An approach for global monitoring of surface water extent variations in reservoirs using MODIS data

Ankush Khandelwal [a,1], Anuj Karpatne [a,1], Miriam E. Marlier [b,1], Jongyoun Kim [b], Dennis P. Lettenmaier [b], Vipin Kumar [a,*]

[a] *Department of Computer Science and Engineering, University of Minnesota, 4-192 Keller Hall, Minneapolis, MN 55455, United States*
[b] *Department of Geography, University of California, Los Angeles, 1255 Bunche Hall, Los Angeles, CA 90095, United States*

ARTICLE INFO

ABSTRACT

Freshwater resources are among the most basic requirements of human society. Nonetheless, global information about the space-time variations of the area of freshwater bodies, and the water stored in them, is surprisingly limited. We introduce a new approach that uses MODIS multispectral data to map the global areal extent and temporal variations of known reservoirs at 500 m spatial resolution at nominal eight-day intervals from 2000 to 2015. We evaluate the performance of the approach on 94 reservoirs by comparing the variations in surface area extents with satellite radar altimetry measurements. Furthermore, we present detailed case studies for five reservoirs on four continents to demonstrate the impact of different challenges on the performance of the approach. For three of these case studies, we also evaluate surface area estimates using Landsat-based surface extent reference maps. Altimetry based height measurements for the 94 reservoirs show higher correlation with surface area computed using our approach, compared to surface area computed using previous approaches. One of the main reasons for these improvements is a novel post-processing technique that makes use of imperfect labels produced by supervised classification approaches on multiple dates to estimate the elevation structure of locations and uses it to enhance the quality and completeness of imperfect labels. However, effective estimation of this elevation structure requires that the water body shows sufficient area variations. Hence, the post-processing approach will not be effective for water bodies that are mostly unchanged or are too small to have sufficient variation. The approach is still challenged in regions with frequent cloud cover, snow and ice coverage, or complicated geometries that will require remote sensing data at finer spatial resolution. The surface area estimates we describe here are publically available.

© 2017 Published by Elsevier Inc.

1. Introduction

Despite the importance of information about the space-time variations and long-term trends in surface freshwater resources, current understanding at the global scale is remarkably limited (Döll et al., 2015). Improved monitoring of freshwater resources is critical for understanding both anthropogenic impacts and climate change influences (Haddeland et al., 2014), and for better representation of water management effects in Earth system models. Remote sensing data have increasingly been used to monitor variations in water bodies in space and time (Lettenmaier et al., 2015), and these data are especially valuable (and usually the only viable option) in parts of the world where in situ monitoring networks are lacking (Alsdorf et al., 2007).

Existing approaches for monitoring the surface extent of water bodies are limited in geographic scope, temporal extent of the record, and/or temporal frequency of observations. As summarized below, most current remote sensing-based products rely on spectral indices derived from multispectral data to distinguish surface water bodies from surrounding land features, as well as digital elevation information in some cases to address the effect of varied topography on classification algorithms. At the global scale, several surface water extent products currently are available. For example, Carroll et al. (2009) combined the Shuttle Radar Topography Missions (SRTM) Water Body Dataset (SWBD) with 250 m MODerate Resolution Imaging Spectroradiometer (MODIS) reflectance data to produce a global static map of surface water for circa 2000. Feng et al. (2014) recently used 30 m Landsat reflectance data and elevation information to produce the Global Inland Water (GIW) dataset for 2000. Similarly, the GLObal WAter BOdies database (GLOWABO) inventories global lakes for 2000, based on Landsat and SRTM data (Verpoorter et al., 2014). Liao et al. (2014) produced global 30 m water extent maps using data from Landsat and China's

---

* Corresponding author.
 E-mail address: kumar001@umn.edu (V. Kumar).
[1] Co-first authors.

HJ-1 satellite for 2000 and 2010, incorporating a classification methodology that combined pixel-based algorithms, an objected-oriented approach, and elevation. Sheng et al. 2016 produced a static lake extent map for 2015 based on Landsat 8 data by using a new mapping and compositing strategy that aims to address seasonal variability issues in lake mapping at continental and global scales. Yamazaki et al. 2015 developed a static Global 3 arc-second Water Body Map (G3WBM) using an automated algorithm to process 33,890 Landsat images from the Global Land Survey (GLS) database. Sharma et al. (2015) proposed a new water index and produces a static water extent map for 2013 using MODIS data. One common characteristic of all of these methods is that they include limited or no information about temporal variability. At the continental scale, Pekel et al. (2014) used MODIS surface reflectance and land surface temperature to map water bodies across Africa every ten days for 2004–2010 in near real-time. Their method is based on transformation of the RGB color space to HSV (Hue, Saturation, and Value) and uses an extensive training dataset, which the authors note is generalizable to the global scale and other satellites. Haas et al. (2009) produced 1 km^2 spatial resolution maps for sub-Saharan western Africa every ten days for 1999–2007 based on the SPOT VGT sensor small water body (SWB) information. Lu et al. (2011) mapped water bodies using the Chinese HJ-1A/B satellites (available over Asia since 2008) at 30 m spatial resolution based on NDVI and NDWI spectral indices, slope information to remove shadows, and the near-infrared (NIR) band to reduce urban effects. They also showed that NDVI – NDWI is superior to a single index of NDVI or NDWI for enhancing contrast between water bodies and surrounding land areas. Fisher et al. (2016) compared performance of different water indices across different sensors for the eastern Australia region.

Other promising approaches, primarily based on MODIS or Landsat surface reflectance data, have been demonstrated for specific locations. For example, Deus and Gloaguen (2013) used MODIS MNDWI to map surface extent changes of Tanzania's Lake Manyara from 2000 to 2011. Gao et al. (2012) mapped surface water extent with an unsupervised classification based on MODIS NDVI data to estimate historical area changes in 34 global reservoirs. A follow-up study by (Zhang et al., 2014) used similar methods applied to 21 reservoirs in South Asia.

At a finer spatial resolution, Bai et al. (2010) used NDWI from Landsat to calculate changes in Central Asian lakes from 1975 to 2007, with approximately decade-long gaps between estimates. Similarly, Li and Sheng (2012) used Landsat data for 1990, 2000, and 2009 to monitor changes in glacial lakes with image segmentation based on NDWI and terrain information. Jiang et al. (2014) also used Landsat indices and topography to automatically extract water bodies for several tests sites in China in the late 2000s. Sun et al. (2012) compared different methods of delineating water bodies with pixel-based and object-based methods using Landsat data in China. For three diverse case studies in Africa, North America, and Europe, a combination of Landsat MNDWI and manual interpretation was used to map surface water changes, which was later combined with satellite altimetry data to estimate storage (Duan and Bastiaanssen, 2013). Feyisa et al. (2014) showed that a Landsat-based Automated Water Extraction Index (AWEI), helped in areas with shadows or topographic effects and offered improvements over MNDWI and maximum likelihood classification methods. However, this first application was limited to five case studies and one time step. Finally, rather than using surface reflectance information as did the studies summarized above, Brisco et al. (2009) used RADARSAT-1 (C-band) to detect seasonal and annual changes of water bodies above 0.1km^2 minimum size for three test sites in Canada.

Recently, methods that aim to capture the temporal variations in the water body surface extents at global scale have been proposed. Klein et al. 2015a, 2015b developed a methodology based on the dynamic thresholding of Near Infrared (NIR) band for mapping surface water extents at daily temporal resolution using MODIS data at 250 m spatial resolution. The key idea of the approach is to adapt the threshold on NIR band value for each image using training labels from the image itself. Recent approaches using a 30-year Landsat record have produced global maps of surface water occurrence (Pekel et al., 2016; Donchyts et al., 2016). These new products can be used to analyze surface water dynamics over long time periods at fine spatial resolution, although limited to monthly temporal resolution at best, and longer in areas with substantial cloud cover. Here we introduce a global supervised classification based approach to monitor temporal variations in surface water extent of known reservoirs using 500 m resolution MODIS multispectral reflectance data. We produce water classification maps that are global in coverage, span the 15-year period 2000–2015 and estimate variations at a nominal eight-day interval. Our approach differs from previous efforts in several ways: (a) we use supervised learning approaches that take advantage of all seven MODIS spectral reflectance bands for two MODIS products, (b) we train supervised classification models to distinguish between water and land using samples from 99 representative lakes, selected from diverse regions of the world for February 2000, when reasonable quality labels were available, (c) we incorporate a recently developed approach (Khandelwal et al., 2015) that makes use of the multi-temporal imperfect classification maps for estimating the elevation structure of dynamic water bodies and uses it to correct classification errors arising due to noise, outliers, and missing values prevalent in remote sensing data, and (d) in addition to the time-varying global water maps we produce, we also provide an automated extraction method for specific water bodies based on their initial shape available from different sources. The automatic extraction step is of particular interest for use in calculation of storage variations, where relative height data (e.g. from satellite altimeters) is available independently for each water body.

A key contribution of the proposed methodology is to provide more accurate dynamics of known water bodies. The methodology requires a static spatial extent as one of the inputs (see Section 2.2.4 for details). There are various existing datasets that provide static spatial extent maps at global scale. In this paper, Global Reservoir and Dam database (GRanD) (Lehner et al., 2011a, 2011b) and Global Lakes and Wetlands Database (GLWD) (Lehner and Döll, 2004) were used to define initial spatial extents of water bodies.

The remainder of the paper is structured as follows. In Section 2.1, we describe the input remote sensing datasets and preprocessing strategies. Section 2.2 describes the classification, post-processing and water body extraction methodologies. Section 2.3 describes the datasets used for validation. Section 3.1–3.3 provide the summary of performance on 94 global reservoirs and a detailed analysis on 5 reservoirs. Section 3.4 provides a discussion on the impact of various data related challenges on the proposed approach. Finally, we summarize the utility of the algorithm at the global scale and discuss potential applications in Section 4.

2. Methods

2.1. Remote sensing datasets

We incorporated three MODIS products into our classification approach: (a) two multispectral reflectance data products as an input to the water/land classification algorithm, and (b) static water and land classification labels from the MOD44W product to train the classification model. We describe these datasets in detail below.

2.1.1. MODIS reflectance data

We used multispectral reflectance data from the MODIS instruments onboard NASA's Terra and Aqua satellites, which are publicly available via the USGS Land Processes Distributed Active Archive Center (http://lpdaac.usgs.gov). The primary product was Collection 5 MCD43A4, which is available since Feb 18th, 2000 (when Terra data became available) and has been adjusted using a bidirectional reflectance distribution function (BRDF) to simulate a nadir viewpoint (NASA, 2015b). The MCD43A4 product is based on 16 days of observations and is available in sinusoidal projection at 500 m spatial resolution and eight-day

temporal frequency. It includes seven reflectance bands covering visible, infrared, and thermal wavelengths (459–2155 nm). Because observations in these spectral bands frequently suffer from poor data quality (e.g. due to cloud or aerosol effects), the MCD43A4 product is generated using data from both the Terra and Aqua satellites such that the combined data product is of the highest possible quality. However, by ignoring poor data quality pixels, the MCD43A4 product suffers from a high degree of missing values, especially in regions with high precipitation, aerosol concentrations, or snow and ice coverage. Fig. 1 illustrates this for the western U.S., showing the fraction of pixels in the MODIS tile that has missing data in any of the seven bands (MCD43A4 is shown in blue) on any timestamp starting from Feb 18th, 2000. We can see in Fig. 1 that almost 10% of pixels in this MODIS tile have missing values in at least one of the seven bands in certain seasons of the year, and even more before Aqua data became available in 2002. This can introduce a high degree of incompleteness in classification maps.

To alleviate the issue of missing values in MCD43A4, we also used the Collection 5 MOD09A1 product, which is collected solely from the Terra satellite (NASA, 2015c). This eight-day composite product provides information for the same seven reflectance bands at 500 m spatial resolution starting on Feb 18th, 2000. It uses corrections for atmospheric composition and aerosols such that resultant data are intended to reflect ground level irradiances in the absence of atmospheric scattering or absorption. Although the MOD09A1 product has a considerably lower number of missing values than MCD43A4 (Fig. 1, red line), it is generally less reliable because it has not been adjusted using a bidirectional reflectance distribution function (to model the values as if they were taken from nadir view) as in MCD43A4. As we will discuss later, we combined the MCD43A4 and MOD09A1 products to compensate for the primary limitations of each product, the noise and number of missing values. For the rest of the paper, MCD43A4 product will be referred to as NBAR product and MOD09A1 will be referred to as REFL product.

Apart from the seven reflectance bands, we also used quality flags to filter out pixels with snow, ice, or clouds. For the REFL product, information about the data quality is available along with the multispectral values in the 16-bit quality assessment (QA) state flags. For the NBAR product, the quality flags are available as a separate product (MCD43A2).

