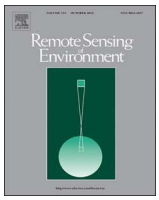
Remote Sensing of Environment 243 (2020) 111792

Contents lists available at ScienceDirect

Remote Sensing of Environment

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



Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series

Amy H. Pickens[a,*], Matthew C. Hansen[a], Matthew Hancher[b], Stephen V. Stehman[c], Alexandra Tyukavina[a], Peter Potapov[a], Byron Marroquin[a], Zainab Sherani[a]

[a] Department of Geographical Sciences, University of Maryland, College Park, MD 20740, USA
[b] Google, Mountain View, CA 94043, USA
[c] Department of Forest and Natural Resources Management, State University of New York, Syracuse, NY 13210, USA

ARTICLE INFO

Edited by Menghua Wang

Keywords:
Surface water
Landsat
Time-series
Area estimation
Change detection
Global

ABSTRACT

Global surface water extent is changing due to natural processes as well as anthropogenic drivers such as reservoir construction and conversion of wetlands to agriculture. However, the extent and change of global inland surface water are not well quantified. To address this, we classified land and water in all 3.4 million Landsat 5, 7, and 8 scenes from 1999 to 2018 and performed a time-series analysis to produce maps that characterize inter-annual and intra-annual open surface water dynamics. We also used a probability sample and reference time-series classification of land and water for 1999–2018 to provide unbiased estimators of area of stable and dynamic surface water extent and to assess the accuracy of the surface water maps. From the reference sample data, we estimate that permanent surface water covers 2.93 (standard error ± 0.09) million km^2, and during this time period an estimated 138,011 (± 28,163) km^2 underwent only gain of surface water, over double the estimated 53,154 (± 10,883) km^2 that underwent only loss of surface water. The estimated area of 950,719 (± 104,034) km^2 that experienced recurring change between land and water states is far greater than the area undergoing these unidirectional trends. From a probability sample of high resolution imagery, an estimated 10.9% (± 1.9%) of global inland surface water is within mixed pixels at Landsat resolution indicating that monitoring of surface water changes requires improved spatial detail. We provide the first unbiased area estimators of open surface water extent and its changes with associated uncertainties and illustrate the challenges of tracking changes in surface water area using medium spatial and temporal resolution data.

1. Introduction

Surface water presence is highly variable, with diverse trends around the world reflecting direct and indirect human as well as natural drivers. Meandering natural rivers and their floodplains support high biodiversity and provide nutrient-rich soil for agriculture, but are increasingly rare (Tockner and Stanford, 2002). Extensive areas of floodplains and natural wetlands have been engineered for food production, for example to rice and aquaculture (Davidson, 2014; Tessler et al., 2016; Zhao et al., 2008). Globally almost half of our river systems are moderately to severely fragmented by dams, levees, and other structures, affecting both ecosystems and economies (Grill et al., 2015). Climate change and diversion of rivers for irrigated agriculture have led to dramatic declines in the surface area of large saline lakes (Wurtsbaugh et al., 2017). Climatic changes are also intensifying rates of glacier melt, causing the lakes of the Tibetan plateau to expand and river discharge patterns to shift (Chevallier et al., 2011; Lutz et al., 2014).

Improved quantification of historical surface water trends will help us to better understand the impacts of such changes and to protect water resource-related ecosystem services. Historical archives of data from earth observation satellites are the only viable means to quantify these dynamics at a global scale and through time. Various efforts have advanced our understanding of historical surface water trends. Pekel et al. (2016) is the most comprehensive surface water product to date with 30 m monthly water/non-water layers, annual maps of seasonal and year-round water, and various multi-temporal maps initially for the period of March 1984–October 2015. In mid-2019, Pekel et al. (2016) released an update extending the mapped period through December 2018, which is available through the original sources. The water mask produced by Hansen et al. (2013) represents persistent water over 2000–2012, with all pixels having water in ≥50% of all growing season

* Corresponding author.
E-mail address: ahudson2@umd.edu (A.H. Pickens).

https://doi.org/10.1016/j.rse.2020.111792
Received 16 December 2019; Received in revised form 18 March 2020; Accepted 20 March 2020

