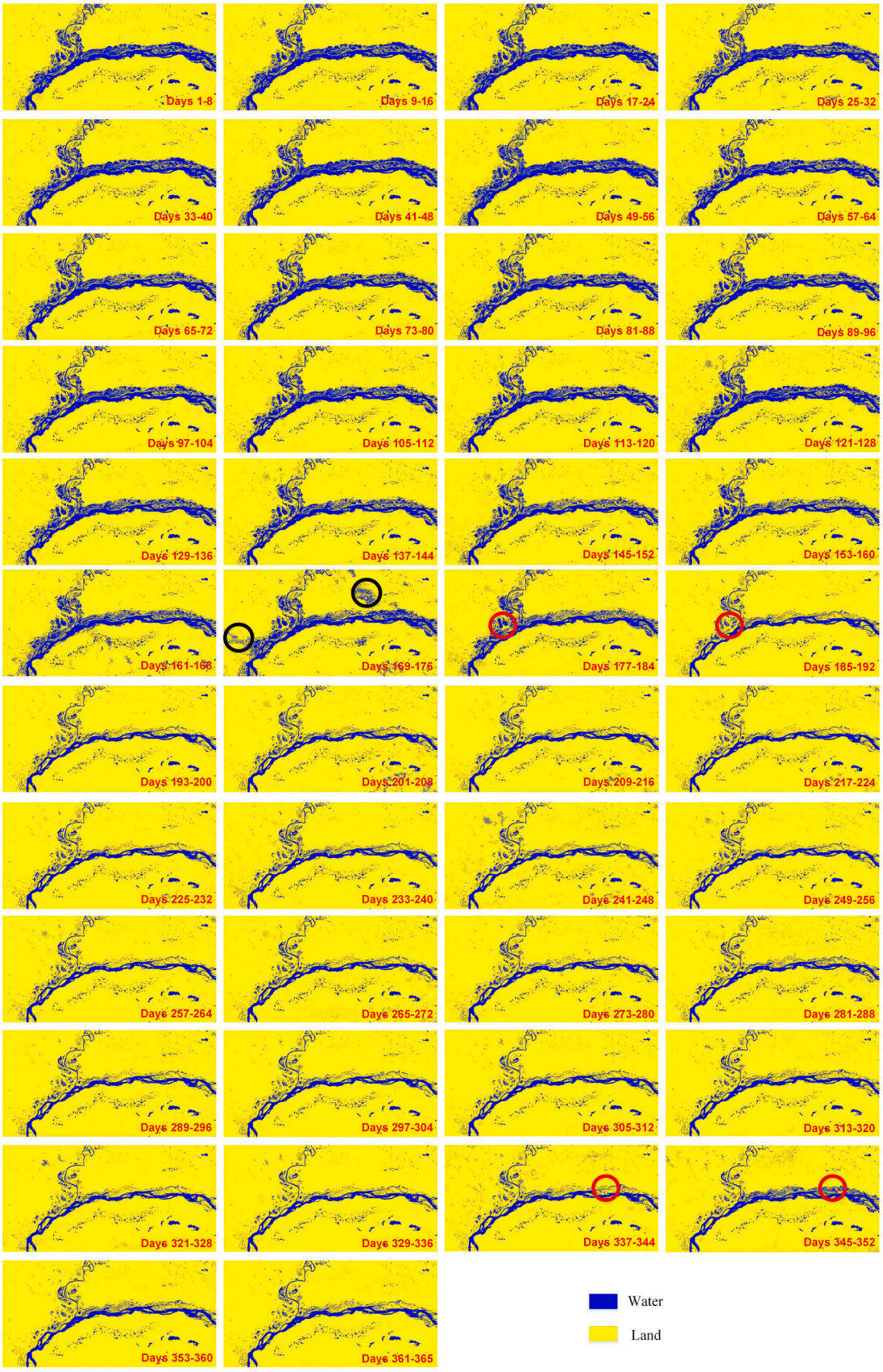

Fig. 10. STSWM predicted surface water maps according to the 8-day composite MODIS images taken in 2010 for the Parana River experiment. The red circles indicate regions that experience water change according to the STSWM maps. The black circles indicate the water extents in the STSWM map resulting from shadows in the MODIS image. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