2.1.2. MOD44W land/water mask

We used static water extent masks from the MODIS MOD44W product to train the supervised classification models to distinguish between land and water bodies (Carroll et al., 2009; NASA, 2015a). The MOD44W product builds primarily upon the Shuttle Radar Topography Mission's (SRTM) Water Body Dataset (SWBD) (NASA, 2005), which mapped all water bodies from 60°S to 60°N at 250 m spatial resolution, but for the short duration of 11 days in February 2000. The MOD44W product combined MODIS 250 m reflectance data with SWBD; reflectance data only was used poleward of 60°N. The SWBD data suffer from inaccuracies in some regions (especially in high altitude regions that were covered by snow/ice in February 2000), but the cloud-penetrating capabilities of SRTM (C-band, 5.6 cm) (Farr et al., 2007) offer an advantage over reflectance data alone for mapping surface water. Nonetheless, the MOD44W product provides a reasonable approximation of the surface extent of water bodies and can be used for training classification models. Although there are other high-resolution global masks of land/water (e.g. Feng et al., 2014), they are generally available at annual scales and do not provide exact information about their date, making them difficult to use for training. The MOD44W product was aggregated from 250 m to 500 m to match the other MODIS products. Specifically, if a 500 m pixel has all of its 4 pixels at 250 m labeled as water in MOD44W product, then we consider the pixel as a water pixel. On the other hand, if a 500 m pixel has all of its 4 pixels labeled as land, we consider the pixel as a land pixel. Partial pixels are excluded from the training set pool.

2.2. Classification methodology

Fig. 2 provides a schematic overview of our classification methodology: 1) supervised classification, 2) merging of classification maps, and 3) post-classification label correction. We describe each of these steps in the following subsections.

2.2.1. Supervised classification

Our supervised classification approach assigned a binary label of water or land to every 500 m pixel every eight days, using the multispectral values observed at that location and time. The binary (water/land) classification maps over time were then utilized to monitor surface water extent variations.

To construct the training set for training the binary classifier, we used a set of 99 global lakes and reservoirs from the MOD44W dataset, which were selected to represent a variety of heterogeneous land and water categories along with relatively low numbers of missing values in the corresponding Feb 18th, 2000 MODIS products. The locations of
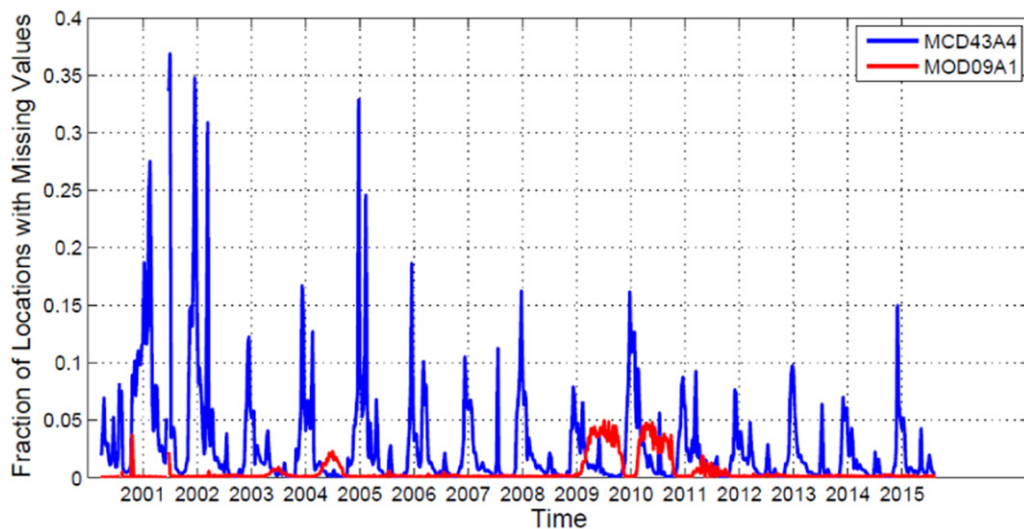

Fig. 1. Time series of the fraction of pixels (at 500 m resolution) in the MODIS tile h08v05 (in California, USA) with a missing value in any of the seven bands for any date from 2000 to 2015. The blue curve shows the fraction of missing values for the MCD43A4 product, while the red curve shows the fraction of missing values for the MOD09A1 product.
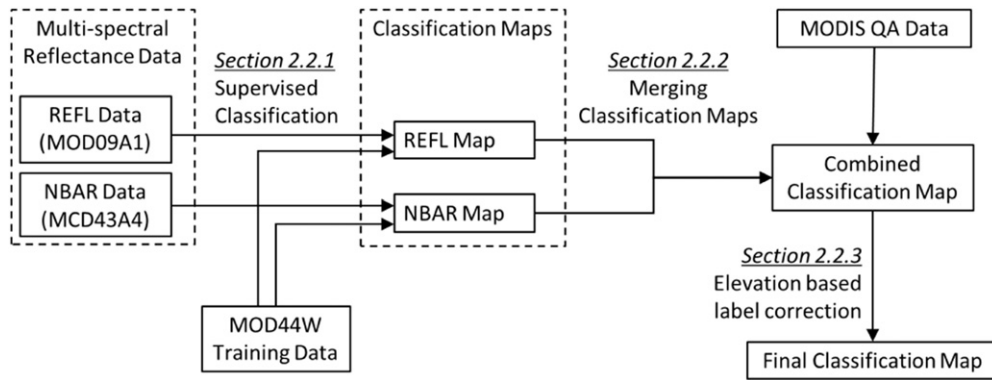
Fig. 2. Overview of the classification methodology.

the 99 water bodies from across the globe are shown in Fig. S1 in the Supplementary material. We also performed visual inspection of multi-spectral images corresponding to Feb 18th, 2000 to ensure the accuracy of the MOD44W dataset for these global 99 water bodies. Then for each water body, we created a buffer region of 20 pixels at 500 m resolution around the periphery of the water body, and used the pixels in the buffer region as well as the interior of every water body to construct the global dataset of water and land pixels. This resulted in a training pool of approximately 1.3 million pixels (after removing pixels that were at the immediate boundaries of the water bodies), where every pixel had an associated binary label of water or land, as obtained from the MOD44W dataset.

Next, we trained separate classification models for the REFL and NBAR products. For each product, we randomly sampled 5000 water and 5000 land pixels from the pool of training pixels to construct the training dataset, after ignoring pixels with a missing value in any of the seven reflectance bands of that product. Using these two training datasets, we used supervised classification approaches (Pang-Ning et al., 2006) to train a model that maps the seven-dimensional attribute space (the set of seven MODIS spectral bands) to the space of binary (water/land) class labels. Supervised learning methods have seen wide applicability in remote sensing applications (Mountrakis et al., 2011; Pal and Mather, 2005), where the goal is to use automated algorithms for learning the representation of a target class (e.g. water or land) using input features (e.g. multispectral signals). A wide variety of supervised learning algorithms have been explored for distinguishing water from land, such as the use of maximum likelihood classifiers (Adam, 1997; Feyisa et al., 2014), decision trees (Li et al., 2013; Verpoorter et al., 2012), artificial neural networks (Bishop et al., 1999), and support vector machines (Huang et al., 2011). They have been shown to provide better performance than unsupervised approaches, especially in regions where ground-truth labels are available for training. An extensive survey on supervised and unsupervised methods for mapping surface water bodies is presented in (Karpatne et al., 2016). We used support vector machines (SVMs) as our preferred choice of supervised classification algorithm. SVMs are widely used for classification in the machine learning and data mining community, since they can control for the complexity of the learned classification model, in accordance with the structural risk minimization theory (Vapnik, 1999). When SVMs are used with non-linear kernel functions, they are able to learn non-linear functions that are highly expressive. We used the radial basis function (RBF) as the preferred choice of kernel function for learning a non-linear mapping of the binary class labels using the multispectral bands. The optimal hyper-parameters of SVM using RBF kernel were chosen using 5-fold cross-validation on the training sets. These classification models were then applied to every 500 m pixel location at every 8-day time-step, starting on Feb 18th, 2000, to produce REFL product and NBAR product based binary (water/land) classification maps, with missing label assigned to a pixel

if any of the seven spectral bands of a data product at a particular time were missing values.

2.2.2. Merging of MODIS classification maps

The classification maps obtained from the previous step suffer from noise and missing values due to issues with both MODIS products, as described in Section 2.1.1 (primarily missing data for NBAR product, and data quality issues for REFL product). REFL product also contains cloud contamination that can impact the classification performance. In addition, a number of water bodies and their surrounding land areas are seasonally covered by snow or ice in the winter months. Since, snow can cover both land and water, snow pixels can neither be considered as land pixels nor water pixels and hence cannot be used while training the classification model. Since, the classification model has no information about snow pixels, they can lie on either side of the classification boundary. Based on the training samples used for our model, snow pixels tend to lie on the water side of the decision boundary and hence get classified as water. Information about snow, ice, and cloud observations can be obtained from the QA flags described previously. Here we describe the methodology for merging the REFL product and NBAR product based classification maps with QA flags to produce a combined classification map with fewer missing values, correctly flagged snow, ice, and clouds pixels. Fig. 3 gives the overview of our methodology.

Fig. 4 illustrates the methodology with an example of a region containing lake Mead and its surrounding area on Dec 27th, 2010. Fig. 4a and Fig. 4e shows the false color composite (FCC) images created using bands 7, 5 and 4 for the NBAR product and the REFL product respectively. In this composite, water generally appears in black color. Relatively poor quality of REFL product is evident through stripes of noisy data in the FCC image (Fig. 4a). Light blue pixels generally represents snow in this false color composite (highlighted by a red circle in Fig. 4e). Any pixel with missing value in any of the bands is represented in white color (highlighted by a red square in Fig. 4e). REFL based classification map (W_r) and NBAR based classification map (W_n) are shown in Fig. 4b and f respectively, with Water (W), Land (L), and Missing (M) labels. The W and L labels were produced by the supervised classification approach presented in Section 2.2.1, while M labels were given if any of the seven reflectance bands had a missing value. We see that W_n (Fig. 4f) has more missing values than W_r (Fig. 4b).

Since clouds contribute to the poor data quality in the REFL product, we used a cloud mask based on the 16-bit QA state flags to remove cloud observations from the W_r classification map. Specifically, we used the 'internal cloud algorithm flag' (bit 10) and 'pixel is adjacent to cloud flag' (bit 13). Apart from these pre-determined cloud flags, we also designed an additional cloud filter to better detect small cirrus clouds (that can sometimes look similar to snow or ice in reflectance values) that were not identified by the MODIS cloud filters and were instead labeled as water in W_r. To identify such clouds, we considered a pixel to be present in the cloud mask if its 'small cirrus flag' (bit 9) is equal to 1, the
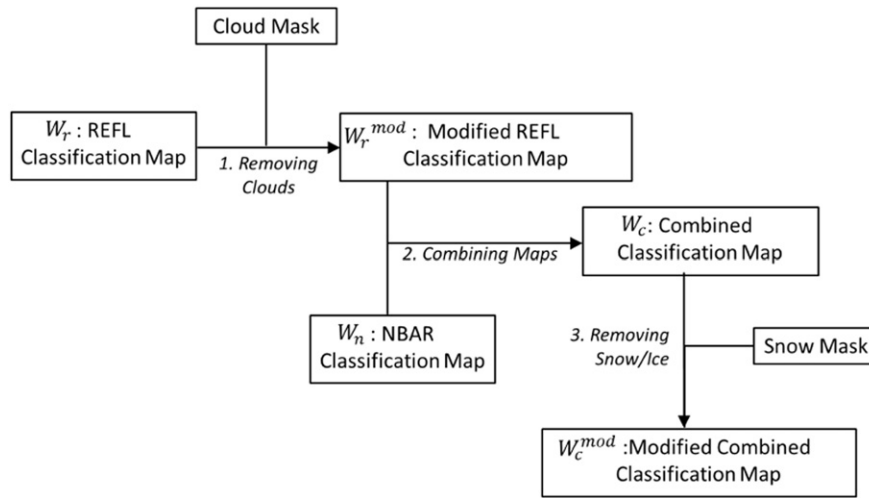
Fig. 3. Overview of our methodology for merging classification maps.

'MOD35 snow/ice flag' (bit 12) is equal to 1, the 'internal snow algorithm flag' (bit 15) is equal to 1, and the W_r label is equal to W. Fig. 4c shows this cloud mask for our example. Then using this cloud mask, we obtained the modified REFL classification map, W_r^{mod}, by assigning an M label to any pixel belonging to the cloud mask (Fig. 4d).