Landsat 7 observations. Other global maps at 30-90 m (Chen et al., 2015; Feng et al., 2016; Yamazaki et al., 2015) have been produced but for isolated years and based on one or a few Landsat scenes per path-row as found in the Global Land Survey collection (Gutman et al., 2008; Tucker et al., 2004). Other products have evaluated intra-annual and inter-annual surface water dynamics but at subnational or national scale (Mueller et al., 2016; Tulbure et al., 2016; Zou et al., 2018). However, the existing global maps that identify change are only based on two time periods. Change in Pekel et al. (2016) was defined in two different ways. In one map, change was defined as the difference in open water occurrence percent between 1984–1999 and 2000–2018. In the second map, change was defined as the transition between permanent water, seasonal water, and land between a first year and the last year, 2018. The first year was selected between 1984 and 2000 on a per-pixel basis as the first year which had sufficient observations through the year to characterize the water presence. All intervening years were ignored for pixels in which water was identified in one or both of the start and end years (Pekel et al., 2016). Due to the large extent of fluctuations between water and land both seasonally and inter-annually (Papa et al., 2010; Pekel et al., 2016; Prigent et al., 2012; Yamazaki et al., 2015), it is necessary to evaluate the entire time-series to accurately assess surface water dynamics.

Previous quantifications of global water area have been based on "pixel counting" of the map (i.e., summing the area of pixels mapped as the target class). In contrast, current good practice guidelines recommend estimating area based on a probability sample of reference data (Eggleston et al., 2006; Olofsson et al., 2014; Penman et al., 2016). Because the reference class labels determined for the sample units have greater accuracy than the map classification, the area estimate based on the reference class labels is less subject to bias due to class labeling error. This greater accuracy of the reference class labels is achieved by using higher quality source data, such as by interpreting higher resolution imagery, or if using the same source data, by implementing an intensive interpretation effort to determine the reference class labels for the relatively small number of sample units (Olofsson et al., 2014). The variability of a sample-based area estimate is measured by the standard error, whereas, there is no measure of uncertainty associated with area derived from pixel counting. According to the Intergovernmental Panel on Climate Change (IPCC) of the United Nations (Eggleston et al., 2006), it is essential for international reporting to follow these guidelines in order to obtain unbiased area estimators of known uncertainty as quantified by standard errors. Good practice methods provide definitive information on area extent and change that can inform science applications and policy initiatives.

For this study we use 1999 as the start date because of low Landsat coverage prior to the launch of Landsat 7 and to the Long-Term Acquisition Plan (Arvidson et al., 2001). There are particular areas such as the United States and Australia that have sufficient observation density going back to 1985 to analyze the seasonal and inter-annual dynamics of surface water, but for other areas our first observations are not until much later, such as Siberia in 2000 and New Zealand in 1998 (Wulder et al., 2016). While early isolated observations provide valuable insights, the large gaps that exist in the early history of the Landsat archive further limit its utility in the consistent assessment of surface water dynamics prior to 1999. Problematic features include 5–8 year gaps with no acquisition in Central Africa and parts of Asia, and 12% of the continental area between 56°S-78°N has a median of zero acquisitions per year prior to 1999. From 1999 to 2012 both Landsat 5 and 7 were collecting data with revisit rate of up to every 8 days, though in reality this is much lower due to cloud cover and data transfer limitations of the satellites. Since 2013, both Landsat 7 and 8 have been collecting data, again with a revisit rate of up to every 8 days, though now the satellites are effectively achieving acquisitions every 8 days (Wulder et al., 2016).

This study advances our understanding of and capability for monitoring open inland surface water. We present a new map characterization of permanent open surface water and of various open surface water change dynamics from 1999 to 2018 based on the entire Landsat archive for this period, consisting of 3.4 million scenes. Using a probability sample of reference data, we assess the accuracy of our global surface water maps for the 1999–2018 interval as well as the accuracy of the maps produced by Pekel et al. (2016). We also use these reference data from the probability sample to provide unbiased estimators of area of global open surface water extent and change. We define open surface water as water on the ground surface that (1) is visible from above and not obscured by objects on or above the water surface, for example, forest, floating aquatic vegetation, or bridges and that (2) covers ≥50% of a 30 m pixel.

Here we analyze a more temporally dense time series than previous studies for both change maps and reference data, employing the full Landsat archive in improving the characterization of the dynamics of global open surface water extent. We employ the first probability-based sample that targets changes in global surface water extent, providing area estimates of dynamics derived from unbiased estimators. The results are less susceptible to the bias encountered when reporting change from pixel counts. The associated standard errors from the reference samples are an improvement over map product areas that lack measures of statistical uncertainty. Our accuracy assessment is another advance, covering the entire time-series of our map products as well as the entire global land surface, making it spatially and temporally comprehensive. Further, the accuracy assessment corroborates the quality of the maps of Pekel et al. (2016) based on a more complete reference sample data set than was used by Pekel et al. (2016) in their evaluation of their map. Lastly, we perform a sub-pixel analysis of 30 m Landsat maps based on 5 m spatial resolution reference sample data from RapidEye imagery, quantifying the global presence of mixed pixels containing land and water at the Landsat observational scale.