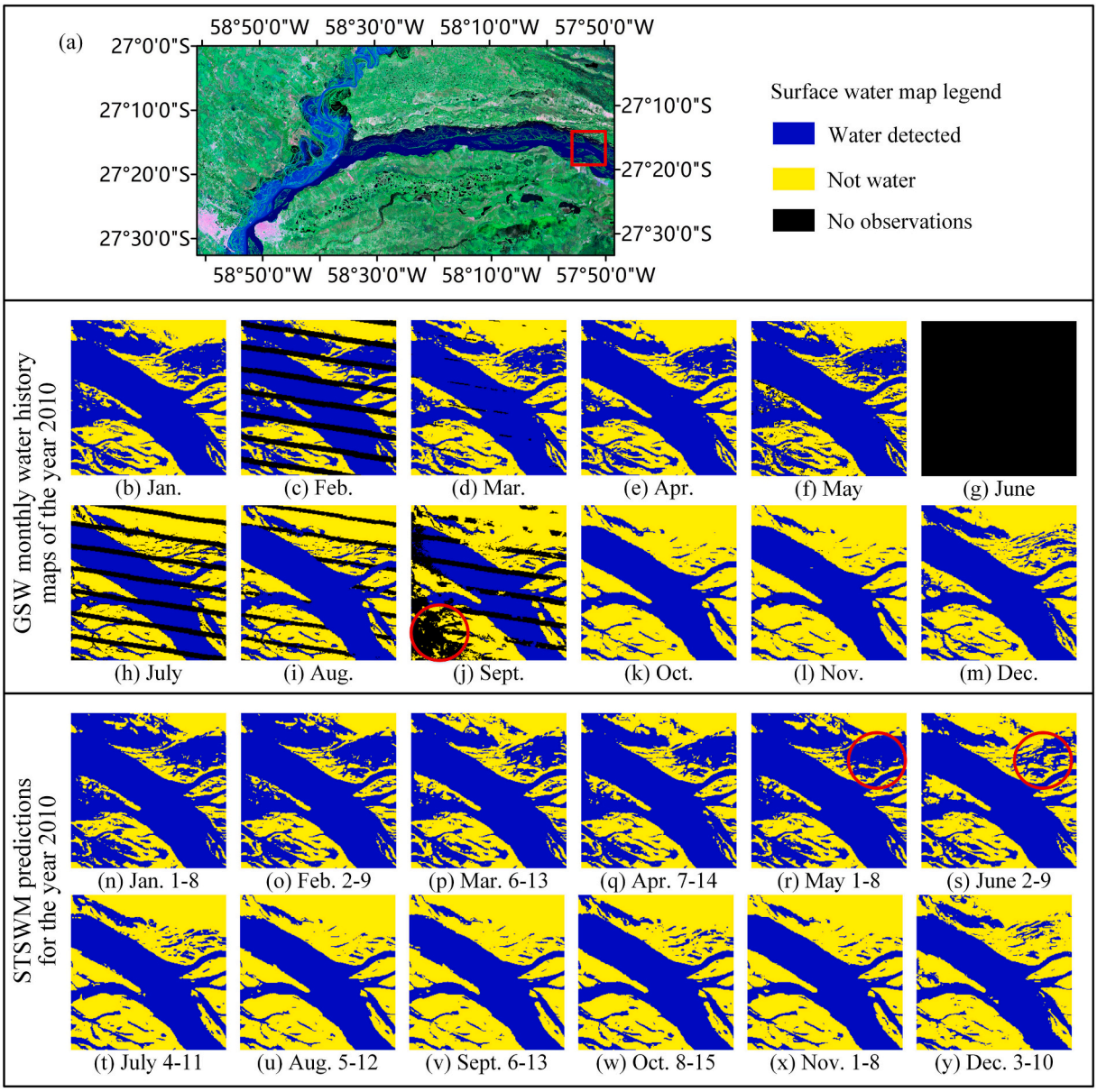

Fig. 11. Comparison of the GSW monthly water history maps and the STSWM predictions for selected dates in 2010. Fig. 11 (b)-(y) shows the surface water maps in the zoomed-in area in the red square in Fig. 11(a). The STSWM maps were selected if 1) the start and end dates of the input MODIS image at t_p were within a month and 2) the dates represented by the input MODIS data are the earliest for that month. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

First, in the spatial domain, a substantial difference was found in the spatial distributions of surface water predicted by the set of methods used. STARFM and most other spatiotemporal image fusion methods (Li et al., 2020a; Zhu et al., 2010; Zhu et al., 2016) assume that the same land cover class pixels are similar in the spectral values in the reflectance image. STARFM selects pixels with similar spectral values at the prior/post-dated time, assuming these pixels belong to the same class, and use these pixels to filter the predicted reflectance image so that they have similar spectral values at the prediction time. In the water body mapping, the water pixels with similar spectral values at t_0 are usually predicted with similar spectral values at t_p by STARFM. Since pixels with similar spectral values are usually classified into the same water or land class using the thresholding method in STARFM, the shape of the water body predicted by STARFM is usually intrinsic in the prior/post-dated Landsat image. In comparison, for STPSA and most spatiotemporal superresolution mapping methods (Li et al., 2017; Li et al., 2019b; Wang et al., 2016), the spatial distribution of objects is mainly determined based on the spatial dependence rule. As a result, the objects are predicted with a rounded shape. The main reason is that the spatial dependence rule is suitable for objects that are larger than the size of coarse spatial resolution (MODIS pixel in this experiment) pixel, but is not suitable for mapping a heterogeneous landscape (Atkinson, 2009; Ge et al., 2009). However, in spatiotemporal surface water mapping, both STARFM and STPSA failed to incorporate topography, which has a substantial influence on water distribution. In contrast, STSWM uses the ancillary SWO map of Pekel et al. (2016) and DEM to provide topographic information. The results show that STSWM more accurately predicted the spatial distribution of water bodies in heterogeneous landscapes in the study regions.

Second, in the temporal domain, major differences were found in the predicted map outputs from the different methods with regard to water change which can be relatively abrupt in situations such as flooding. STARFM assumes the reflectance change to be gradual and models this in a linear way. Therefore, STARFM and most spatiotemporal fusion