We then perform the merging step using the NBAR classification map (W_n), as the base map. As discussed previously, even though W_n contains more accurate class labels, it may also suffer from a large number of missing labels. To address this, we borrowed labels from the REFL classification map, W_r^{mod}, whenever we encountered a missing label in
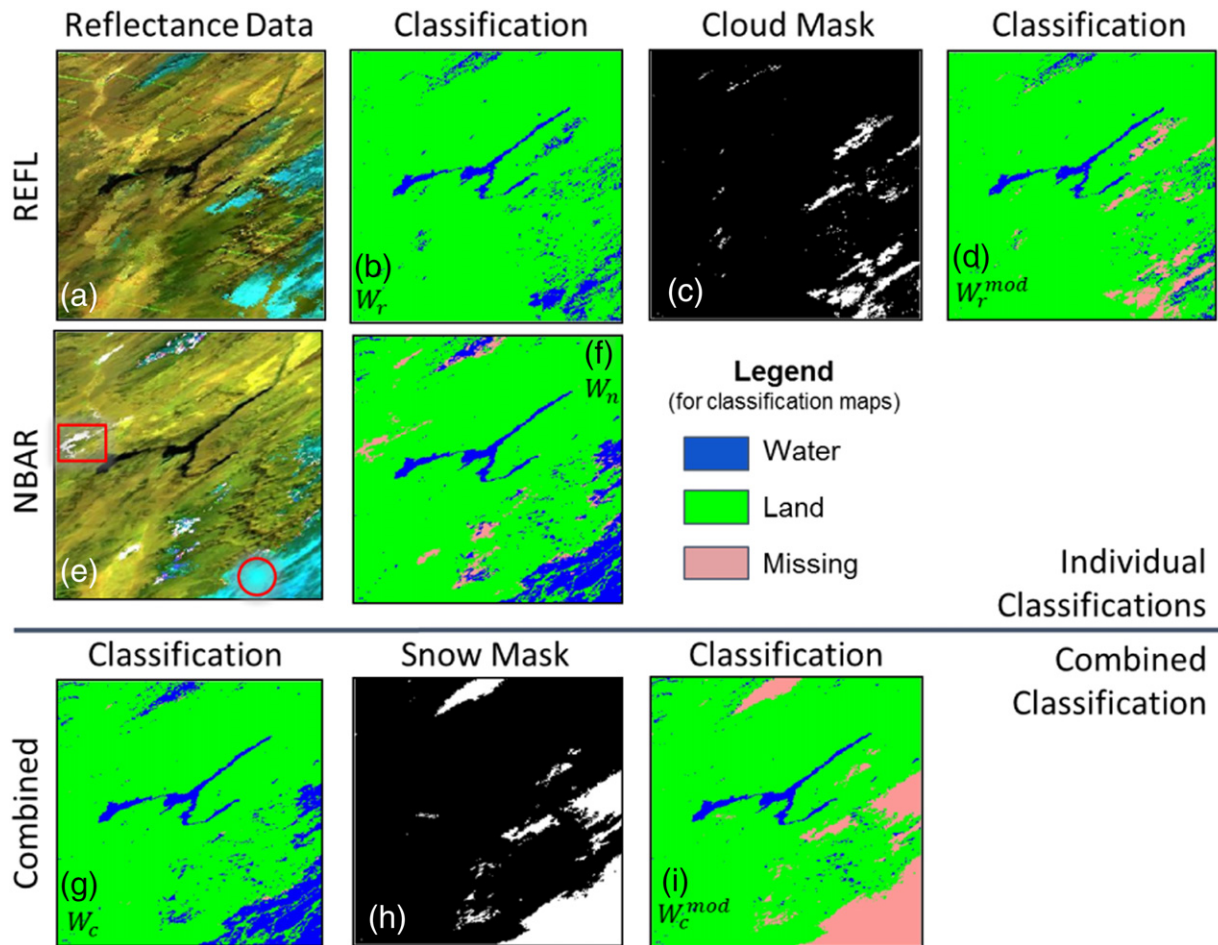

Fig. 4. An illustrative example demonstrating the methodology for merging classification maps for Lake Mead and its surrounding area. (a) REFL product based FCC image. (b) REFL classification map (W_n). (c) Cloud Mask. (d) Modified REFL classification map (W_r^{mod}). (e) NBAR product FCC image. Region enclosed in red circle shows an example region covered with snow. Region enclosed in red square shows an example region with missing information. (f) NBAR classification map (W_n). (g) Combined classification map (W_c). (h) Snow Mask. (i) Modified combined classification map (W_c^{mod}).

the NBAR classification map. The combined classification map, W_c, that combines the labels from W_n and W_r^{mod} can thus be obtained using the following equation:

$$W_c(l,t) = \begin{cases} W & \text{if } W_n(l,t) = W \text{ or } \left(W_n(l,t) = M \text{ and } W_r^{mod}(l,t) = W\right) \\ L & \text{if } W_n(l,t) = L \text{ or } \left(W_n(l,t) = M \text{ and } W_r^{mod}(l,t) = L\right) \\ M & \text{if } W_n(l,t) = M \text{ and } W_r^{mod}(l,t) = M \end{cases}$$

where, (l,t) denotes the pixel location (l), and time of observation (t), respectively. Fig. 4g shows the combined classification map (W_c) produced for the example data by merging W_r^{mod} and W_n.

Then, to account for snow and ice observations in W_c, we constructed a snow mask, S, which was equal to 1 if the snow BRDF albedo flag from the QA product is equal to 1, or any of the two snow flags in the REFL product are equal to 1, namely the 'MOD35 snow/ice flag' (bit 12) and 'internal snow algorithm flag' (bit 15). The snow mask for our example is shown in Fig. 4h. Any pixel that was labeled as snow or ice by the snow mask S was assigned a missing label, thus resulting in the modified combined classification map, W_c^{mod}, obtained as follows:

$$W_c^{mod}(l,t) = \begin{cases} W & \text{if } W_c(l,t) = W \text{ and } S(l,t) = 0 \\ L & \text{if } W_c(l,t) = L \text{ and } S(l,t) = 0 \\ M & \text{if } W_c(l,t) = M \text{ or } S(l,t) = 1 \end{cases}$$

Fig. 4i shows the modified combined classification map (W_c^{mod}) for our example, after applying the snow mask on W_c. Note that this classification map has fewer missing values than the classification map W_n, filters out poor quality pixels due to clouds, snow/ice, or missing values. For brevity, we will refer to the modified combined classification map as the initial classification map. Post-processing methodology described in the next section is applied on these initial classification maps to obtain final classification maps.

Note that we have used only the Terra product (MOD09A1) and not the Aqua product (MYD09A1). In our approach, we trained a separate classification model for both MOD09A1 and MCD43A4. Since, training information from MOD44W dataset is available for Feb 2000 from the SRTM mission, we could not use MYD09A1 product because it starts from 2002 and therefore could not train a classification model for this product.

2.2.3. Post-processing techniques

Water classification maps obtained from the methodology described above are affected by two primary issues – missing and incorrect labels. In the data preprocessing stage, some pixels were automatically excluded from the analysis because of poor data quality. These missing labels limit the estimation of water extent, and ancillary information is required to assign classification labels to these pixels. Moreover, even after the pre-processing stage, some poor quality pixels may not get excluded due to errors in measurement of quality flags. Classification

models applied to these (un-flagged) poor quality pixels might result in incorrect class labels.

Because poor data quality is often due to clouds, aerosols, and other related environmental factors, missing and incorrect labels are spatially and temporally autocorrelated. For example, Fig. 5a shows a false color composite of Lake Las Adjuntas in Mexico, with four major areas of missing data (shown in white), which appear as contiguous patches of pixels. Fig. 5b shows the initial classification map obtained from the classification methodology explained in Sections 2.2.1 and 2.2.2.

Traditional post-processing methods reported in the remote sensing literature (Liu et al., 2006; Paul et al., 2004; Liu et al., 2008; Griffiths et al., 2010; Chen et al., 2012; Rozenstein and Karnieli, 2011; Mithal et al., 2013) often assume that the errors in the labels are independent and hence occur as "salt and pepper" noise, but these techniques are not applicable here because of spatial and temporal autocorrelation in both missing and incorrect labels. Recently, Khandelwal et al. (2015) has proposed a new methodology to improve the accuracy of classification labels by using an implicit ordering through elevation among data instances to detect inconsistencies in classification labels. Water bodies are formed when water fills in depressions on the earth's surface, and locations within and surrounding a water body will have varying elevations. This bathymetry (elevation) information introduces an inherent ordering in the locations which determines how a water body grows or shrinks. If a location is filled with water, then all of the locations in the basin that have lower elevation than the given location should have filled with water beforehand. Thus, bathymetry information can be used to detect inconsistent classification labels that do not adhere to this physical constraint. Similarly, if we are given perfect classification labels (i.e., they perfectly agree to the physics constraint), then the growing and shrinking of a lake over time can be used to extract correct elevation ordering.

Fig. 6 demonstrates the utility of the elevation-based constraint. Fig. 6a shows a simplistic cross section of a lake's bathymetry. Clearly, the deepest location (G) will fill first and the shallowest location (A) last. Fig. 6b shows locations that have ordered according to their elevation with their corresponding input class labels. Fig. 6c provides the physical interpretation of the labels with obvious inconsistencies. Deeper locations (C and D) do not have water but a shallower location (B) has water, which is physically impossible. Hence, the bathymetry constraint was able to detect the inconsistencies in class labels.

But in reality, bathymetry information is available for only a few lakes in the world, and from in situ, rather than remote sensing sources, which limits their utility for analysis at global scale. So, we assumed that even though inherent ordering through elevation exists among locations of a given lake, it is not available or observed. In other words, the ordering is a latent or hidden information. In order to overcome this challenge, Khandelwal et al., 2015 developed an Expectation Maximization framework that iterates between estimating the ordering information from the classification labels and estimating correct classification labels from the estimated ordering. The algorithm iterates until elevation information cannot be further improved, which is then used to
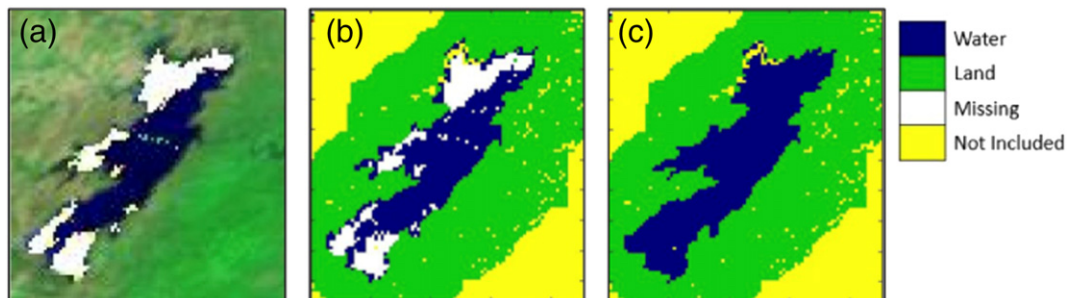

Fig. 5. An illustrative example showing the spatially correlated pattern of missing data. (a) January 25, 2011 MCD43A4 false color composite of Lake Las Adjuntas, Mexico (23.96°N, 98.73°W). (b) Corresponding initial classification map. (c) Classification map after post-processing.
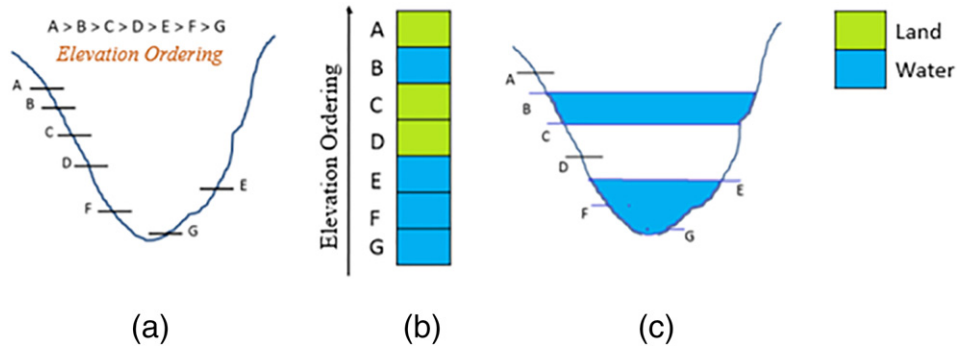
Fig. 6. Example demonstrating the utility of elevation constraint. (a) An illustrative example of lake bathymetry. (b) Locations and their input class labels. (c) Physical interpretation of the class labels.

correct inconsistent class labels. For the rest of the paper, this step will be referred to as elevation based label correction step.

Because the approach does not rely on using spatial and temporal autocorrelation in class labels, it was able to remove spatial and temporally auto-correlated errors effectively. For example, even though errors or missing labels due to clouds can show spatial and temporal autocorrelation, their appearance and disappearance is not related to the growing and shrinking of a water body. Hence the errors are expected to get fixed by the elevation based label correction method. Fig. 5c shows the corrected and imputed labels after the post-processing step. For further details, the reader is referred to Khandelwal et al. (2015).

2.2.4. Extraction of specific water bodies

As mentioned before, our goal is to monitor the surface area variations of individual water bodies. Hence, we needed to isolate the locations within and around the water body that contribute to its dynamics. For this purpose, we used spatial information from the Global Reservoir and Dam database (GRanD) (Lehner et al., 2011a, 2011b) and Global Lakes and Wetlands Database (GLWD) (Lehner and Döll, 2004) as the initial shape, which we call the initial region of interest (ROI). We created a buffer of 20 pixels at 500 m resolution around the initial ROI to include nearby locations that might contribute to the dynamics of the given water body. This assumes that the maximum water body

extent will not exceed this buffer region, which is sufficient for vast majority of water bodies. We call the initial ROI together with the automated buffer the max ROI. Fig. 7a, b and c show this process for Lake Kisale, Zaire. Fig. 7a shows the MCD43A4 based false color composite of Lake Kisale (marked by a red star) on May 25th, 2013. Water appears in black color in this false color composite. Fig. 7b shows initial ROI obtained from the GRanD dataset. Pixels with brown color represent the locations that are part of the initial ROI. Fig. 7c shows max ROI after the growing process.