2. Data and methods

2.1. Per scene classification

A time-series of open water presence was created through the automated implementation of per Landsat sensor ensembles of classification trees. For each sensor, classification tree models of observation quality and land or water state (Potapov et al., 2015) were developed in the Global Land Analysis and Discovery (GLAD) laboratory. These models were then implemented in Google Earth Engine (Gorelick et al., 2017) and applied to the entire 1999–2018 Landsat 5, 7, and 8 archive, classifying each scene into land, water, cloud, shadow, haze, and snow and ice. In each scene we aim to map as water all pixels with ≥50% water cover according to the definition of open surface water given above.

Due to the diversity of reflectance patterns for open surface water targets, hierarchical, bagged classification trees (Breiman, 1996; Breiman et al., 1984) were used to discriminate clear observations from those contaminated by cloud, shadow, and ice and discriminate water from land. All images were first converted to top of atmosphere (TOA) reflectance (Chander et al., 2009). Water is sufficiently separable from land to employ TOA units in discrimination without conversion to surface reflectance or application of other normalization methods (Pekel et al., 2016; Tulbure et al., 2016; Yamazaki et al., 2015). The classification models utilize all the Landsat bands (excluding the panchromatic band), normalized difference ratios of each pair of bands, and 3 × 3 pixel spatial averages of all bands and ratios, as well as, utilize topographic inputs of elevation and derived slope, aspect, and hillshade data. There are 21 normalized difference ratios for Landsat 5 and 7 and 36 for Landsat 8 due to three additional spectral bands. These ratios include commonly applied water indices such as the Normalized Difference Water Index (NDWI, (Green-NIR)/(Green + NIR)) (McFeeters, 1996) and the Modified Normalized Difference Water Index (MNDWI, (Green-SWIR1)/(Green + SWIR1)) (Xu, 2006). Elevation was

taken from Shuttle Radar Topography Mission (SRTM) (SRTM, 2017) elevation data for areas south of 60°N, and Global Multi-resolution Terrain Elevation Data 2010 (USGS, n.d) for areas north of 60°N.

Five sets of seven bagged classification trees were built hierarchically for each sensor from a training set of fully classified scenes. These scenes were individually classified through manual training and using the same classification tree framework as the globally applied model. The sets of classified scenes were used to train the global models, for example 165 Landsat 5 scenes were mapped with the six categories of water, land, cloud, shadow, and snow/ice and used in training a global Landsat 5 water model. The global models were iterated by testing images not included in the training, visually identifying errors and adding problematic scenes to the training set and adjusting sampling rates until results were deemed satisfactory at the global scale. Scenes were selected from the year of launch through 2011, 2013, and 2015 for Landsat 5, 7, and 8, respectively, and within 125°W to 162°E and 60°N to 89°S. The final number of scenes in the training set was 165, 164, and 120 per Landsat 5, 7, and 8, respectively, and these scenes were sampled with a rate of 0.5–1.5% depending on the class and the set, resulting in over a billion pixels used as training for the global models.

The first set of seven bagged classification trees separates cloud and snow and ice from haze, land, shadow and water; the second set separates water from haze, land, and shadow; the third set separates shadow from haze and land; and the fourth set separates haze from land; and the final set separates cloud from snow and ice. Thus, clear water observations are identified after the first two sets and clear land observations after four sets. The set of seven trees per sensor that discriminated water from clear land, shadow, and haze relied strongly on the normalized difference ratio of the near infrared and green bands (NDWI), accounting for 89%, 85%, and 81% of deviance decrease for the Landsat 5, 7 and 8 models respectively. For this set of trees separating water, all of the 3 × 3 pixel spatial average metrics accounted for 4.9%, 6.7%, and 7.9% of the total deviance decrease, but for the first set of trees identifying cloud, snow, and ice the 3 × 3 metrics contributed 86.5%, 82.9%, and 74.4%, which is likely due to the diffuse nature of clouds. However, all of the Landsat bands and ratios and the topographic inputs contributed to discriminating water, land, and bad data. The models are thus complex, with 50–1000 nodes per tree. However, overfitting is avoided by using a large training sample and aggregating each set of trees by selecting the median output probability. This enables the identification of open surface water across many different states, from sediment-laden to clear to eutrophic and from shallow to deep; water with the surface obscured from above by vegetation or other obstructions such as bridges is excluded.