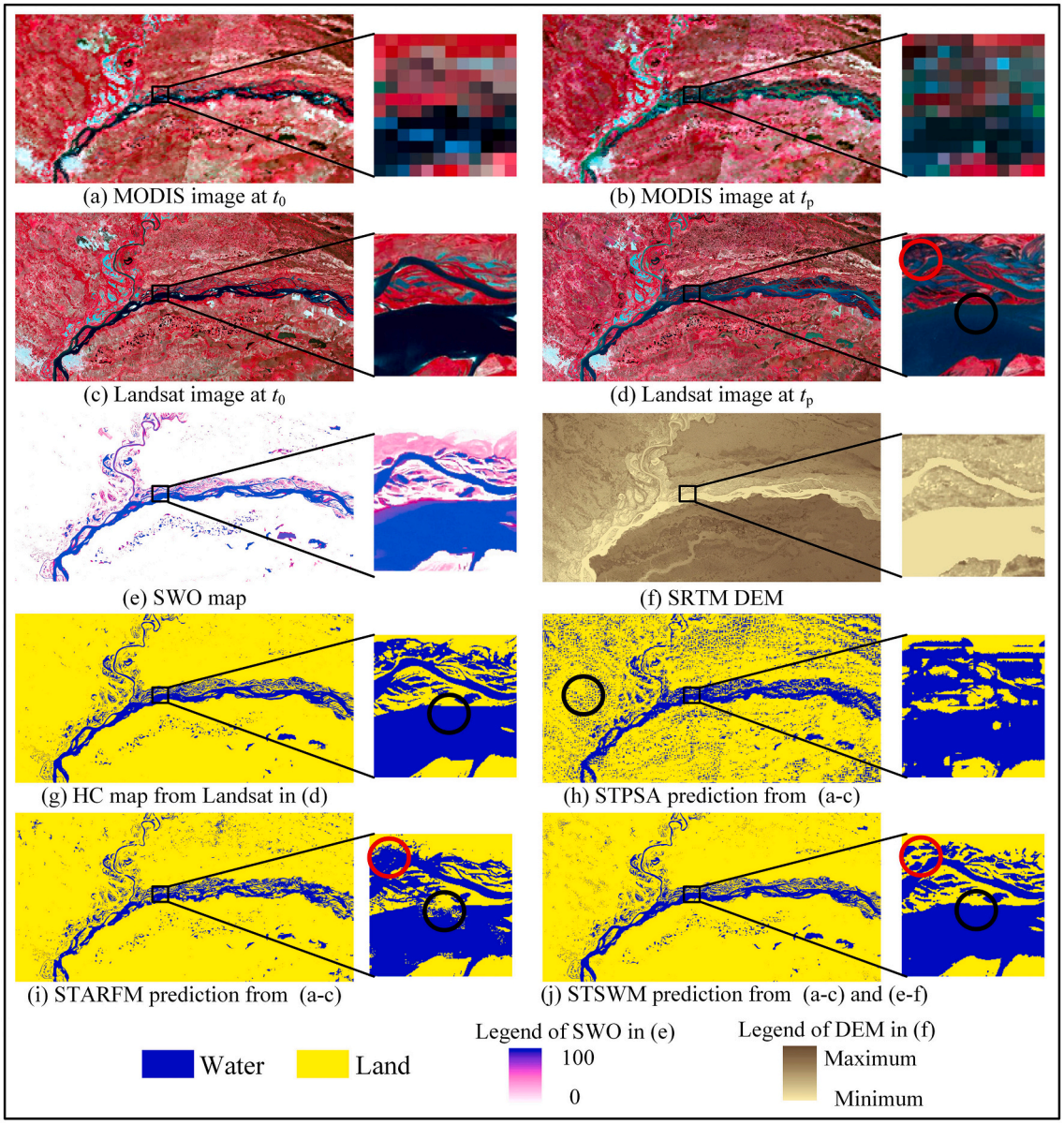

Fig. 12. Input, reference and output maps of the different methods in the Parana River experiment. (a) MODIS 8-day composite image at t_0 (starts from 2010/11/09 to 2010/11/16), (b) MODIS 8-day composite image at t_p (starts from 2010/12/11 to 2010/12/18), (c) Landsat image taken on 2010/11/13, (d) Landsat image taken on 2010/12/15, (e) SWO map, (f) DEM map, (g) HC map derived from the Landsat image in (d), (h) STPSA map, (i) STARFM map, (j) STSWM map. In the MODIS and Landsat composite images, RBG = SWIR-2, NIR, and Green bands. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

methods are good at predicting any reflectance change caused by vegetation phenology and the sensor observation angle/elevation change. However, they are limited in predicting abrupt reflectance change occurring in, for example, land cover conversions such as flooding or forest clearance (Zhu et al., 2016). Predicting abrupt reflectance change is one of the biggest challenges for existing spatiotemporal fusion methods (Zhu et al., 2018). In comparison, for STPSA, the temporal link between objects at different dates is mainly based on the temporal dependence rule, which assumes that objects are unchanged and temporally correlated. As a result, STPSA predicted relatively poor accuracy for changed land cover objects than unchanged objects (Li et al., 2019b). In contrast, STSWM is not restricted to the constraints that the reflectance change is linear or that objects at different times are temporally dependent. STSWM first determines the change of water fractions at different dates. For unchanged water fractions, the prior/post-dated surface water map at t_0 is used in surface water mapping. For changed water fractions, the SWO and DEM that provide the topographic information is used in predicting the changed water distribution. STSWM correctly predicted the abrupt water change in Figs. 10 and 12.

Although the fusion methods of STARFM and STPSA can be used in surface water mapping, they are more suitable in predicting reflectance imagery (such as STARFM) or a land cover map of multiple classes (such as STPSA), but fail to consider the spatial and temporal information about surface water. In contrast, the proposed STSWM considers the specific surface water spatial and temporal information. Results show that STSWM is more suitable in spatiotemporal surface water mapping in spatially heterogeneous landscapes and predicting abrupt surface water change.

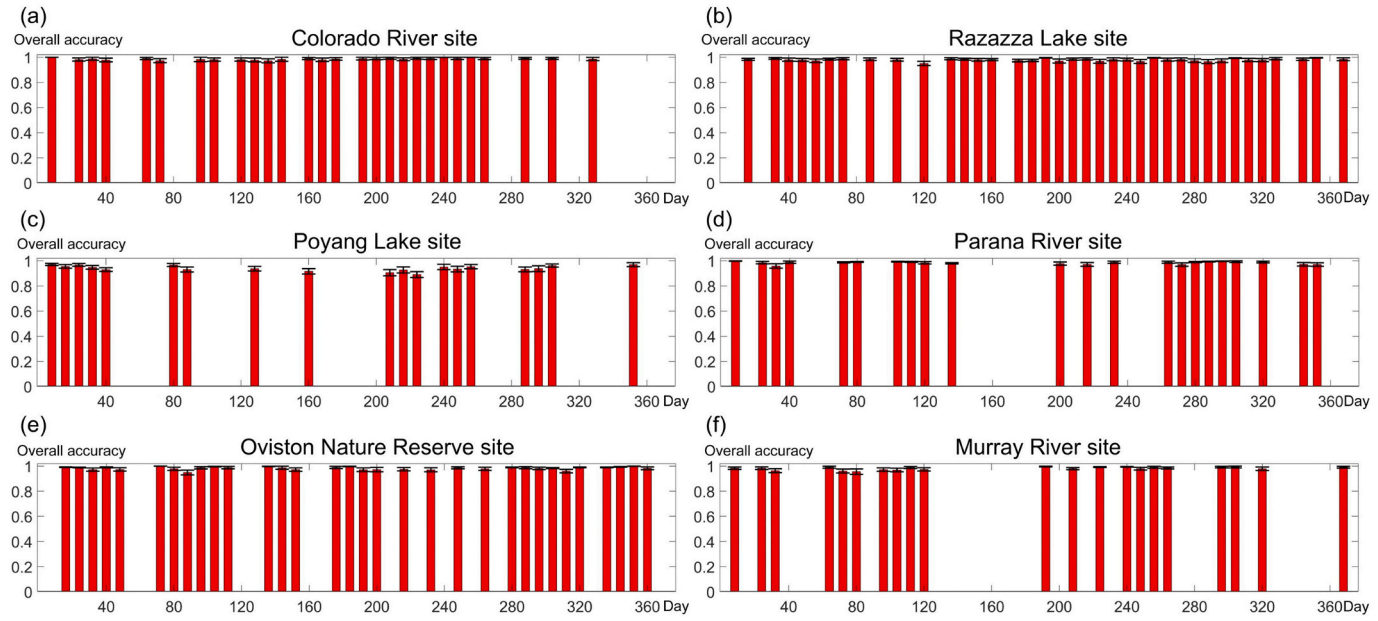

Fig. 13. The stratified estimator of overall accuracies of the STSWM maps in the six sites. The margin of error in the bar indicates the standard error. A total of 200 samples were used in the validation. The samples covered by cloud, shadow or striping in the Landsat image within the period represented by the MODIS data were excluded in estimating the accuracy of the corresponding STSWM map. The overall accuracy of an STSWM image was not estimated if more than 40% of the samples, i.e., 80 samples, were excluded.