However, there can also be scenarios where the water body of interest has other water bodies in close proximity and therefore the max ROI might also include other water bodies. If there are other water bodies in max ROI, then total number of water pixels in any given time step would not be a correct representation of the area of the specific water body of interest. Fig. 7d shows the above scenario for Lake Kisale as some surrounding water bodies have also become part of the buffer. Hence, we used a procedure to detect other water bodies and exclude them from the max ROI.

First, in order to detect water bodies in a region, we computed a fractional map that represents the fraction of time steps on which a location is classified as water. Fig. 7e shows the fractional map for Lake Zaire, scaled to lie between 0 and 100, with locations outside of the max ROI or that have zero fractional water in yellow. Generally, a water body
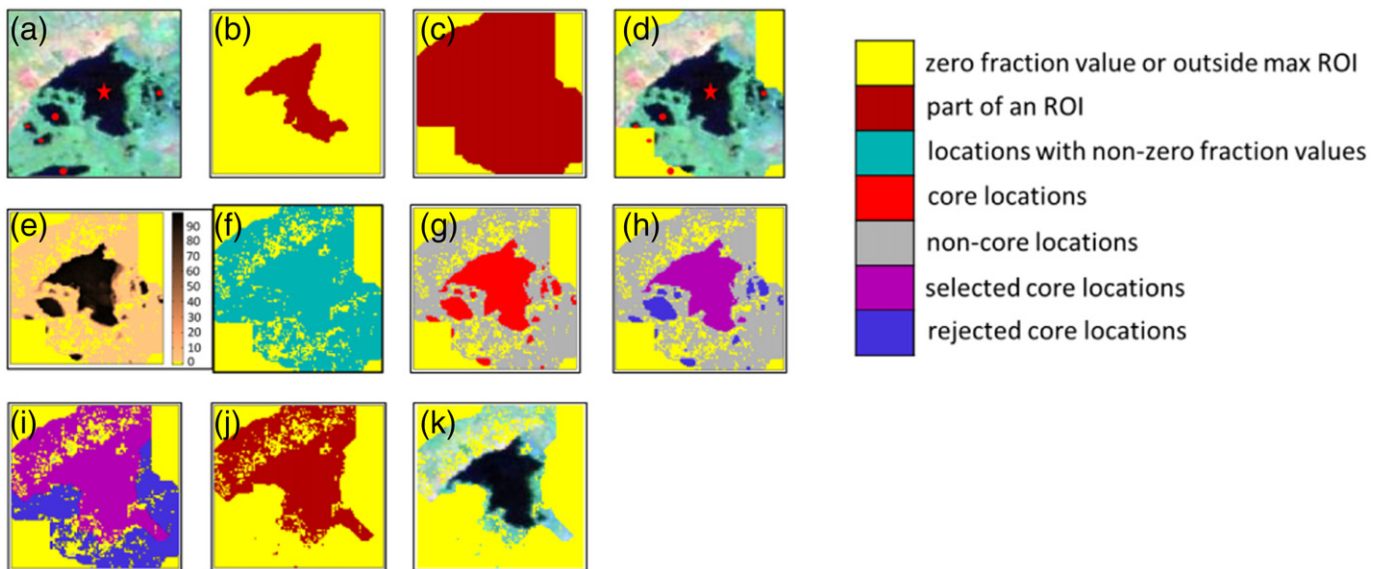

Fig. 7. Example of automated extraction of extent of Lake Kisale, Zaire (8.27°S, 26.50°E). (a) False color composite from MCD43A4 product with Lake Kisale marked by red star and surrounding water bodies marked by red circles. (b) Initial ROI from the GRanD database. (c) Max ROI from 20 pixel spatial buffer. (d) False color composite corresponding to maximum ROI. (e) Fractional map (0–100%) of water presence across all time steps. (f) Zero fractional threshold. (g) 10% fractional threshold, along with non-core components. (h) Selected core locations, rejected core locations, and non-core locations. (i) Selected and rejected components after region growing. (j) Final buffer, or pruned ROI. (k) False color composite within the pruned ROI.

grows and shrinks around the edges, so the inner regions of water bodies will have high fractional values whereas boundary locations will have lower fractional values. Moreover, locations that are always land will have a fractional value of zero and can be considered as locations that separate different water bodies. Ideally, if there are no classification errors, then different water bodies could be detected by simply thresholding the fractional map to obtain a binary map where locations with zero fractional value are land and locations with non-zero fractional values are locations of interest. The binary map could then be used to extract different connected components, or adjacent water pixels, and each connected component could be considered as a separate water body. However, the classification labels can have inaccuracies that might lead to incorrect estimation of water bodies. For example, Fig. 7f shows the binary map obtained using zero threshold (i.e. all locations in max ROI that have non-zero fractional values as locations of interest). As we can see, due to classification inaccuracies in the initial classification maps (obtained by merging NBAR and REFL product based classification maps as explained in Section 2.2.2), a vast majority of locations that are completely land in the image have non-zero fractional value, so extracting connected components from these locations would not generate separate water bodies. Instead, all the locations would be returned as a single water body. Alternative approaches include finding the time step with the maximum spatial extent and then selecting the largest connected component that overlaps with the initial ROI. However, this solution can also lead to poor selection of the water body if the time step with maximum extent has lots of classification errors which is common in our dataset (see Fig. S2). Hence, we needed a more robust way to threshold the fractional map. To address this, we first detected the inner regions (core locations) of water bodies by using a strict threshold on the fractional value, which was then grown to include the shallow regions around the core locations. For example, Fig. 7g shows the binary map with a 10% threshold, where the core locations (shown in red) corresponding to different water bodies were separated from the surrounding non-core locations (shown in grey). Connected components extracted from this map would correspond to the core region of multiple water bodies.

Next, if a core component intersects with the initial ROI, then it is considered as the core of interest. Fig. 7h shows locations that are part of cores of interest in purple and locations that are part of remaining cores (rejected cores) are shown in blue color. As mentioned above, inner regions or core regions of a water body are surrounded by shallow and more dynamic locations which are not part of any core. So, in order to assign the non-core locations to these two types of core locations, both types of core locations were grown simultaneously until all the non-core locations were assigned to one of the core types. Fig. 7i shows the locations of interest and rejected locations after the region growing process. The locations of interest obtained in this step were considered as the final ROI which we call as the pruned ROI as shown in Fig. 7j. Fig. 7k shows the false color composite image within the pruned ROI. As we can see, the pruned ROI contains only Lake Kisale and other water bodies have been removed from the ROI. The buffer extraction process not only removes unwanted water bodies from the buffer but can also handle mountain shadows if they do not directly fall on the water body of interest. In other words, spurious water regions created by mountain shadows (or snow) will be excluded if they do not overlap with the initial ROI (see Fig. S10).

In some situations, a single threshold on the fractional map may not lead to correct extraction of water bodies. As noted above, if the threshold is too low then spurious locations will become part of the core locations. On the other hand, if the threshold is too high then some of the water bodies might not have any locations among the core set. In order to make this process independent of the threshold, we estimated cores of interest and rejected cores for different thresholds (0 to 90, in 10 increments). Locations that became part of the rejected cores for any threshold were considered as belonging to the final set of rejected core locations. In other words, we took the union of all rejected cores

to obtain the final set of rejected core locations. To determine the final set of locations of interest, we selected locations that have fractional values greater than 10% but were not part of the final set of rejected locations. Finally, these locations were grown, as described above, to obtain the pruned ROI.

In some situations, a water body can be a group of multiple water components. The buffer extraction step allows multiple water components to be a part of a water body if the water components are represented in the initial ROI (see Fig. S3). The buffer extraction process excludes only those components that do not intersect with the initial ROI. However, if the water body has multiple water components, then the current methodology assumes that all water components grow and shrink while maintaining the same water level. This will be a correct assumption if the all the constituent water components are connected via a channel which may or may not visible at MODIS 500 m spatial resolution. However, there can be situations when water components are disconnected and hence can have different water levels. For such cases, it is better to do elevation based label correction separately for each water component of the water body and then aggregate their area estimates. In our current implementation, we assume that all the water components in the initial ROI are connected and hence elevation based label correction is applied on all the components simultaneously. In future, we will analyze each component separately and aggregate their area estimates.

2.3. Validation data

We have evaluated the performance of the proposed approach on the global set of 94 reservoirs comparing temporal variations in the surface extent maps with changes in relative water level height from satellite altimetry observations. Furthermore, for the specific 5 case studies, we have also used in situ observations and higher resolution Landsat based reference maps for isolated time steps to compare with the MODIS-estimated area. Next, we describe these validation datasets in detail.

2.3.1. Satellite altimetry

We obtained water levels from three different satellite altimetry databases: 1) U.S. Department of Agriculture (USDA) Global Reservoir and Lake Monitoring (GREALM) (http://www.pecad.fas.usda.gov/cropexplorer/global_reservoir) (Birkett, 1995; Birkett and Beckley, 2010), 2) the Laboratoire d'Etudes en Géophysique et Océanographies Spatiales (LEGOS/Hydroweb) in France (http://hydroweb.theia-land.fr/) (Crétaux et al., 2011), and 3) the Database for Hydrological Time Series of Inland Waters (DAHITI; http://dahiti.dgfi.tum.de/en/) (Schwatke et al., 2015).

The GREALM database mainly utilizes data from Topex/Poseidon (T/P), Jason-1, Jason-2, and GFO, as well as ENVISAT satellites. GREALM merges T/P, Jason-1, and Jason-2 (hereafter referred to as GREALM-10) time series of relative water level variation with respect to the 9-year mean T/P level at 10-day intervals. An ENVISAT time series of relative water level variation with respect to the mean level of a given ENVISAT reference cycle at 35-day intervals is also provided (hereafter referred to as GREALM-35). For each lake or reservoir included in GREALM, there are two time series of water level variations (specifically, the raw data and the smoothed data with a median type filter to eliminate outliers and reduce high frequency noise); the raw data were used in this study. Also, GREALM-10 altimetry sometimes includes overlap dates for a portion of the time series during the overlap period for two missions (such as T/P and Jason-1). Thus to generate a single time series for GREALM-10 altimetry, we found the date showing minimum difference between two altimeters during overlapping period and merged the two datasets based on that date. Other altimetry databases, such LEGOS and DAHITI, use additional altimeter missions than the GREALM database, including T/P, Jason-1, Jason-2, Jason-3, GFO, Envisat, ERS-1, ERS-2, Cryosat-2, and SARAL/AltiKa. They also differ by generating a

single water level time series by merging through several processing steps (e.g., multi-mission bias correction, outlier removal, Kalman filtering). Although there are promising applications to monitor water level changes in smaller water bodies using Cryosat-2 (Jiang et al., 2017), we do not use individual Cryosat-2 data except as already incorporated into the products described above.

Water levels from the three products are given relative to different datum/reference systems which precludes combining the absolute values from the different products. Therefore, we adopted the method of Birkett and Beckley (2010), where water levels from different altimeters are simply shifted vertically (adding or subtracting a shift constant) to correct for the different geoids or references and are then fit to the surface water areal extent. To compare with surface area, the water level estimates, available every 10 or 35 days (depending on the altimeter), were overlaid with the corresponding eight-day window of the MODIS composite data to analyze temporal variations.

2.3.2. In situ observations

We also obtained gauge observations of reservoir height for several of our case study sites (Table 1). Lake Mead water levels are recorded monthly by the U.S. Bureau of Reclamation (http://www.usbr.gov/lc/region/g4000/hourly/mead-elv.html). Monthly data for Lake Williston, Canada is archived by the Water Survey of Canada (https://wateroffice.ec.gc.ca/); the water levels observed at three separate gauge stations were averaged for comparisons with the surface area estimates Water levels for Lake Nasser were manually estimated for four months per year from 2000 to 2010 based on information reported by El-Shazli and Hoermann (2016), which was based on measurements from the Egyptian Ministry of Water Resources and Irrigation. We were unable to locate in situ data for Kremenshugskoye and Nova Ponte.

2.3.3. Landsat classification

In order to validate surface area estimates directly, we also created Landsat based reference classification maps. Landsat 5 and Landsat 7 data were downloaded from the USGS archive (http://earthexplorer.usgs.gov/). Both datasets provide 6 multispectral bands (red, green, blue, NIR, SWIR-1, and SWIR-2) at 30 m spatial resolution, and a thermal band at 120 m spatial resolution in Landsat 5 and 60 m in Landsat 7, which we resampled to 30 m resolution. For three reservoirs (Mead, Nova Ponte and Kremenshugskove; see Table 1 for details), we identified two cloud-free dates with high and low surface areas. First, each Landsat image was classified using an SVM classifier with training data collected from each image to ensure high classification accuracy. These classification maps were further improved through visual inspection. Traditionally, such reference maps are created completely by a human expert which is time consuming and cannot be done for multiple large water bodies at different points of time. Using a classification approach to obtain the initial reference map allows the human expert to focus only on the error-prone regions, hence making the reference map generation process relatively more scalable. An example of the Landsat classification is shown in Fig. S4 in the Supplementary material.