These models were then applied to all 3.4 million scenes, totaling 2.4 petabytes of data with the computing power of Google Earth Engine. In order to composite the results into global maps all classified data were resampled from the source UTM zone to a geographic projection (Lat/Long) with a spatial resolution of 0.00025°, equivalent to 28 m at the equator.

2.2. Time series analysis

2.2.1. Creation of annual percent water time-series

The land and water observations of a given pixel were summed per month and aggregated into water presence frequency at various timescales, measured by the percent of clear observations flagged as water (water / (water + land)), hereafter referred to as water percent. To create a more representative and stable measure of water percent, the individual observations were filtered and weighted, a practice also implemented by Pekel et al. (2016). First, clear observations (water + land) over the full study period were examined and if < 12.5% of observations were in an opposite state of water or land (not covered with clouds, shadows, haze, snow or ice), and the total number of these observations was ≤3, they were removed as outliers. This was done because cloud shadow over dense tree cover or other dark surfaces can erroneously be flagged as water, and image artifacts or undetected haze over water can sometimes be erroneously flagged as land. However the probability of these errors repeatedly happening over the same pixel is low, and we found that these thresholds removed considerable noise while not erasing real change.

We created a seasonally-normalized annual water percent to account for intra-annual variation in clear observation frequency (e.g. fewer observations in the far north in winter due to low sun elevations or in tropical monsoon environments due to high cloud cover). The water and land observations were summed per meteorological seasons (December–February, March–May, June–August, September–November) of each year, and used to calculate the percent of water observations out of all clear observations per season. Months with < 5 observations over the 20 years were excluded to remove a potentially irregular impact on the annual time-series due to different portions of the year being observed. The average of the four seasons with data was used to calculate percent of water per year. Given that the start day of hydrological years varies around the world, we selected a start day that corresponded with the meteorological seasons and that mostly closely aligned with a traditional calendar year. A year was thus defined as December 1 of the previous year through November 30 of the given year (e.g. 2003 was defined as December 2002–November 2003). Seasonal weighting was done to account for varying number of clear observations during different seasons of the year due to seasonal prevalence of weather events which often both obscure the surface and cause more surface water, seasonal snow cover, and varying acquisition rates related to sun angle for regions at high latitudes.

2.2.2. Dynamic type classification

An inter-annual water dynamics model was developed to characterize and visualize the changes occurring over the study period (Fig. 1). To reduce short-term annual anomalies and inter-annual cloud-free observation variability, the annual open water percent time-series was smoothed using a 3-year mean moving window. Next, the range and mean of the annual percent water time-series data were calculated per pixel. The range is representative of the difference between the maximum rate and the minimum rate of open water presence. Pixels with a range ≤33% and a mean ≤10% or ≥90% were labeled permanent land and water, respectively, providing a stable target with low sensitivity to possible omission or commission effects due to image artifacts or atmospheric conditions. Short-duration, anomalous flood events are ignored in this definition of permanent land. Change pixels were identified as pixels with a range ≥50%, and all other pixels labeled as stable seasonal, characterized by water presence having little or no inter-annual variation and consistent intra-annual variation. Further characterization was applied to pixels labeled as change to identify typologies defined by an analysis of all local maxima and minima in the time-series. These local extrema were used to segment the 20-year time-series; segments with an amplitude < 30% of the overall time-series amplitude were removed. Remaining dominant change segments were used to characterize the following main change types: gain, loss, dry period (water-land-water), wet period (land-water-land), and high frequency (3 or more) land-water transitions. All of these change types were mapped and validated along with the stable seasonal, permanent land, and permanent water classes. If a pixel had < 10 years with observations or < 15 total observations and did not meet the criteria for permanent water or land, it was marked as sparse-data.

2.2.3. Dynamic class mapping

To visualize all of the classes in a continuum we reduced the time-series to three time-sequential values of water percent mapped in an R-G-B color space (Fig. 2). Specifically, the aforementioned classification rules were refined to characterize the timing of monotonic changes and the intensity of all change types. For pixels with monotonic loss or gain of water, the red band value was taken from the maximum or minimum