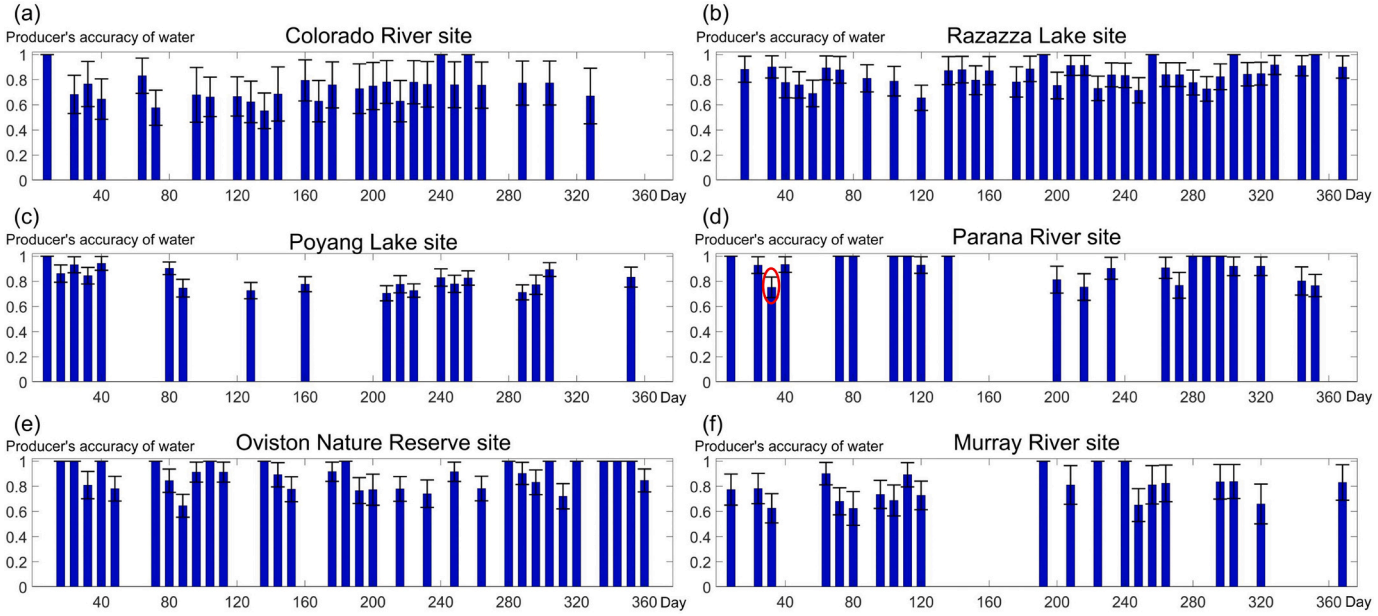

Fig. 14. The stratified estimator of producer's accuracy of water of the STSWM maps in the six sites. The margin of error in the bar indicates the standard error. A total of 200 samples were used in the validation. The samples covered by cloud, shadow or striping in the Landsat image within the period represented by the MODIS data were excluded in estimating the accuracy of the corresponding STSWM map. The producer's accuracy of water of an STSWM image was not estimated if more than 40% of the samples, i.e., 80 samples, were excluded. The red ellipse indicates the accuracy of STSWM that had a drastic surface water change during the period between the prediction time and the time of ancillary data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.2. Comparison between the STSWM and the methods that used only Landsat images

In this paper, the proposed STSWM was used to predict surface water maps at an 8-day step. Although the Landsat images can be acquired every eight days in some parts of the world using a combination of Landsat 5, 7, and 8, about 35% of Landsat images are contaminated by clouds (Ju and Roy, 2008). In contrast, the STSWM can predict the presence of surface water even for cloud pixels in the MODIS composite image by using an adaptive moving window to select similar pixels and labeling the 30 m pixels within the MODIS cloud pixels. In addition, the Landsat 7 images are affected by striping, while the STSWM is capable of selecting ancillary Landsat images that are free of striping. In the Parana River experiment, less than ten cloud-free and stripe-free Landsat images

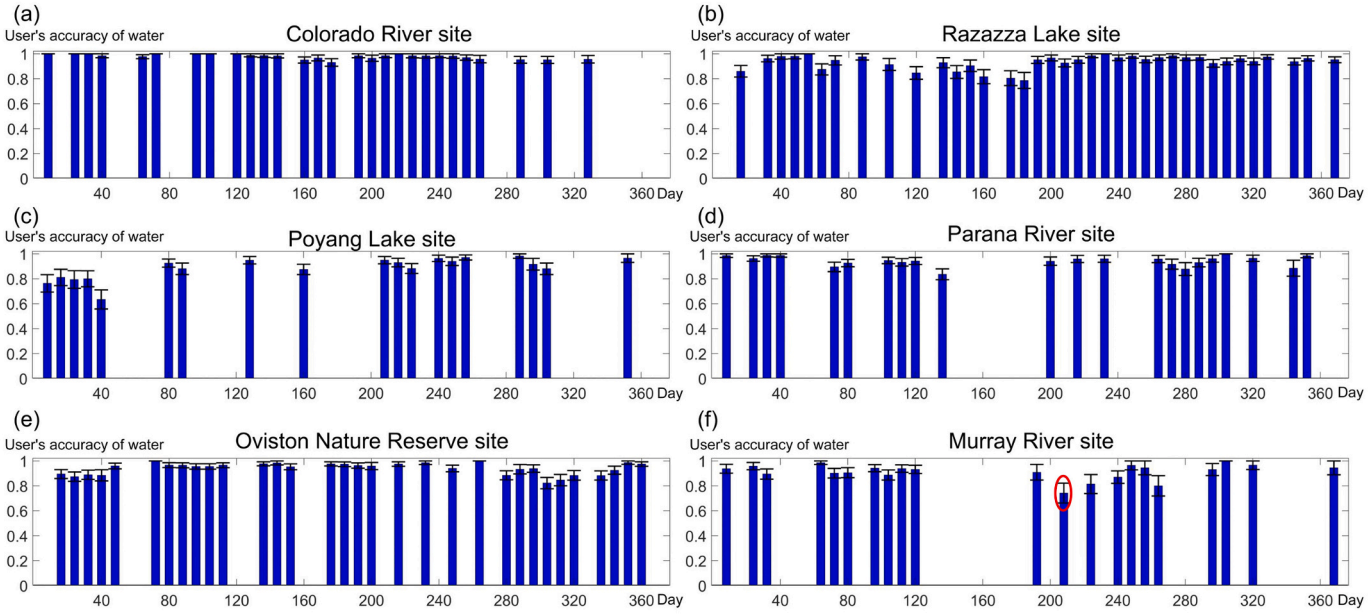

Fig. 15. The stratified estimator of user's accuracy of water of the STSWM maps in the six sites. The margin of error in the bar indicates the standard error. A total of 200 samples were used in the validation. The samples covered by cloud, shadow or striping in the Landsat image within the period represented by the MODIS data were excluded in estimating the accuracy of the corresponding STSWM map. The user's accuracy of water of an STSWM image was not estimated if more than 40% of the samples, i.e., 80 samples, were excluded. The red ellipse indicates the accuracy of STSWM that had a drastic surface water change during the period between the prediction time and the time of ancillary data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

were available in the chosen year, which affected the monitoring of the surface water at a high temporal resolution. The STSWM was able to predict 30 m surface water maps at an 8-day time step, demonstrating the advantage of fusing MODIS and Landsat data when creating high spatiotemporal resolution water maps.