The Landsat-based reference maps were converted to polygons and reprojected to match the sinusoidal projection of MODIS. Then for each MODIS pixel, the intersection of the pixel with the reference polygon was used to calculate the fraction of the given MODIS pixel that is covered with water according to the Landsat based reference maps. This allows us to compute a MODIS scale reference map for each Landsat based reference map. For example, Fig. 8 shows the MDC43A4 based false color composite of Lake Mead on May 1st, 2001 and the MODIS scale reference map derived using Landsat based reference map on Apr 28th, 2001 (closest date to May 1st, 2001 for which Landsat data was available). White pixels represent locations that are not part of the pruned ROI.

Since the classification methodology returns binary labels, the classification maps cannot be compared directly with MODIS scale reference maps. To facilitate the comparison, we converted MODIS scale reference maps to binary maps using a threshold of 0.5 (i.e. locations in MODIS scale reference maps that have fraction values greater than or equal to 0.5 are labeled as water and locations with fraction values less than 0.5 are labeled as land). Comparing the classification maps with MODIS scale binary reference maps gives us quantitative information about the performance of the classification maps generated by the proposed approach. We used user's accuracy and producer's accuracy for quantitative evaluation of the results.

3. Results

To evaluate the performance at global scale, we have applied our approach on 94 reservoirs where altimetry data was available. Fig. S6 shows the geographical locations of these 94 reservoirs. These reservoirs belong to very diverse regions of the world across six different continents. We have used linear correlation between surface area estimates and altimetry data values as the evaluation metric. Furthermore, we have presented case studies for 5 reservoirs on four different continents where we discuss the performance in detail. We have also added Landsat based evaluation for 3 out of these 5 reservoirs.

3.1. Case studies

In the following section, we present and evaluate the performance of the proposed approach for five reservoirs across the globe (Table 1). The aim of these case studies is to demonstrate how the performance of the proposed framework gets impacted by various factors such as complicated geometries, varying sizes, geographic locations, and temporal dynamics. The altimetry satellite passing tracks and location of the gauge stations are shown in Fig. S5.

Each case study reservoir was also included by Gao et al. (2012) in their MODIS-based analysis, which allowed us to quantify

Table 1
Details of the five case studies presented in Section 3.1.

| Reservoir	Location	Area (km^2)	Dynamic region width	Linear correlation (R)					# Time-steps analyzed	% Missing in analyzed time-steps	% Missing in non-analyzed time-steps
				GREALM-10	GREALM-35	LEGOS	DAHITI	Gauge			
Mead (USA)	36.17°N, 114.40°W	610	6.7	–	0.93 (76)	0.95 (119)	0.96 (142)	0.98 (187)	710	0.09	72
Nasser (Egypt)	23.18°N, 32.75°E	5380	3.44	0.96 (503)	0.98 (74)	0.98 (312)	0.98 (338)	0.94 (47)	712	0.02	100
Kremenshugskoye (Ukraine)	49.25°N, 32.58°E	1850	2.23	0.44 (342)	0.42 (49)	0.48 (181)	0.45 (324)	–	463	0.84	67
Nova Ponte (Brazil)	19.10°S, 47.52°W	240	8.50	0.92 (149)	–	0.84 (30)	–	–	568	0.46	46
Williston (Canada)	56.00°N, 124.00°W	1620	1.8	0.53 (75)	–	0.61 (42)	0.52 (44)	0.68 (65)	206	0.47	63
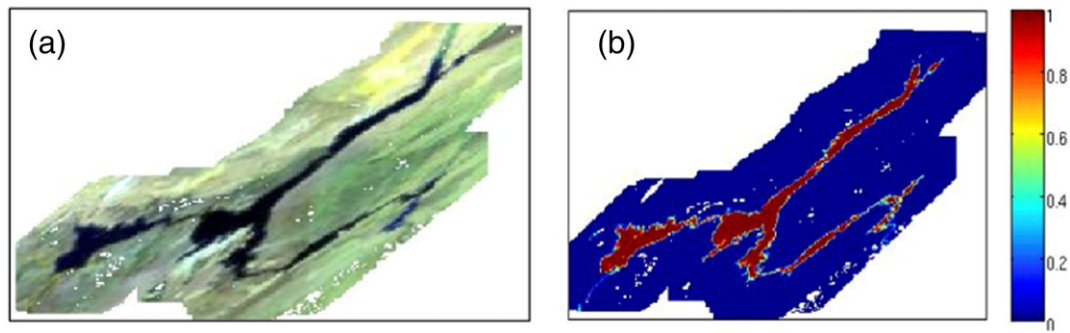
Fig. 8. a) False color composite of Lake Mead on May 1st, 2001. b) MODIS scale reference map on April 28th, 2001. Missing values are shown in white.

improvements in the correlations between the surface area estimates and radar altimetry-based relative height variations. Here, we have calculated a distance metric (henceforth referred to as Dynamic Region Width) which captures the level of dynamics in a given water body. To study the nature of dynamics in every water body, we can categorize locations in a water body into two types – static water locations (locations that remain water in all time steps) and dynamic locations (locations that were classified as both water and land at different times). If all the dynamic locations are very close to static locations, then we can say that the water body has not shown much variation and vice-versa. Dynamic Region

Width refers to the average distance (in 500 m pixels) from dynamic to static water pixels. For example, the Dynamic Region Width value of 1 would mean that on an average only 1 pixel around the static water pixels have shown dynamics i.e., changes are happening only at the boundaries of the static water pixels.

Table 1 provides the details of the five case studies. Areas given in column 2 are averages from the GRanD database (Lehner et al., 2011a, 2011b). Individual linear correlation coefficients are from water height datasets described in Sections 2.3.1 and 2.3.2. For each correlation, the numbers in parentheses represent the number of time steps for which both height and area data were available.
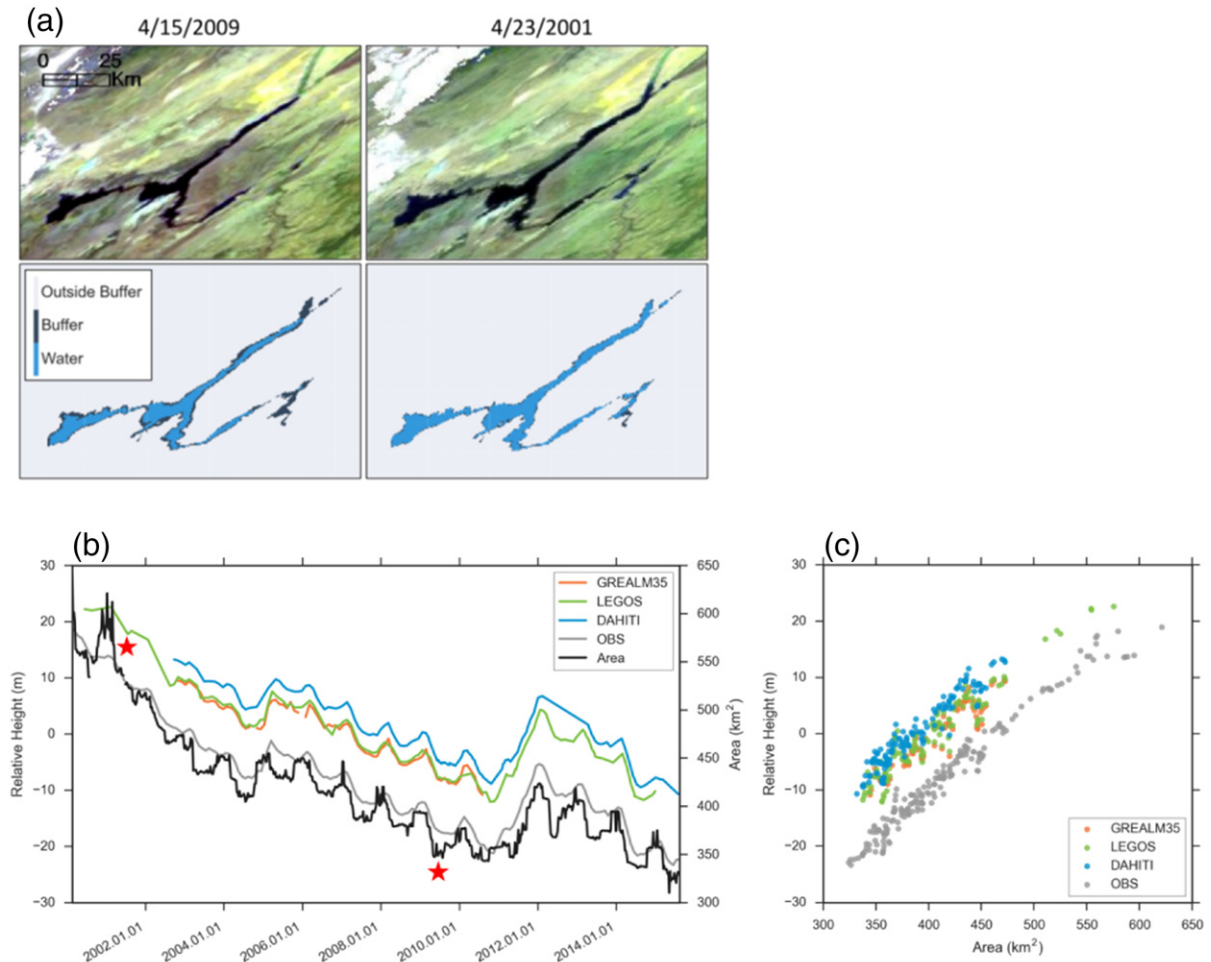

Fig. 9. (a) Comparison of 500 m MCD43A4 false color composite (top row) and classification results (bottom row) for Lake Mead, USA. The classification shows water pixels, non-water pixels within the maximum extent buffer, and pixels outside of the buffer. (b) Time series of surface area (km^2) from 2000 to 2015, with dates of classification results in (a) marked by stars. (c) Comparison of surface area (km^2) with relative elevation (m). OBS refers to in situ observations, other datasets refer to altimetry products described in Section 2.3.1.

3.1.1. Lake Mead, USA

Lake Mead, located between Nevada and Arizona in the USA, is approximately 610 km² in area (Lehner et al., 2011a, 2011b). As seen in Fig. 9a, this reservoir is challenging for classification algorithms because of multiple narrow reaches of the water body and its highly dynamic temporal variations, with Dynamic Region Width value of 6.7. Fig. 9b shows the surface area and relative height time series from 2000 to 2015, which indicates declining area over time, along with seasonal variations. Out of 713 potential time steps at the time of analysis (from 2/2000 to 8/2015), only three were removed because of large amounts of missing values (Table 1). While the Gao et al. (2012) classification performed well, with $R = 0.83$, our algorithm improved this to an average of $R = 0.94$ relative to the radar altimetry datasets (GREALM-35, LEGOS, and DAHITI), and $R = 0.98$ relative to monthly in situ observations (Fig. 9c).

3.1.2. Nasser, Egypt

Lake Nasser is a large reservoir (Fig. 10a) in southern Egypt and northern Sudan with an approximate area of 5380 km² (Lehner et al., 2011a, 2011b). It has strong seasonal dynamics (Dynamic Region Width of 3.5) that was captured well by the classification algorithm (Fig. 10b). Only one of the 713 potential time steps was removed due to missing values (Table 1). The classification improved on the $R = 0.79$ reported by Gao et al. (2012) to an average of 0.97 relative to

GREALM-10, GREALM-35, LEGOS, and DAHITI, and $R = 0.94$ relative to monthly in situ observations (Fig. 10c).

3.1.3. Kremenshugskoye, Ukraine

Kremenshugskoye is a large reservoir in central Ukraine with an approximate area of 1850 km² (Lehner et al., 2011a, 2011b). As seen in Fig. 11, although the reservoir has a more compact shape than the other reservoirs and relatively low dynamics (Dynamic Region Width of 2.23), it is challenging for the classification algorithm due to frequent missing values. Classification maps were produced for 463 of 713 potential time steps. Despite these limitations, our classification algorithm improved on $R = 0.23$ reported by Gao et al. (2012) to $R = 0.44$. However, due to the small variations in relative height (± 2 m) and large surface area and dynamic shallow regions in the northern part of the water body, strong relationships between area and height are difficult to achieve. Moreover, some time steps that had large fraction of locations with poor quality data were included in the analysis because the noise filter was not able to detect poor quality pixels in some of the time steps. The performance of the algorithm was limited in these time steps and lead to sudden area decreases/increases.