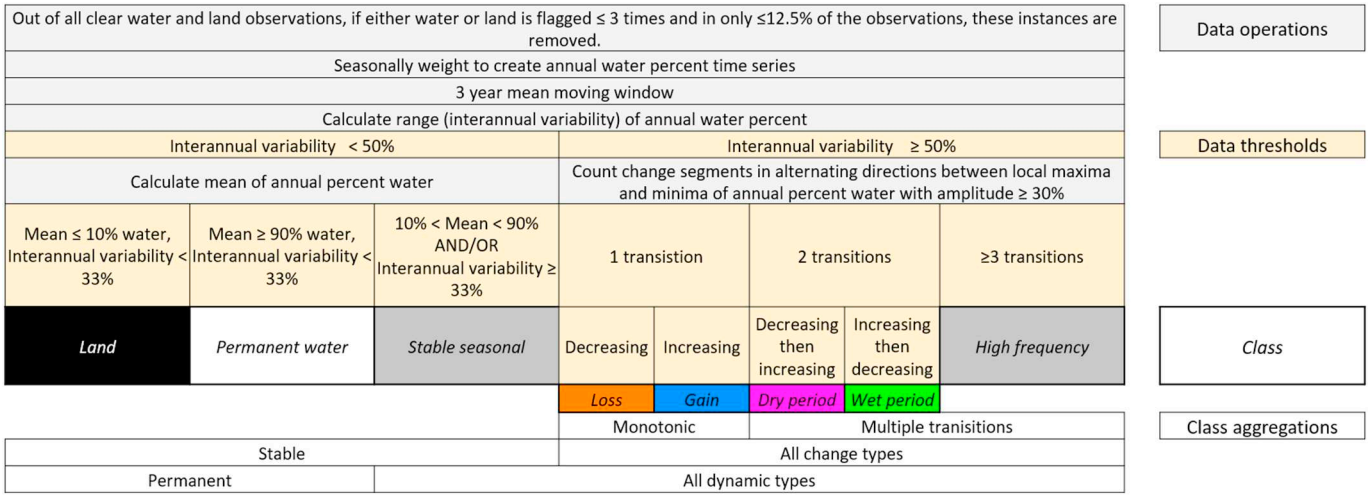


Fig. 1. Workflow of time-series analysis starting with the monthly water and land counts and resulting in dynamic class labels.

water percent, respectively, at the beginning of the period. The blue band value was taken from the minimum or maximum water percent, respectively, at the end of the period. The green band value was the mean of the entire period, which enabled a graphical representation of whether the change was comparatively early or late in the study period. For pixels with two changes, there was either a change from water to land to water (a dry period), or a change from land to water to land (a wet period). For pixels with a dry period, there was a local maximum, followed by the minimum, and then by a local maximum which were assigned time-sequentially to the R-G-B space. The minimum was assigned to the green band and the two maxima were assigned to the red and blue bands with the lesser maximum averaged with the beginning or end value of the time-series depending on whether it was before or after the minimum. For example, a pond that had year-round water that dried up for a few years and then filled up but only for half the year would be represented in R-G-B space by its initial maximum value of 100% (r), its minimum value of 0% (g), and its final seasonal value of 50% (b). Likewise, for pixels with a wet period, there was a local