The STSWM was compared with the GSW monthly water history maps for the study areas. The GSW datasets were affected by clouds, stripping, and the availability of Landsat data in some months. For instance, there is no Landsat image of the zoomed-in area in Fig. 9(f) for June in the Parana River experiment. In contrast, the STSWM provided water maps with a denser temporal frequency, which is helpful for monitoring surface water dynamics.

5.3. Limitations and future research

Although STSWM can predict water maps for cloud and shadow contaminated regions, a higher area percentage of cloud and shadow in the MODIS scene would reduce the accuracy of the STSWM maps. The detection of cloud and shadow from MODIS images contains uncertainty. If the cloud and shadow are not detected in the MODIS image, STSWM would probably unmix the shadow MODIS pixels as water area since water and shadow are usually spectrally similar (Li et al., 2013b). For instance, the water extent highlighted in the black circles in Fig. 10 is actually the result of shadow in the MODIS image. The high percentage of cloud and shadow would probably result in high commission error of water. This is found in Fig. 15 in the quantitative results, in which STSWM generated a relatively lower user's accuracy of water in the Poyang Lake site with the highest cloud percentage shadow in the six sites. In addition, if the cloud and shadow are correctly detected in the MODIS image, the uncertainty brought by the cloud and shadow is not negligible in water mapping from STSWM. STSWM uses prior information about water distribution to map water for cloud and shadow MODIS pixels. The STSWM prior information, i.e., the water level should be relatively uniform for inundated surface water areas in a local region, may not suit if the cloud and shadow extent is too large. In this case, STSWM has to search for local similar pixels that are spatially far away (Section 2.1), and the uncertainty in STSWM increases correspondingly. In general, although the water maps from STSWM are temporally more dense than those produced from only Landsat data such as the GSW (Pekel et al., 2016) and Global Land Analysis & Discovery (GLAD) (Pickens et al., 2020) data in which the cloud and shadow pixels are viewed as bad data, the uncertainty brought by the cloud and shadow in STSWM may result in high commission error of water than the GSW and GLAD data. Advanced methods to correct the cloud and shadow contaminated water maps, such as the method based on the histogram of the occurrence image (Zhao and Gao, 2018), should be developed to further enhance the STSWM model.

The STSWM performance is affected by the SWO data that provides bathymetry information. STSWM used the SWO from GSW layers produced according to the 30 years of Landsat data to represent the bathymetry. The bathymetry could change due to the land cover type changes caused by the activities such as the deliberate human river diversion, the building of man-made reservoirs, and the lake's transformation to land. In such circumstances, the SWO data produced from historical Landsat observations may be improper to represent the bathymetry used in STSWM. In addition, the SWO data used in this paper is produced from the JRC GSW monthly water history maps, which represent if water is detected in a pixel in a month. Other SWO data may introduce different information about surface water occurrence information for STSWM. For instance, the GLAD global surface water dynamics data provides the month water percent layer calculated each month for 1999–2018 (https://www.glad.umd.edu/dataset/global-surface-water-dynamics), which could be explored in STSWM in the future. The GLAD data also mapped surface waters in high latitudes in Dec. and Jan., which could provide different bathymetry information than the JRC GSW data for STSWM.

The performance of the STSWM is also affected by the accuracy of water fraction images from spectral unmixing. First, only the mean spectral values of each sub-class were used as the endmembers. Second, the linear spectral unmixing was applied to the MODIS image for simplicity and computational efficiency. The spectral response from different classes could be mixed non-linearly, and the potential of other

unmixing algorithms could be explored to reduce the spectral unmixing error in the production of the water fraction image (Yuan et al., 2020). In addition, the STSWM is also affected by the accuracy of the 30 m surface water map derived from the ancillary Landsat data at t_0. In this study, the 30 m surface water map was derived using an automatic method which first derived an MNDWI map from the Landsat data and then used Otsu thresholding to identify the water and land. In addition to the MNDWI, the use of a different water indexes in the STSWM model could be explored in the context of water extraction from the Landsat image (Fisher et al., 2016). In addition, the STSWM model did not use any ancillary information about land covers in the study region. If there are available training samples about water and other land covers in the Landsat image at t_0, more advanced classification methods based on machine learning and deep learning can also be applied to generate the surface water map at t_0 (Isikdogan et al., 2017).

In this study, the inputs of the proposed STSWM only include one pair of MODIS and Landsat images. This ensures that the STSWM can monitor flooding and inundation at a nearly real-time frequency using one pair of MODIS and Landsat images acquired temporally close to and pre-dating the prediction period. The STSWM can also be extended to include two pairs of MODIS-Landsat images, one acquired before and the other acquired after the prediction date. In the solution, the STSWM can use both the surface water maps generated from the pre- and post-dated Landsat images to spatially filter the initial surface water map using the similar water likelihood pixels, assuming the surface water maps from each Landsat image can provide complementary surface water distribution information.

6. Conclusions

In this paper, a new STSWM method that uses a pair of Landsat-MODIS images products and SWO and DEM data was proposed to generate 30 m spatial resolution surface water maps at an 8-day time step. Unlike other spatiotemporal superresolution land cover mapping algorithms used for multi-class mapping and spatiotemporal image fusion algorithms used to generate Landsat-like reflectance images, the STSWM directly maps the 30 m resolution surface water data and its changes. The proposed method both visually and quantitatively outperforms STPSA and STARFM methods. The key to its performance is the inclusion of SWO and DEM data, which provide the topographic information required for predicting the spatial distribution of the water. The STSWM could predict water maps even though cloud cover was present in the MODIS images using local topology information, while the STPSA and STARFM could not map the surface water in cloudy regions.

STSWM incorporated the high temporal frequency MODIS data, and generated temporally more continuous surface water maps than those based only on Landsat data. In contrast to the GSW monthly water history maps, which may fail to detect the surface water changes due to the presence of cloud cover and the lack of available Landsat data, the STSWM was able to predict 30 m surface water maps at an 8-day time step that more clearly represent the surface water dynamics. The STSWM is especially capable of monitoring the changes in the surface water in heterogeneous landscapes. It has a great potential for increasing our understanding of the surface water dynamics in large areas at a fine spatiotemporal resolution, which would facilitate water resources management.