3.1.4. Nova Ponte, Brazil

The Nova Ponte reservoir is located in the Brazilian state of Minas Gerais, with an approximate area of 240 km² (Lehner et al., 2011a, 2011b). It has a highly complicated and dynamic shape (Fig. 12a),
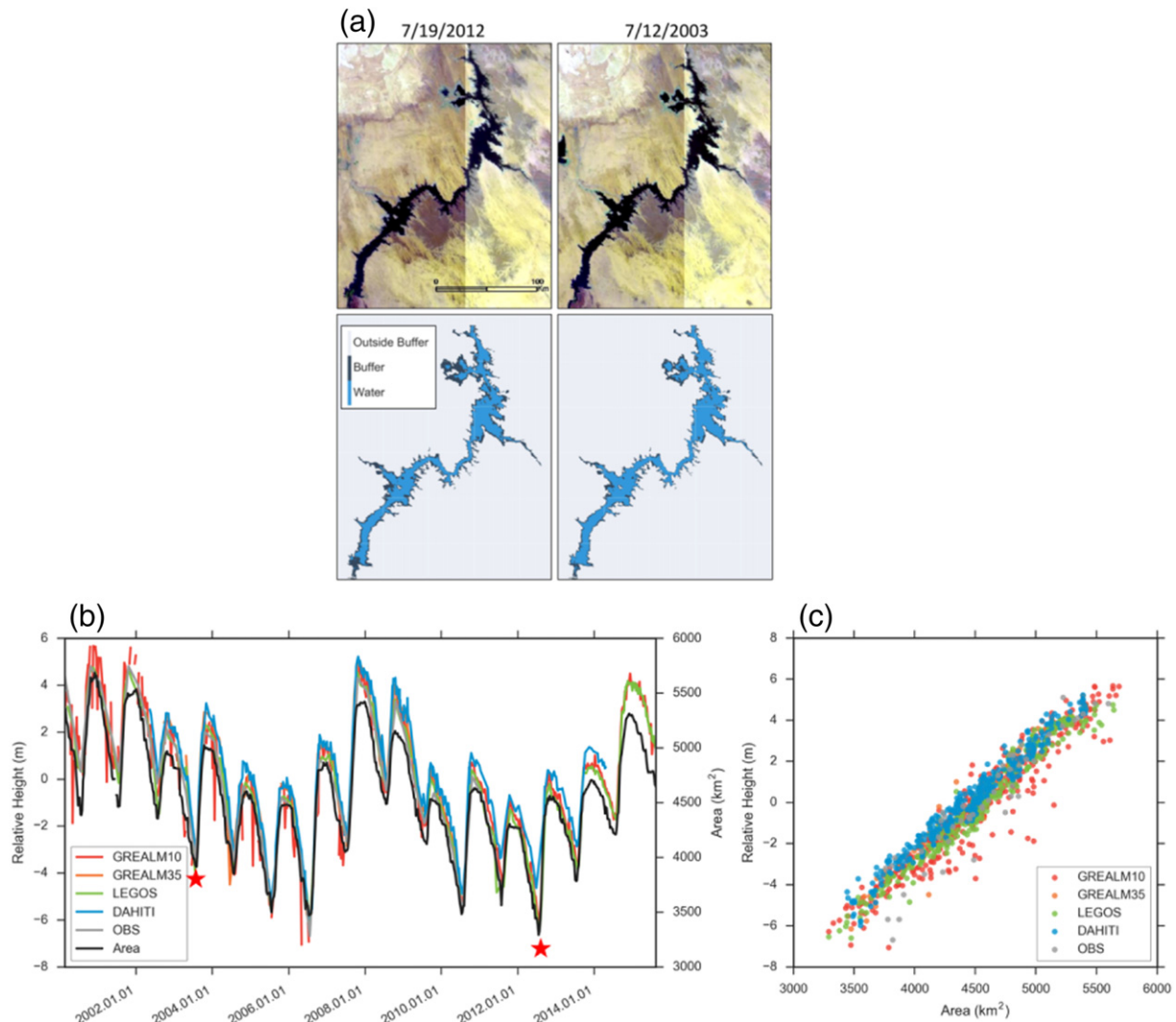

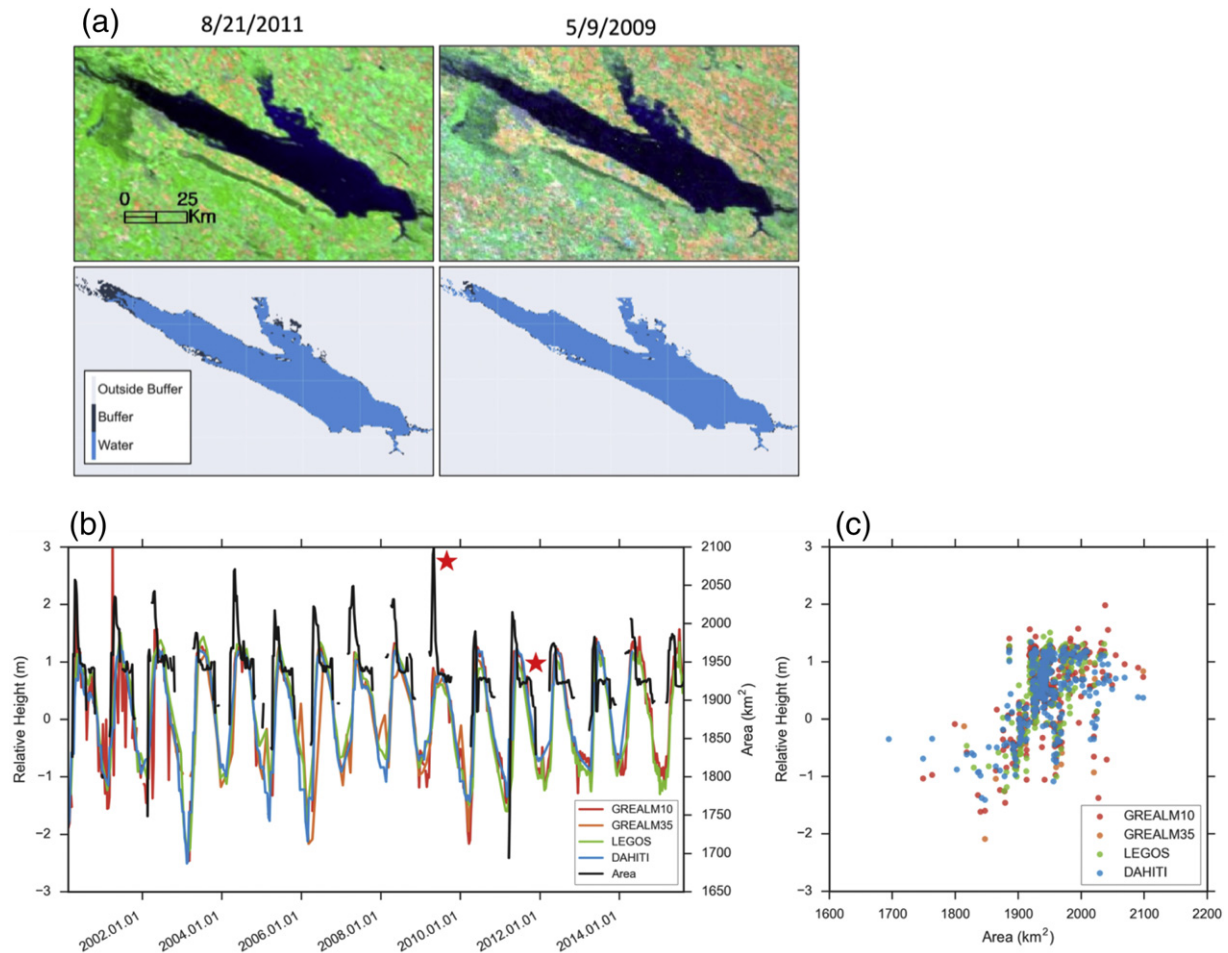Fig. 10. Following Fig. 9 but for Lake Nasser, Egypt.

Fig. 11. Following Fig. 9 but for Kremenshugskoye, Ukraine.

with the Dynamic Region Width of 8.5. Nova Ponte also suffers from a high fraction of missing values, especially due to cloud cover (Fig. 12b). Classification maps were produced for 568 of 713 time steps. However, the results here substantially improved on the $R = 0.08$ reported by Gao et al. (2012) to an average of $R = 0.88$ for the altimetry datasets (Fig. 12c).

3.1.5. Williston, Canada

Lake Williston is located in central British Columbia, with an approximate area of 1620 km² (Lehner et al., 2011a, 2011b). As seen in the classification images and time series of total area (Fig. 13a,b), there are many of dates with missing values (mostly due to ice cover), with maps produced for only 206 of 713 time steps. However, the correlations with relative height are reasonable, with an average of $R = 0.55$ with altimetry-based data and 0.68 with in situ observations (Fig. 13c). This improved on Gao et al. (2012), who reported $R = 0.11$.

3.2. Comparison with Landsat classification

Here, we compare classification maps generated by the proposed methodology with MODIS scale reference binary maps. Table 2 shows the number of 500 m water pixels detected by the proposed methodology in the region of interest for each reservoir and the number of 500 m pixels in the MODIS scale binary reference map that are obtained by thresholding the MODIS scale reference map using the 0.5 threshold. The remaining columns show the user's and producer's accuracy. Both user's and producer's accuracy values were high for Mead and Kremenshugskoye. For Nova Ponte, the results were less accurate for the high date (June 15th, 2011) and producer's accuracy was even low

for the low surface area date (July 13th, 2001). The reason behind the reduced performance on Nova Ponte was its near-linear shape, which results in a large number of pixels being mixed at MODIS scale. Mixed pixels have high uncertainty in classification because they do not have clear spectral signature of either water or land. Hence, these pixels have a high chance of being misclassified. On the date when the reservoir has low area, the impact of mixed pixels becomes more severe as the lake becomes more narrow due to drying.

3.3. Global altimetry based evaluation

The case studies presented above detail the performance of our classification algorithm in five diverse situations. The algorithm has been applied across the globe, and here we present performance on 94 reservoirs for which altimetry data was available. First, we evaluated the algorithm on a subset of 27 global reservoirs, all of which were previously analyzed by Gao et al. (2012). Constraining our results to the time period also analyzed by Gao et al. (2012), our algorithm equaled or improved the correlation between surface area and relative height from radar altimeters for a majority of reservoirs, see Fig. 14a. In Fig. 14a, X-axis represents the average of the correlation between area estimates by the proposed approach and height information from the four altimetry datasets. Y-axis represents the correlation value reported by Gao et al. (2012). For many of them, the correlation improved from a very low value (less than 0.4) to greater than 0.85. Fig. S7 shows the comparison between the proposed approach and Gao et al. (2012) when the maximum correlation among the 4 altimetry datasets was chosen for the proposed approach.
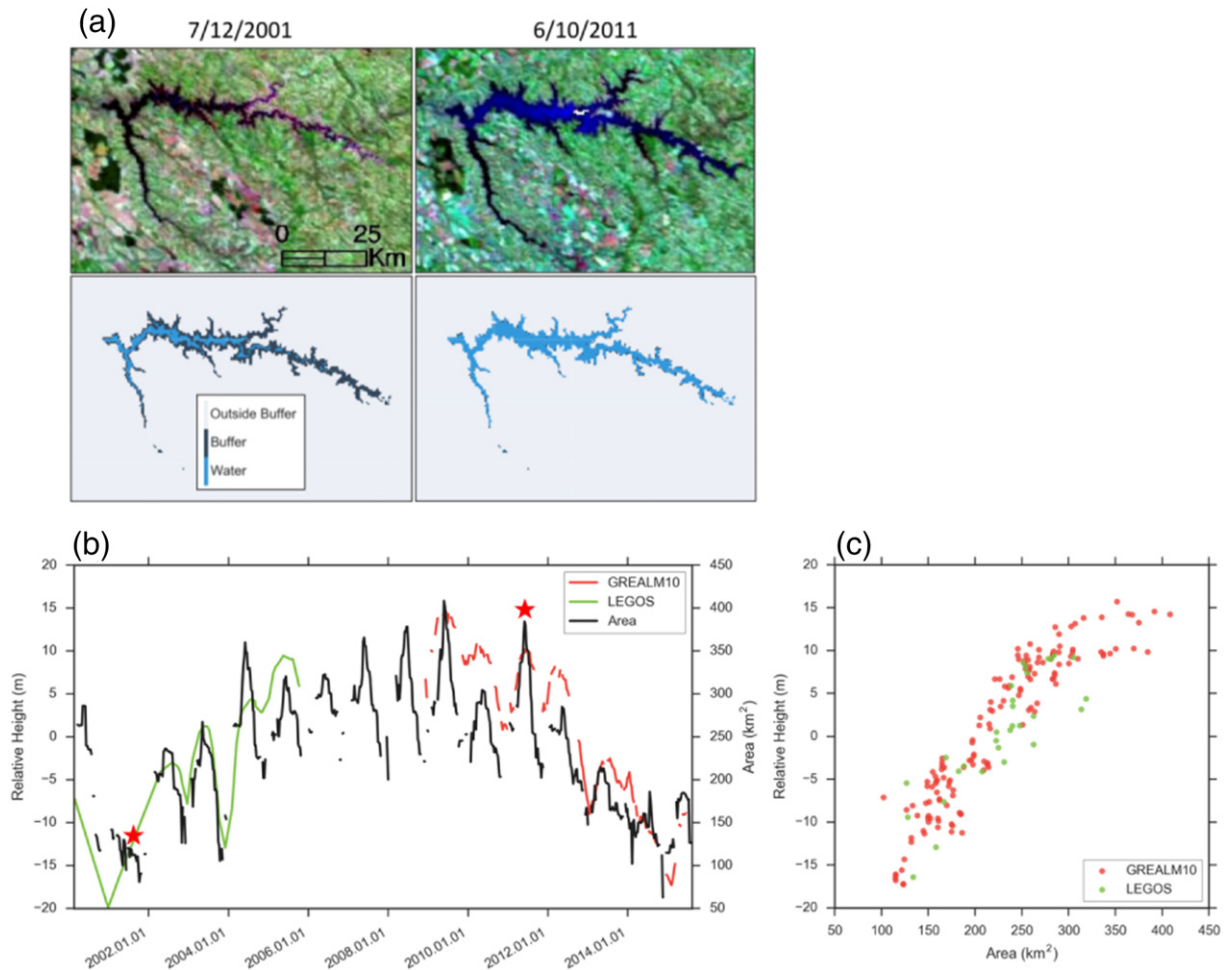
Fig. 12. Following Fig. 9 but for Nova Ponte, Brazil.