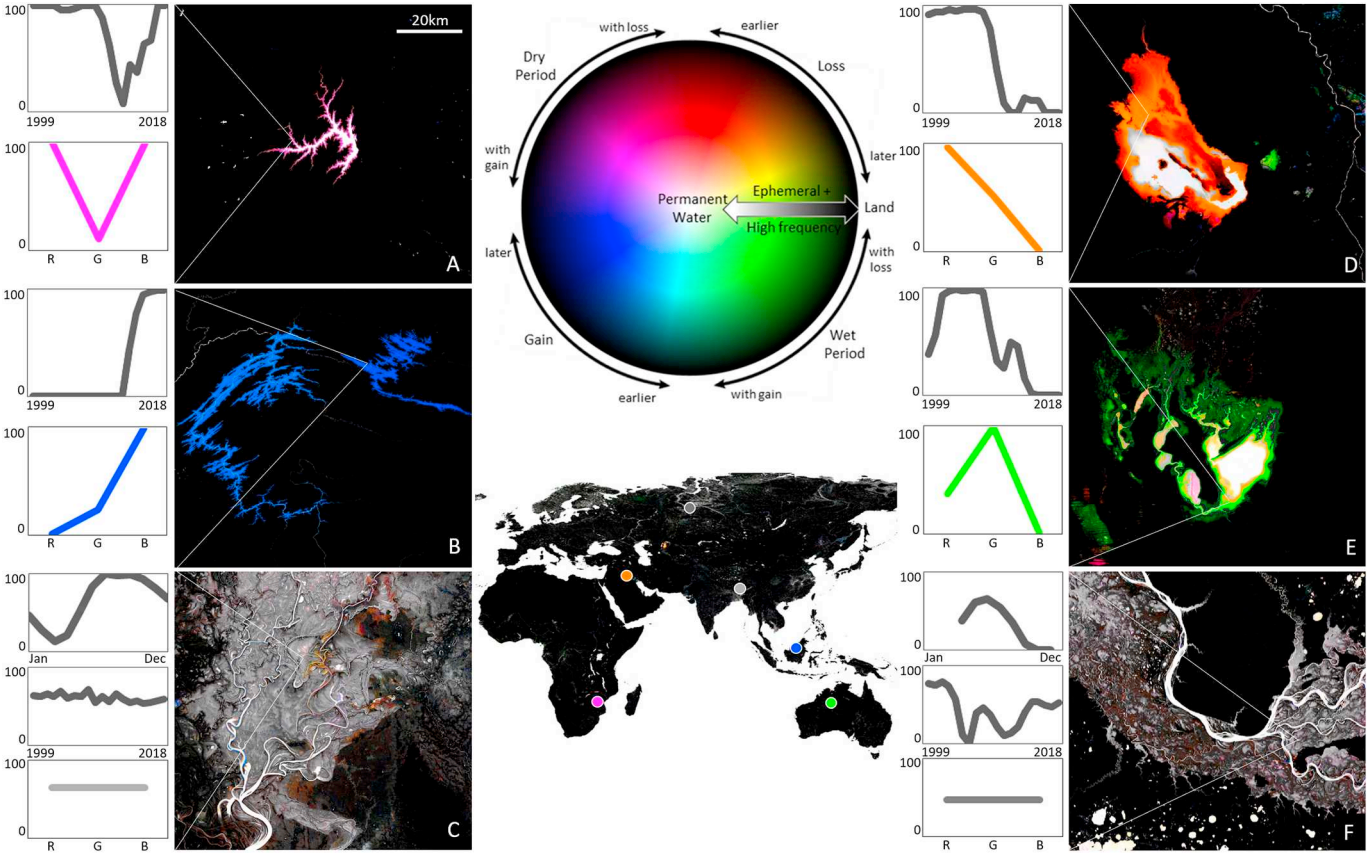


Fig. 2. Examples of classes of the water dynamics map. Each example has the annual water percent time-series and resulting R-G-B reduction for a given pixel. For the stable seasonal and high frequency examples, the 20-year monthly mean water percent is also shown. (A) Dry period: Chicamba Real Dam, Mozambique. (B) Gain: Bakun Dam, Malaysia. (C) Stable seasonal: Meghna River floodplain, Bangladesh. (D) Loss: Razazza Lake, Iraq. (E) Wet period: Lake Gregory, Australia. (F) High frequency: Ob River floodplain, Russia. (Please see web version of this article for color figures.)

minimum then the maximum and then a local minimum which were assigned time-sequentially to the R-G-B space. The maximum was assigned to the green band and the two minima were assigned to the red and blue bands with the greater minimum averaged with the beginning or end value of the time-series depending on whether it was before or after the maximum. For pixels with more than two changes, it was not possible to maintain the structure of the trend in an R-G-B reduction and the mean of the entire period was taken for all three points to show the average annual percent of time for which water was present. These R-G-B reduction heuristics were also applied to the stable seasonal class to show the continuum of changes with intensity < 50%, resulting in pixels with color saturation < 50%.

2.3. Area estimation and accuracy assessment

Two complementary reference datasets were collected: one for accuracy assessment of the water and land classification within a single month and one for accuracy assessment of the classification of surface water dynamics and area estimation of categories of change. The first reference dataset was created at 5 m resolution from RapidEye imagery and was used to assess the accuracy of the base monthly surface water maps in discriminating land and water and to examine spatial heterogeneity of surface water within the Landsat pixels (Section 2.3.1). The 5 m resolution sample data were not used to evaluate change accuracy or to estimate area. To evaluate inter-annual dynamics, we performed a separate analysis. The only source of reference data available at 30 m resolution or finer with global-scale observations since 1999 is the Landsat archive. Consequently, we were constrained to use Landsat to produce area estimates of surface water extent and change for the 1999–2018 monitoring period and to quantify map accuracy (Section 2.3.2). An overview of the work flow can be seen in Figure 3.

2.3.1. Single-date RapidEye (5 m) sample

To quantify accuracy for mapping water at a given instance at a subpixel scale, we used a stratified sample of 5 m resolution reference imagery from RapidEye. This sample was used to estimate accuracy of water and land classifications only at the individual month time scale and was not suitable for evaluating change. To create strata, we divided the global land surface into 20x20km blocks and calculated the percent water cover in each block based on the water mask created by Hansen et al. (2013) that they termed “datamask” within their global forest change product (Fig. 4). This mask represents all pixels with water detection percent ≥50% for all clear growing-season observations of 2000–2012 from Landsat 7. Blocks that had no water pixels, no pixels with > 0% tree cover, and all pixels with ≥95% bare ground were defined as desert blocks and excluded from the sampling frame (gray areas, Fig. 4). Blocks that were entirely water in all the observations, found exclusively in very large lakes and seas, were also excluded. The remaining 307,195 eligible blocks were divided into four near equal size strata corresponding to block water cover of 0%, 0–0.08%, 0.08%–2%, and > 2% (Fig. 2).