Credit author statement

Xiaodong Li: Conceptualization, Methodology, Data curation, Formal analysis, Writing - Original Draft, Project administration.
Feng Ling: Conceptualization, Writing - Review & Editing, Funding acquisition.
Giles M. Foody: Conceptualization, Writing - Review & Editing,
Doreen S. Boyd: Conceptualization, Writing - Review & Editing.
Lai Jiang: Data curation, Funding acquisition.
Yihang Zhang: Data curation, Funding acquisition.
Pu Zhou: Data curation, Software.
Yalan Wang: Data curation, Software.
Rui Chen: Data curation, Software.
Yun Du: Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported in part by the Natural Science Foundation of China (62071457, 41801292), in part by the Application Foundation Frontier project of Wuhan (2020020601012283), in part by the key scientific research projects of water conservancy in Hubei Province, China (HBSLKY202103), in part by the technology innovation special project of Hubei Province (2019ACA155), in part by the Innovation Group Project of Hubei Natural Science Foundation (2019CFA019), in part by the Key Research Program of Frontier Sciences, Chinese Academy of Sciences (ZDBS-LY-DQC034), and in part by the Youth Innovation Promotion Association CAS (2017384). The authors are grateful to the anonymous reviewers for their comments on this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2021.112680.

References

Allen, G.H., Pavelsky, T.M., 2015. Patterns of river width and surface area revealed by the satellite-derived North American River Width data set. Geophys. Res. Lett. 42, 395–402. https://doi.org/10.1002/2014gl062764.
Armon, M., Dente, E., Shmilovitz, Y., Mushkin, A., Cohen, T.J., Morin, E., Enzel, Y., 2020. Determining bathymetry of shallow and ephemeral desert lakes using satellite imagery and altimetry. Geophys. Res. Lett. 47 https://doi.org/10.1029/2020gl087367.
Assuncao, J., Gandour, C., Pessoa, P., Rocha, R., 2017. Property-level assessment of change in forest clearing patterns: The need for tailoring policy in the Amazon. Land Use Policy 66, 18–27. https://doi.org/10.1016/j.landusepol.2017.04.022.
Atkinson, P.M., 2009. Issues of uncertainty in super-resolution mapping and their implications for the design of an inter-comparison study. Int. J. Remote Sens. 30, 5293–5308. https://doi.org/10.1080/01431160903131034.
Bates, P.D., Marks, K.J., Horritt, M.S., 2003. Optimal use of high-resolution topographic data in flood inundation models. Hydrol. Process. 17, 537–557. https://doi.org/10.1002/hyp.1113.
Bates, P.D., Neal, J.C., Alsdorf, D., Schumann, G.J.P., 2014. Observing global surface water flood dynamics. Surv. Geophys. 35, 839–852. https://doi.org/10.1007/s10712-013-9269-4.
Belgiu, M., Stein, A., 2019. Spatiotemporal Image Fusion in Remote Sensing. Remote Sens-Basel. 11 https://doi.org/10.3390/rs11070818.
Berry, P.A.M., Garlick, J.D., Smith, R.G., 2007. Near-global validation of the SRTM DEM using satellite radar altimetry. Remote Sens. Environ. 106, 17–27. https://doi.org/10.1016/j.rse.2006.07.011.
Busker, T., de Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., Pekel, J.F., Cottam, A., 2019. A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. Hydrol. Earth Syst. Sc. 23, 669–690. https://doi.org/10.5194/hess-23-669-2019.
Chen, B., Chen, L., Huang, B., Michishita, R., Xu, B., 2018. Dynamic monitoring of the Poyang Lake wetland by integrating Landsat and MODIS observations. ISPRS J. Photogramm. 139, 75–87. https://doi.org/10.1016/j.isprsjprs.2018.02.021.
Cook, A., Merwade, V., 2009. Effect of topographic data, geometric configuration and modeling approach on flood inundation mapping. J. Hydrol. 377, 131–142. https://doi.org/10.1016/j.jhydrol.2009.08.015.
Dao, P., Mong, N., Chan, H., 2019. Landsat-MODIS image fusion and object-based image analysis for observing flood inundation in a heterogeneous vegetated scene. Gisci. Remote Sens. 56, 1148–1169. https://doi.org/10.1080/15481603.2019.1627062.
Deng, C., Zhu, Z., 2020. Continuous subpixel monitoring of urban impervious surface using Landsat time series. Remote Sens. Environ. *238* https://doi.org/10.1016/j.rse.2018.10.011.
Dietz, A.J., Klein, I., Gessner, U., Frey, C.M., Kuenzer, C., Dech, S., 2017. Detection of water bodies from AVHRR data——a TIMELINE thematic processor. Remote Sens. 9 https://doi.org/10.3390/rs9010057.

Du, Z., Bin, L., Ling, F., Li, W., Tian, W., Wang, H., Gui, Y., Sun, B., Zhang, X., 2012. Estimating surface water area changes using time-series Landsat data in the Qingjiang River Basin. China. *J. Appl. Remote Sens.* 6 https://doi.org/10.1117/1.Jrs.6.063609.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The shuttle radar topography mission. Rev. Geophys. 45 https://doi.org/10.1029/2005rg000183.

Fisher, P., 1997. The pixel: A snare and a delusion. Int. J. Remote Sens. 18, 679–685. https://doi.org/10.1080/014311697219015.

Fisher, A., Flood, N., Danaher, T., 2016. Comparing Landsat water index methods for automated water classification in eastern Australia. Remote Sens. Environ. 175, 167–182. https://doi.org/10.1016/j.rse.2015.12.055.

Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. IEEE T. Geosci. Remote 44, 2207–2218. https://doi.org/10.1109/tgrs.2006.872081.

Ge, Y., Li, S., Lakhan, V.C., 2009. Development and testing of a subpixel mapping algorithm. IEEE T. Geosci. Remote 47, 2155–2164. https://doi.org/10.1109/tgrs.2008.2010863.

Ge, Y., Chen, Y., Stein, A., Li, S., Hu, J., 2016. Enhanced subpixel mapping with spatial distribution patterns of geographical objects. IEEE T. Geosci. Remote 54, 2356–2370. https://doi.org/10.1109/tgrs.2015.2499790.

Getirana, A., Jung, H.C., Tseng, K.H., 2018. Deriving three dimensional reservoir bathymetry from multi-satellite datasets. Remote Sens. Environ. 217, 366–374. https://doi.org/10.1016/j.rse.2018.08.030.

He, D., Zhong, Y., Zhang, L., 2020. Spectral-spatial-temporal MAP-based sub-pixel mapping for land-cover change detection. IEEE T. Geosci. Remote 58, 1696–1717. https://doi.org/10.1109/tgrs.2019.2947708.

Heimhuber, V., Tulbure, M.G., Broich, M., 2018. Addressing spatio-temporal resolution constraints in Landsat and MODIS-based mapping of large-scale floodplain inundation dynamics. Remote Sens. Environ. 211, 307–320. https://doi.org/10.1016/j.rse.2018.04.016.

Holgerson, M.A., Raymond, P.A., 2016. Large contribution to inland water CO2 and CH4 emissions from very small ponds. Nat. Geosci. 9, 222–U150. https://doi.org/10.1038/ngeo2654.