In Fig. 14b, we report the correlation values between area estimates of our approach and altimetry observations as a function of Dynamic Region Width values for the 94 reservoirs. Specifically, the Y-axis in Fig. 14b shows the maximum correlation among 4 altimetry datasets and X-axis shows the value of the Dynamic Region Width. As previously mentioned, a low Dynamic Region Width value corresponds to changes that are happening only in the boundaries of the static region of the lake. Due to the coarse spatial resolution of the MODIS data, the effect of mixed pixels is more prominent in water bodies with low Dynamic Region Width values, which can lead to poor correlation values with relative height variations. For the lakes with a high Dynamic Region Width value, the algorithm consistently performs much better.

3.4. Impact of data related challenges on the performance of proposed approach

Several challenges (such as noise, missing data, snow, clouds, mountain or cloud shadows, and turbidity) arise in mapping surface water extents and their dynamics from remote sensing data. In this section, we discuss the performance of the approach in the presence of these challenges. We have selected a wide range of examples from the set of 94 reservoirs to show both the robustness and limitations of the proposed approach to these challenges.

3.4.1. Impact of algae

In the 7-dimensional multispectral space, algae looks similar to vegetation when it covers the water surface. Hence, the presence of algae can lead to incorrect classification of water as land. This limitation in

the approach can be potentially handled by training a separate classification model to distinguish between vegetation and algae. However, in certain situations, the elevation based label correction method will be able to fix the classification errors, e.g., when the appearance and disappearance of algae is not consistent with the surface extent dynamics of the water body (see Fig. S9).

3.4.2. Impact of mountain and cloud shadows

Mountain and cloud shadows often get misclassified as water in the 7-dimensional feature space. Since our approach is currently focused on mapping dynamics of known surface water bodies, we avoid executing our approach where there is no known water body. However, some water bodies can be near mountains, and shadows from mountain can lead to overestimation of water in the buffer region around the water body. As explained before, if the shadows do not directly fall on the water body, they will be excluded from the analysis. In this way, buffer extraction process provides robustness to errors due to mountain shadows in certain situations (see Fig. S10).

Since, the errors due to cloud shadows are not related to the variation of area in a water body, the elevation based label correction method is able to detect classification errors due to cloud shadows as inconsistent and hence corrects them.

3.4.3. Impact of missing data

Data can be missing due to the presence of snow, clouds or sensor related issues. Accurate calculation of surface area is not possible when some of the labels are missing. Missing data can be highly correlated in both space and time and hence cannot be simply fixed by majority
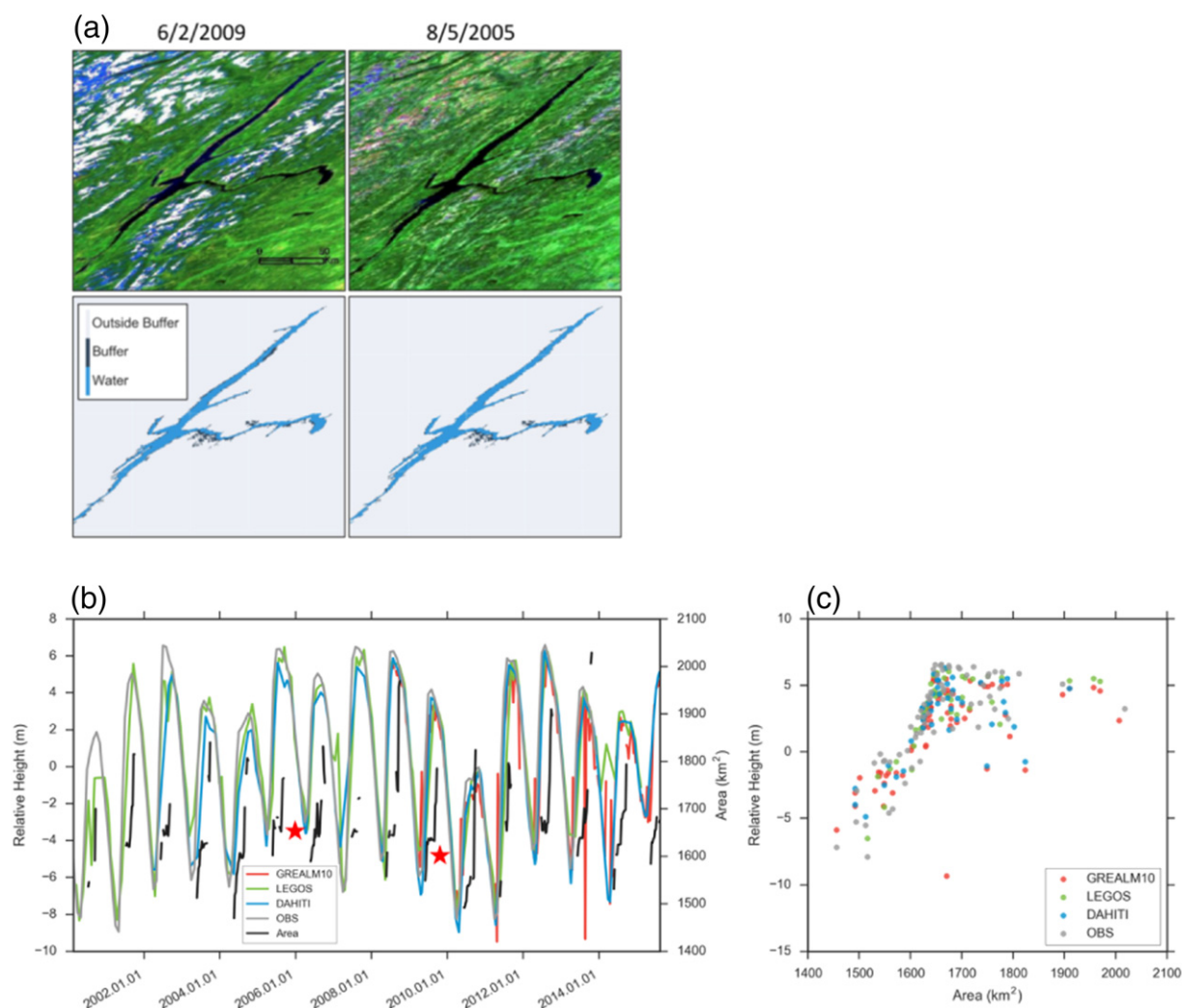
Fig. 13. Following Fig. 9 but for Lake Williston, Canada.

filtering based methods. Moreover, data can be missing on the boundaries which can be even more challenging. Since elevation based label correction method does not rely on spatial and temporal autocorrelation, it can impute labels very robustly by learning inherent elevation structure between locations (Figs. S11, S12 and S13).

3.4.4. Impact of incorrect labels by supervised classification

Due to the noise in data, spectral inseparability in the feature space, or lack of training samples, certain locations can get misclassified by the classification methodology. Again, these errors can be highly correlated in space and time, and cannot be removed by majority filtering based methods. However, if the occurrence of these erroneous labels does not agree with the physically constrained growing and shrinking of a water body, then these labels can be detected and subsequently corrected. Figs. S14–S20 shows the ability of the proposed approach to

recover the correct labels even in the presence of large amounts of imperfect labels.

There can also be situations when the approach is not able to correct the labels (Figs. S21 and S22). In Fig. S21, the water body was broken into pieces. The newly exposed land between the two pieces did not have any representative samples in the training data and hence was misclassified as water. The water body retained its size in the initial classification map and the classification results suggest that the lake did not shrink while the false color composites show that water body has shrunk. The elevation based label correction method was not able to correct the labels in this case because the spurious extent matches with the extent when the lake is really filled and hence these pixels are not detected as inconsistent. Similarly in Fig. S22, in reality the lake has shrunk but the initial classification is spuriously detecting the newly exposed land as water.

Table 2
Validation of MODIS-based classification maps with higher spatial resolution Landsat-based reference maps. For each of the three reservoirs, dates with high and low surface area extent were selected. Number of water pixels refers to 500 m MODIS pixels.

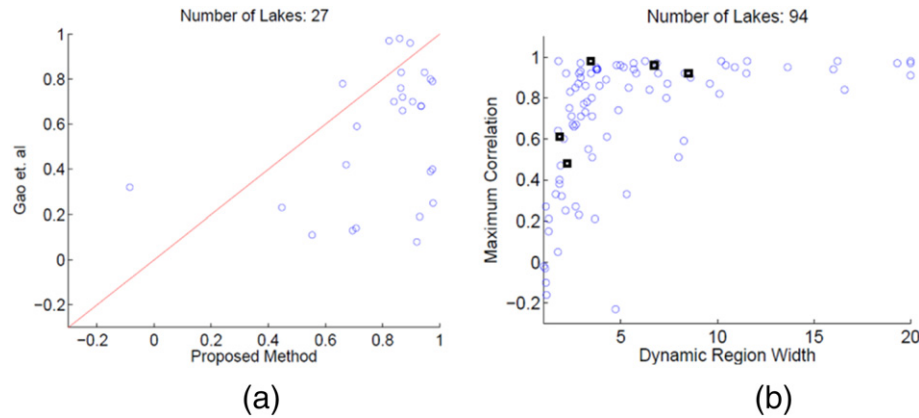| Reservoir	Date	MODIS water pixels	Landsat water pixels	User's accuracy	Producer's accuracy
Mead	4/18/2009 (low)	1698	1662	0.93	0.95
	4/28/2001 (high)	2523	2498	0.95	0.96
Kremenshugskoye	8/26/2011 (low)	8250	8369	0.99	0.97
	5/16/2009 (high)	8784	8653	0.98	0.97
Nova Ponte	7/13/2001 (low)	512	805	0.96	0.61
	6/15/2011 (high)	1665	1685	0.89	0.88

Fig. 14. (a) Comparison in the linear correlation coefficient (R) for 27 overlapping reservoirs between Gao et al. (2012) and results based on the proposed approach. b) Maximum R among the four possible altimetry datasets as a function of Dynamic Region Width, for a set of 94 global reservoirs. Black square boxes highlight the case studies and a map of locations is in Fig. S6.

4. Conclusions and future work

We have developed an approach that detects changes in the surface area extent of known reservoirs over time on a global scale. Our results improve upon previous efforts by: 1) providing water classification results at high temporal frequency (every eight days pending high quality data availability) and moderate 500 m spatial resolution across the globe, and 2) extracting surface area for specific water bodies of interest. We presented a summary of our global scale results on 94 reservoirs for which altimetry measurements were available, and show that our surface area estimates match well with relative height variations and show significant improvements over previous estimates on a subset of 27 reservoirs that were reported in (Gao et al., 2012). We also assessed algorithm performance for five reservoirs by comparing with radar altimetry and in situ observations of relative height variations over time and finer spatial resolution information from Landsat at isolated time steps. In general, correlations between the relative height variations and surface area extent were strong and offered significant improvements over previous estimates (Gao et al., 2012). However, surface area variations showed relatively poor correlation with relative height in areas where there were small variations in water depth over time (small value of Dynamic Region Width). Landsat-based classification results generally show a close correspondence with our MODIS-based results.

Based on these results, we conclude that:

1. Our MODIS-based classification for mapping global surface water extent variations performs considerably better than previous mapping efforts. Despite the somewhat coarse spatial resolution (500 m), the algorithm performs well when compared against relative height measurements and Landsat-based classification maps.
2. Our approach improves on previous efforts largely because of the effectiveness of the post-processing of classification results (that are performed using implicit elevation ordering information) for detecting and correcting erroneous classification labels.
3. Although all results presented in this paper are at MODIS scale, this post-processing approach can be used to correct classification labels generated at any resolution (e.g. Landsat or Sentinel) as long as we have satellite observation data for the water body on many dates where it has sufficient amount of variation in surface extent. This is required for effective estimation of elevation ordering that is used for post-processing of classification results. Hence, this post-processing approach will not be effective for water bodies that have very little variation (insufficient Dynamic Region Width) or are too small (e.g. less than 10 pixels) to have sufficient variation.

By automating both the classification algorithm and the ability to extract the area of a specific water body of interest, our classification estimates can be rapidly updated as new data becomes available. These two classification outputs broaden the use of this data to a wider range of the scientific community, including groups who are interested in classification of all water bodies at the MODIS scale or the results for specific lakes or reservoirs. The surface area estimates presented here are available to the public (http://z.umn.edu/monitoringwaterRSE).