Thirty-five blocks were randomly selected per stratum, and a 5 m multispectral image was obtained from RapidEye for each block. Since we mapped each RapidEye image individually, no radiometric correction was necessary, other than to manually remove cloud or other artifacts from the image. However, RapidEye images were mis-registered by up to 40 m when using reported ephemeris data. To overcome this issue, we implemented a post-processing step of shifting the RapidEye classifications to the x-y offset that yielded the greatest water overlap with the water mask of Hansen et al. (2013).

All sample RapidEye images were from 2010 to 2013 growing seasons based on availability, and each sample image was compared against the monthly percent water layer from this current study corresponding to the month in which the image was taken. This time range allowed for data from Landsat 5, 7, and 8 to all be represented. The RapidEye single date images and monthly aggregate map products were not coincident given the varying acquisition rates of Landsat and RapidEye data. However, the majority of surface water is stable at monthly time scales and, in general, any non-matching data will lower the accuracy estimates rather than inflate them. Some blocks for which the RapidEye data could not be obtained or that had over 25% cloud cover were replaced by selecting the next eligible block from a randomly ordered list of sample blocks up to 15 blocks. Some blocks did not have any Landsat data from the corresponding month and were excluded. The final sample sizes were 28, 33, 29, and 26 for the very low, low, medium, and high water cover strata, respectively, due to missing data.

Each RapidEye image was individually trained and classified through an iterative process of delineating water and land training polygons and then running an image-specific set of seven bagged classification trees to classify the entire sample block. The classification trees were built on the five spectral bands in RapidEye imagery and clouds and shadows were manually masked. For many blocks, we iterated and added training polygons numerous times to obtain high quality maps at 5 m. These maps are suitable as reference data as the higher spatial resolution enables a more discrete mapping of surface water that is readily identifiable through a supervised mapping approach (Olofsson et al., 2014). The result is an independent and better characterization of water extent for the respective date than the corresponding Landsat map made using a global algorithm.

The resulting 5 m discrete map of water, land, and no data was the reference set compared against the Landsat monthly percent water layers to obtain user's (corresponding to commission) and producer's (corresponding to omission) accuracies at 5 m and 30 m resolutions. For this analysis, the Landsat monthly percent water was thresholded so that all pixels above 50% were labeled as water and all pixels below 50% were labeled as land, as the dominant state of land or water was more likely to match the surface conditions of when the RapidEye scene was imaged. Pixels with equal land and water observations were excluded to prevent introducing a bias, since there is no dominant state.

To estimate user's accuracy of the monthly mapped water class we used a ratio estimator (Stehman, 2013):

$$\widehat{R} = \frac{\sum_{h=1}^{H} y_h}{\sum_{h=1}^{H} x_h} \tag{1}$$

where H is the number of strata, Y_h is the total area of intersection between the Landsat monthly map of water and the reference classification from all sampled blocks of stratum h with the area from block i denoted y_i, X_h is the total area of mapped water within the Landsat map from all sampled blocks of stratum h with the area from block i denoted x_i, and $\widehat{R}$ is the estimator for user's accuracy. For producer's accuracy we used the same formula keeping the same definition for Y_h and y_i but now defining X_h to be the total area classified as water in the reference data from all sampled blocks of stratum h and x_i the area of this region in block i. The estimated variance of the ratio estimator is:

$$\widehat{V}(\widehat{R}) = \frac{1}{\widehat{X}^2} \sum_{h=1}^{H} N_h^2 \left(1 - \frac{n_h}{N_h}\right)(s_{yh}^2 + \widehat{R}^2 s_{xh}^2 - 2\widehat{R} s_{xyh})/n_h \tag{2}$$

where ${s_{yh}}^2$ and ${s_{xh}}^2$ are the sample variances of y and x in stratum h, $s_{xyh} = (\sum_{i=1}^{n_h} (x_i - \bar{x}_h)(y_i - \bar{y}_h))/(n_h - 1)$ is the sample covariance between x and y in stratum h, n_h is the sample size in stratum h, N_h is the total number of blocks in stratum h, and $\widehat{X} = \sum_{h=1}^{H} N_h \bar{x}_h$. The discrete 5 m reference layers were also upscaled to continuous 30 m maps of water with each 30 m pixel representing the percent of 5 m pixels within it that were labeled water. These new maps were also thresholded at 50% to create the 30 m resolution, binary classified (land and water) reference set and we applied the ratio estimator to obtain user's and producer's accuracies of the monthly Landsat maps at 30 m.