Huang, C., Chen, Y., Wu, J.P., 2014. DEM-based modification of pixel-swapping algorithm for enhancing floodplain inundation mapping. Int. J. Remote Sens. 35, 365–381. https://doi.org/10.1080/01431161.2013.871084.

Isikdogan, F., Bovik, A.C., Passalacqua, P., 2017. Surface Water Mapping by Deep Learning. IEEE J. Sel. Top. Appl. 10, 4909–4918. https://doi.org/10.1109/jstars.2017.2735443.

Ju, J., Roy, D.P., 2008. The availability of cloud-free Landsat ETM plus data over the conterminous United States and globally. Remote Sens. Environ. 112, 1196–1211. https://doi.org/10.1016/j.rse.2007.08.011.

Keshava, N., Mustard, J.F., 2002. Spectral unmixing. *IEEE Signal. Proc. Mag.* 19, 44–57. https://doi.org/10.1109/79.974727.

Li, S., Sun, D., Goldberg, M., Stefanidis, A., 2013a. Derivation of 30-m-resolution water maps from TERRA/MODIS and SRTM. Remote Sens. Environ. 134, 417–430. https://doi.org/10.1016/j.rse.2013.03.015.

Li, S., Sun, D., Yu, Y., 2013b. Automatic cloud-shadow removal from flood/standing water maps using MSG/SEVIRI imagery. Int. J. Remote Sens. 34, 5487–5502. https://doi.org/10.1080/01431161.2013.792969.

Li, L., Chen, Y., Xu, T., Liu, R., Shi, K., Huang, C., 2015a. Super-resolution mapping of wetland inundation from remote sensing imagery based on integration of back-propagation neural network and genetic algorithm. Remote Sens. Environ. 164, 142–154. https://doi.org/10.1016/j.rse.2015.04.009.

Li, X., Du, Y., Ling, F., 2015b. Sub-pixel-scale land cover map updating by integrating change detection and sub-pixel mapping. Photogramm. Eng. Rem. S. *81*, 59–67. https://doi.org/10.14358/pers.81.1.59.

Li, X., Ling, F., Foody, G.M., Ge, Y., Zhang, Y., Du, Y., 2017. Generating a series of fine spatial and temporal resolution land cover maps by fusing coarse spatial resolution remotely sensed images and fine spatial resolution land cover maps. Remote Sens. Environ. 196, 293–311. https://doi.org/10.1016/j.rse.2017.05.011.

Li, L., Skidmore, A., Vrieling, A., Wang, T., 2019a. A new dense 18-year time series of surface water fraction estimates from MODIS for the Mediterranean region. Hydrol. Earth Syst. Sc. 23, 3037–3056. https://doi.org/10.5194/hess-23-3037-2019.

Li, X., Ling, F., Foody, G.M., Ge, Y., Zhang, Y., Wang, L., Shi, L., Li, X., Du, Y., 2019b. Spatial-temporal super-resolution land cover mapping with a local spatial-temporal dependence model. IEEE T. Geosci. Remote 57, 4951–4966. https://doi.org/10.1109/tgrs.2019.2894773.

Li, Y., Gao, H., Jasinski, M.F., Zhang, S., Stoll, J.D., 2019c. Deriving high-resolution reservoir bathymetry from ICESat-2 prototype photon-counting Lidar and Landsat imagery. IEEE T. Geosci. Remote 57, 7883–7893. https://doi.org/10.1109/tgrs.2019.2917012.

Li, X., Foody, G.M., Boyd, D.S., Ge, Y., Zhang, Y., Du, Y., Ling, F., 2020a. SFSDAF: An enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion. Remote Sens. Environ. 237. https://doi.org/10.1016/j.rse.2019.111537.

Li, Y., Gao, H., Zhao, G., Tseng, K., 2020b. A high-resolution bathymetry dataset for global reservoirs using multi-source satellite imagery and altimetry. Remote Sens. Environ. 244 https://doi.org/10.1016/j.rse.2020.111831.

Lin, P., Pan, M., Allen, G.H., de Frasson, R.P., Zeng, Z., Yamazaki, D., Wood, E.F., 2020. Global estimates of reach-level bankfull river width leveraging big data geospatial analysis. Geophys. Res. Lett. 47 https://doi.org/10.1029/2019gl086405.

Ling, F., Xiao, F., Du, Y., Xue, H., Ren, X., 2008. Waterline mapping at the subpixel scale from remote sensing imagery with high-resolution digital elevation models. Int. J. Remote Sens. 29, 1809–1815. https://doi.org/10.1080/01431160701802489.

Ling, F., Li, W., Du, Y., Li, X., 2011. Land cover change mapping at the subpixel scale with different spatial-resolution remotely sensed imagery. IEEE Geosci. Remote S. 8, 182–186. https://doi.org/10.1109/lgrs.2010.2055034.

Ling, F., Boyd, D., Ge, Y., Foody, G.M., Li, X., Wang, L., Zhang, Y., Shi, L., Shang, C., Li, X., Du, Y., 2019. Measuring river wetted width from remotely sensed imagery at the subpixel scale with a deep convolutional neural network. Water Resour. Res. 55, 5631–5649. https://doi.org/10.1029/2018wr024136.

Ling, F., Li, X., Foody, G.M., Boyd, D., Ge, Y., Li, X., Du, Y., 2020. Monitoring surface water area variations of reservoirs using daily MODIS images by exploring sub-pixel information. ISPRS J. Photogramm. 168, 141–152. https://doi.org/10.1016/j.isprsjprs.2020.08.008.

Luo, S., Song, C., Liu, K., Ke, L., Ma, R., 2019. An effective low-cost remote sensing approach to reconstruct the long-term and dense time series of area and storage variations for large lakes. Sensors 19. https://doi.org/10.3390/s19194247.

Muad, A.M., Foody, G.M., 2012a. Impact of land cover patch size on the accuracy of patch area representation in HNN-based super resolution mapping. IEEE J. Sel. Top. Appl 5, 1418–1427. https://doi.org/10.1109/jstars.2012.2191145.

Muad, A.M., Foody, G.M., 2012b. Super-resolution mapping of lakes from imagery with a coarse spatial and fine temporal resolution. Int. J. Appl. Earth Obs. 15, 79–91. https://doi.org/10.1016/j.jag.2011.06.002.

Mueller, N., Lewis, A., Roberts, D., Ring, S., Melrose, R., Sixsmith, J., Lymburner, L., McIntyre, A., Tan, P., Curnow, S., Ip, A., 2016. Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia. Remote Sens. Environ. 174, 341–352. https://doi.org/10.1016/j.rse.2015.11.003.