Possible extensions for algorithm improvement include implementation of improved supervised classification models that can handle the rich heterogeneity in the characteristics of land and water bodies in multispectral data at a global scale (Karpatne and Kumar, 2015; Karpatne et al., 2015). This would result in more accurate REFL and NBAR classification maps, which are used for obtaining the combined classification map. Currently, the elevation constraint based post processing method gives equal importance to each location and similarly equal importance to each time step. But in reality, some locations are more difficult to classify and similarly some time steps might have more errors due to noise. To incorporate this property, we plan to develop an extension of the proposed scheme where locations and time steps will be weighted according to their classification quality. This will make the estimation of elevation more robust to noise and hence improved label correction. We also plan to update our results to the new collection 6 data and the latest available MODIS data beyond August 2015. Our goal is to continually update our results in an automated fashion operating on a near-real time basis, as and when new data is available via the MODIS repository. Further, in order to compute water storage variations over time, surface extent maps based on the methodology presented in this paper will be combined with water level variations from radar altimetry. We aim to combine the area results presented here with height information to monitor storage variation in a global dataset of several hundred lakes and reservoirs. Finally, we aim to increase the coverage of our analysis by using the proposed approach together with the possible extensions on Landsat data to monitor smaller water bodies as well.

reservoir/. We also thank the anonymous reviewers for their constructive feedback that greatly in improve the quality of the paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.rse.2017.05.039.

References

Adam, S., 1997. Glacier snow line mapping using ERS-1 SAR imagery. Remote Sense Environment 61, 46–54.

Alsdorf, D.E., Rodríguez, E., Lettenmaier, D.P., 2007. Measuring surface water from space. Rev. Geophys. 45, RG2002.

Bai, J., Chen, X., Li, J., Yang, L., Fang, H., 2010. Changes in the area of inland lakes in arid regions of central Asia during the past 30 years. Environ. Monit. Assess. 178, 247–256.

Birkett, C.M., 1995. The contribution of TOPEX/POSEIDON to the global monitoring of climatically sensitive lakes. J. Geophys. Res. Atmos. 100 (25179–179–25–204).

Birkett, C.M., Beckley, B., 2010. Investigating the performance of the Jason-2/OSTM radar altimeter over lakes and reservoirs. Mar. Geod. 33, 204–238.

Bishop, M.P., Shroder Jr., J.F., Hickman, B.L., 1999. SPOT panchromatic imagery and neural networks for information extraction in a complex mountain environment. Geocarto Int. 14, 19–28.

Brisco, B., Short, N., Sanden, J., Landry, R., 2009. A semi-automated tool for surface water mapping with RADARSAT-1. Can. J. Remote. Sens. 35, 336–344.

Carroll, M.L., Townshend, J.R., DiMiceli, C.M., Noojipady, P., Sohlberg, R.A., 2009. A new global raster water mask at 250 m resolution. Int. J. Digit. Earth 2, 291–308.

Chen, X., Chen, J., Shi, Y., Yamaguchi, Y., 2012. ISPRS J. Photogramm. Remote Sens. 71, 86–95.

Crétaux, J.F., Jelinski, W., Calmant, S., Kouraev, A., Vuglinski, V., Bergé-Nguyen, M., Gennero, M.C., Nino, F., Abarca Del Rio, R., Cazenave, A., Maisongrande, P., 2011. SOLS: A lake database to monitor in the Near Real Time water level and storage variations from remote sensing data. Adv. Space Res. 47, 1497–1507.

Deus, D., Gloaguen, R., 2013. Remote sensing analysis of lake dynamics in semi-arid regions: implication for water resource management. Lake Manyara, East African Rift, Northern Tanzania. Water 5, 698–727.

Döll, P., Douville, H., Güntner, A., Müller Schmied, H., Wada, Y., 2015. Modelling freshwater resources at the global scale: challenges and prospects. Surv. Geophys. 37, 195–221.

Donchyts, G., Baart, F., Winsemius, H., Gorelick, N., Kwadijk, J., van de Giesen, N., 2016. Earth's surface water change over the past 30 years. Nat. Clim. Chang. 6, 810–813.

Duan, Z., Bastiaanssen, W.G.M., 2013. Estimating water volume variations in lakes and reservoirs from four operational satellite altimetry databases and satellite imagery data. Remote Sens. Environ. 134, 403–416.

El-Shazli, A., Hoermann, G., 2016. Development of storage capacity and morphology of the Aswan High Dam Reservoir. Hydrol. Sci. J. 61, 2639–2648.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodríguez, 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, RG2004.

Feng, M., Sexton, J.O., Channan, S., Townshend, J.R., 2014. A global, high-resolution (30-m) inland water body dataset for 2000: first results of a topographic–spectral classification algorithm. Int. J. Digit. Earth 9, 113–133.

Feyisa, G.L., Meilby, H., Fensholt, R., Proud, S.R., 2014. Automated water extraction index: a new technique for surface water mapping using Landsat imagery. Remote Sens. Environ. 140, 23–35.

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.

Gao, H., Birkett, C., Lettenmaier, D.P., 2012. Global monitoring of large reservoir storage from satellite remote sensing. Water Resour. Res. 48, W09504.

Griffiths, P., Hostert, P., Gruebner, O., van der Linden, S., 2010. Remote sensing of environment. Remote Sens. Environ. 114, 426–439.

Haas, E.M., Bartholomé, E., Combal, B., 2009. Time series analysis of optical remote sensing data for the mapping of temporary surface water bodies in sub-Saharan western Africa. J. Hydrol. 370, 52–63.

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., Konzmann, M., Ludwig, F., Masaki, Y., Schewe, J., Stacke, T., Tessler, Z.D., Wada, Y., Wisser, D., 2014. Global water resources affected by human interventions and climate change. Proc. Natl. Acad. USA 111, 3251–3256.

Huang, L., Li, Z., Tian, B.S., Chen, Q., Liu, J.L., Zhang, R., 2011. Classification and snow line detection for glacial areas using the polarimetric SAR image. Remote Sens. Environ. 115, 1721–1732.

Jiang, H., Feng, M., Zhu, Y., Lu, N., Huang, J., Xiao, T., 2014. An automated method for extracting rivers and lakes from Landsat imagery. Remote Sens. 6, 5067–5089.

Jiang, L., Schneider, R., Andersen, O., Bauer-Gottwein, P., 2017. CryoSat-2 altimetry applications over rivers and lakes. Water 9, 211–220.

Karpatne, A., Kumar, V., 2015. Adaptive heterogeneous ensemble learning using the context of test instances. 2015 IEEE International Conference on Data Mining (ICDM), pp. 787–792.

Karpatne, A., Khandelwal, A., Kumar, V., 2015. Ensemble learning methods for binary classification with multi-modality within the classes. Proceedings of the 2015 SIAM International Conference on Data Mining, pp. 730–738.

Karpatne, A., Khandelwal, A., Chen, X., Mithal, V., Faghmous, J., Kumar, V., 2016. Global monitoring of inland water dynamics: state-of-the-art, challenges, and opportunities. Computational Sustainability. Springer International Publishing, pp. 121–147.

Khandelwal, A., Mithal, V., Kumar, V., 2015. Post classification label refinement using implicit ordering constraint among data instances. 2015 IEEE International Conference on Data Mining (ICDM), pp. 799–804.

Klein, I., Dietz, A.J., Gessner, U., Kuenzer, C., 2015a. Global WaterPack: Intra-annual assessment of spatio-temporal variability of inland water bodies. Remote Sensing Time Series. Springer International Publishing, pp. 99–117.

Klein, I., Dietz, A., Gessner, U., Dech, S., Gessner, U., 2015b. Results of the Global WaterPack: A novel product to assess inland water body dynamics on a daily basis. Remote Sensing Letters 6.1, 78–87.

Lehner, B., Döll, P., 2004. Development and validation of a global database of lakes, reservoirs and wetlands. J. Hydrol. 296, 1–22.

Lehner, B., Liermann, C.R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J.C., Rödel, R., Sindorf, N., Wisser, D., 2011a. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. Front. Ecol. Environ. 9, 494–502.

Lehner, B., Reidy Liermann, C., Revenga, C., Vorosmarty, C., Fekete, B., Crouzet, P., Doll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J.C., Rodel, R., Sindorf, N., Wisser, D., 2011b. Global Reservoir and Dam Database, Version 1 (GRanDv1): Dams, Revision 01.

Lettenmaier, D.P., Alsdorf, D., Dozier, J., Huffman, G.J., Pan, M., Wood, E.F., 2015. Inroads of remote sensing into hydrologic science during the WRR era. Water Resour. Res. 51.

Li, J., Sheng, Y., 2012. An automated scheme for glacial lake dynamics mapping using Landsat imagery and digital elevation models: a case study in the Himalayas. Int. J. Remote Sens. 33, 5194–5213.

Li, S., Sun, D., Yu, Y., Csiszar, I., Stefanidis, A., Goldberg, M.D., 2013. A new Short-Wave Infrared (SWIR) method for quantitative water fraction derivation and evaluation with EOS/MODIS and Landsat/TM data. IEEE Trans. Geosci. Remote Sens. 51, 1852–1862.

Liao, A., Chen, L., Chen, J., He, C., Cao, X., Chen, J., Peng, S., Sun, F., Gong, P., 2014. High-resolution remote sensing mapping of global land water. Sci. China Earth Sci. 57, 2305–2316.

Liu, D., Kelly, M., Gong, P., 2006. A spatial–temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery. Remote Sens. Environ. 101, 167–180.

Liu, D., Song, K., Townshend, J.R.G., Gong, P., 2008. Using local transition probability models in Markov random fields for forest change detection. Remote Sens. Environ. 112, 2222–2231.

Lu, S., Wu, B., Yan, N., Wang, H., 2011. Water body mapping method with HJ-1A/B satellite imagery. Int. J. Appl. Earth Obs. Geoinf. 13, 428–434.

Mithal, V., Khandelwal, A., Boriah, S., Steinhaeuser, K., Kumar, V., 2013. Change detection from temporal sequences of class labels: application to land cover change mapping. Proceedings of the 2013 SIAM International Conference on Data Mining, pp. 1–9.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66, 247–259.

NASA, 2005. Shuttle Radar Topography Mission Water Body Dataset (SWBD).

NASA, 2015a. Land Water Mask Derived from MODIS and SRTM L3 Global 250 m SIN Grid (MOD44W).

NASA, 2015b. MODIS Nadir BRDF-adjusted Reflectance 16-day L3 Global 500 m (MCD43A4).

NASA, 2015c. MODIS Surface Reflectance 8-day L3 Global 500 m (MOD09A1).

Pal, M., Mather, P.M., 2005. Support vector machines for classification in remote sensing. Int. J. Remote Sens. 26, 1007–1011.

Pang-Ning, T., Steinbach, M., Kumar, V., 2006. Introduction to Data Mining. Pearson Education.

Paul, F., Huggel, C., Kääb, A., 2004. Combining satellite multispectral image data and a digital elevation model for mapping debris-covered glaciers. Remote Sens. Environ. 89, 510–518.

Pekel, J.F., Vancutsem, C., Bastin, L., Clerici, M., Vanbogaert, E., Bartholomé, 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.

Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature.

Rozenstein, O., Karnieli, A., 2011. Comparison of methods for land-use classification incorporating remote sensing and GIS inputs. Appl. Geogr. 31, 533–544.

Schwatke, C., Dettmering, D., Bosch, W., Seitz, F., 2015. DAHITI – an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. Hydrol. Earth Syst. Sci. 19, 4345–4364.

Sharma, R., Tateishi, R., Hara, K., Nguyen, L., 2015. Developing Superfine Water Index (SWI) for global water cover mapping using MODIS data. Remote Sens. 7, 13807–13841.

Sheng, Y., Song, C., Wang, J., Lyons, E.A., Knox, B.R., Cox, J.S., Gao, F., 2016. Representative lake water extent mapping at continental scales using multi-temporal Landsat-8 imagery. Remote Sensing of Environment (ISSN: 0034-4257) 185:129–141. http://dx.doi.org/10.1016/j.rse.2015.12.041.

Sun, F., Sun, W., Chen, J., Gong, P., 2012. Comparison and improvement of methods for identifying waterbodies in remotely sensed imagery. Int. J. Remote Sens. 33, 6854–6875.

Vapnik, V.N., 1999. An overview of statistical learning theory. IEEE Transactions on Neural Networks, p. 10.

Verpoorter, C., Kutser, T., Tranvik, L., 2012. Automated mapping of water bodies using Landsat multispectral data. Limnol. Oceanogr. Methods 10, 1037–1050.

Verpoorter, C., Kutser, T., Seekell, D.A., Tranvik, L.J., 2014. A global inventory of lakes based on high-resolution satellite imagery. J. Geophys. Res. Atmos. 41, 6396–6402.

Zhang, S., Gao, H., Naz, B.S., 2014. Monitoring reservoir storage in South Asia from multisatellite remote sensing. Water Resour. Res. 50, 8927–8943.

Yamazaki, D., Trigg, M.A., Ikeshima, D., 15 December 2015. Development of a global ~ 90 m water body map using multi-temporal Landsat images. Remote Sensing of Environment (ISSN: 0034-4257) 171:337–351. http://dx.doi.org/10.1016/j.rse.2015.10.014.