Ogilvie, A., Belaud, G., Massuel, S., Mulligan, M., Le Goulven, P., Calvez, R., 2018. Surface water monitoring in small water bodies: potential and limits of multi-sensor Landsat time series. Hydrol. Earth Syst. Sc. 22, 4349–4380. https://doi.org/10.5194/hess-22-4349-2018.

Olofsson, P., Foody, G.M., Stehman, S.V., Woodcock, C.E., 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sens. Environ. 129, 122–131. https://doi.org/10.1016/j.rse.2012.10.031.

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

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE T. Syst. Man. Cy-S. 9, 62–69.

Pekel, J.F., Vancutsem, C., Bastin, L., Clerici, M., Vanbogaert, E., Bartholome, E., Defourny, P., 2014. A near real-time water surface detection method based on HSV transformation of MODIS multi-spectral time series data. Remote Sens. Environ. 140, 704–716. https://doi.org/10.1016/j.rse.2013.10.008.

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

Pickens, A.H., Hansen, M.C., Hancher, M., Stehman, S.V., Tyukavina, A., Potapov, P., Marroquin, B., Sherani, Z., 2020. Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series. Remote Sens. Environ. 243 https://doi.org/10.1016/j.rse.2020.111792.

Stehman, S.V., 2009. Sampling designs for accuracy assessment of land cover. Int. J. Remote Sens. 30, 5243–5272. https://doi.org/10.1080/01431160903131000.

Tan, Z., Wang, X., Chen, B., Liu, X., Zhang, Q., 2019. Surface water connectivity of seasonal isolated lakes in a dynamic lake-floodplain system. J. Hydrol. 579 https://doi.org/10.1016/j.jhydrol.2019.124154.

Tulbure, M.G., Broich, M., Stehman, S.V., Kommareddy, A., 2016. Surface water extent dynamics from three decades of seasonally continuous Landsat time series at subcontinental scale in a semi-arid region. Remote Sens. Environ. 178, 142–157. https://doi.org/10.1016/j.rse.2016.02.034.

Vorosmarty, C.J., Green, P., Salisbury, J., Lammers, R.B., 2000. Global water resources: Vulnerability from climate change and population growth. Science 289, 284–288. https://doi.org/10.1126/science.289.5477.284.

Walker, J.J., de Beurs, K.M., Wynne, R.H., Gao, F., 2012. Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. Remote Sens. Environ. 117, 381–393. https://doi.org/10.1016/j.rse.2011.10.014.

Wang, Q., Atkinson, P.M., 2018. Spatio-temporal fusion for daily Sentinel-2 images. Remote Sens. Environ. 204, 31–42. https://doi.org/10.1016/j.rse.2017.10.046.

Wang, Q., Shi, W., Atkinson, P.M., 2016. Spatiotemporal subpixel mapping of time-series images. IEEE T. Geosci. Remote 54, 5397–5411. https://doi.org/10.1109/tgrs.2016.2562178.

Wang, Q., Tang, Y., Tong, X., Atkinson, P.M., 2020. Virtual image pair-based spatio-temporal fusion. Remote Sens. Environ. 249. https://doi.org/10.1016/j.rse.2020.112009.

Xiao, F., Cheng, W., Zhu, L., Feng, Q., Du, Y., 2018. Downscaling MODIS-derived water maps with high-precision topographic data in a shallow lake. Int. J. Remote Sens. 39, 7846–7860. https://doi.org/10.1080/01431161.2018.1474529.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27, 3025–3033. https://doi.org/10.1080/01431160600589179.

Xu, Y., Huang, B., . A spatio-temporal pixel-swapping algorithm for subpixel land cover mapping. IEEE Geosci. Remote S. *11*, *474–478*. https://doi.org/10.1109/lgrs.2013.2268153.

Yang, J., Huang, X., Tang, Q.H., 2020a. Satellite-derived river width and its spatiotemporal patterns in China during 1990-2015. Remote Sens. Environ. 247 https://doi.org/10.1016/j.rse.2020.111918.

Yang, X., Chen, Y., Wang, J., 2020b. Combined use of Sentinel-2 and Landsat 8 to monitor water surface area dynamics using Google Earth Engine. Remote Sens. Lett. 11, 687–696. https://doi.org/10.1080/2150704x.2020.1757780.

Yao, F., Wang, J., Wang, C., Cretaux, J., 2019a. Constructing long-term high-frequency time series of global lake and reservoir areas using Landsat imagery. Remote Sens. Environ. 232 https://doi.org/10.1016/j.rse.2019.111210.

Yao, F., Wang, J., Wang, C., Cretaux, J.F., 2019b. Constructing long-term high-frequency time series of global lake and reservoir areas using Landsat imagery. Remote Sens. Environ. 232 https://doi.org/10.1016/j.rse.2019.111210.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241 https://doi.org/10.1016/j.rse.2020.111716.

Zhang, L., Wu, K., Zhong, Y., Li, P., 2008. A new sub-pixel mapping algorithm based on a BP neural network with an observation model. Neurocomputing 71, 2046–2054. https://doi.org/10.1016/j.neucom.2007.08.033.

Zhang, F., Zhu, X., Liu, D., 2014. Blending MODIS and Landsat images for urban flood mapping. Int. J. Remote Sens. 35, 3237–3253. https://doi.org/10.1080/01431161.2014.903351.

Zhao, G., Gao, H.L., 2018. Automatic Correction of Contaminated Images for Assessment of Reservoir Surface Area Dynamics. Geophys. Res. Lett. 45, 6092–6099. https://doi.org/10.1029/2018gl078343.

Zhu, X., Helmer, E.H., 2018. An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions. Remote Sens. Environ. 214, 135–153. https://doi.org/10.1016/j.rse.2018.05.024.

Zhu, Z., Woodcock, C.E., 2014. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change. Remote Sens. Environ. 152, 217–234. https://doi.org/10.1016/j.rse.2014.06.012.

Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. Remote Sens. Environ. 114, 2610–2623. https://doi.org/10.1016/j.rse.2010.05.032.

Zhu, X., Helmer, E.H., Gao, F., Liu, D., Chen, J., Lefsky, M.A., 2016. A flexible spatiotemporal method for fusing satellite images with different resolutions. Remote Sens. Environ. 172, 165–177. https://doi.org/10.1016/j.rse.2015.11.016.

Zhu, X., Cai, F., Tian, J., Williams, T.K.A., 2018. Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions. Remote Sens-Basel. 10 https://doi.org/10.3390/rs10040527.

Zou, Z., Xiao, X., Dong, J., Qin, Y., Doughty, R., Menarguez, M.A., Zhang, G., Wang, J., 2018. Divergent trends of open-surface water body area in the contiguous United States from 1984 to 2016. P. Natl. Acad. Sci. USA 115, 3810–3815. https://doi.org/10.1073/pnas.1719275115.