To assess the spatial heterogeneity of surface water, we used the continuous 30 m reference data compute the percent of pixels, excluding pure land pixels, that were mixed, defined as having both water

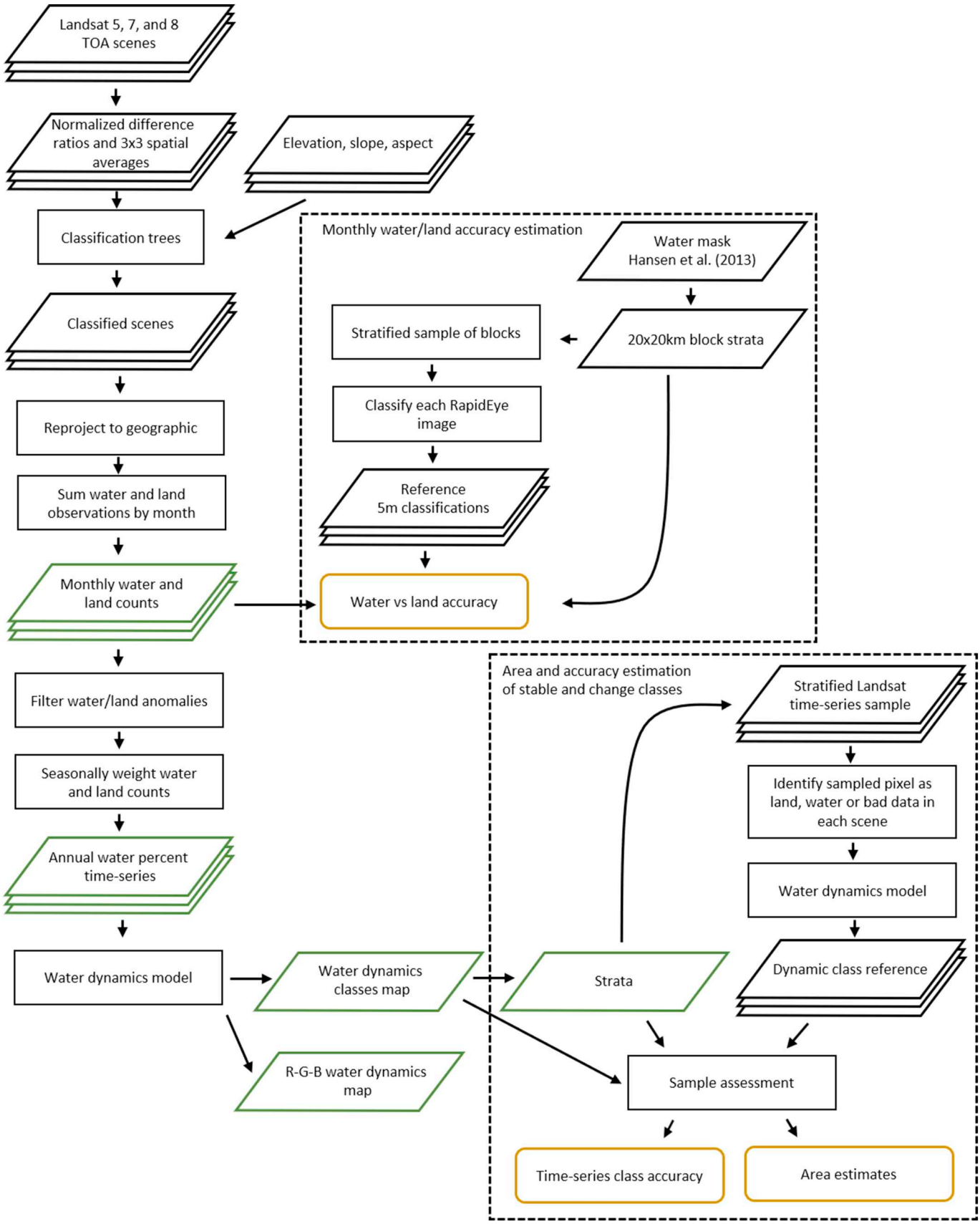


Fig. 3. Overview of workflow with the maps created in this study in green boxes and output area estimates and accuracies in yellow boxes. All maps and reference data are available at: www.glad.umd.edu/dataset/global-surface-water-dynamics. (Please see web version of this article for color figures.)

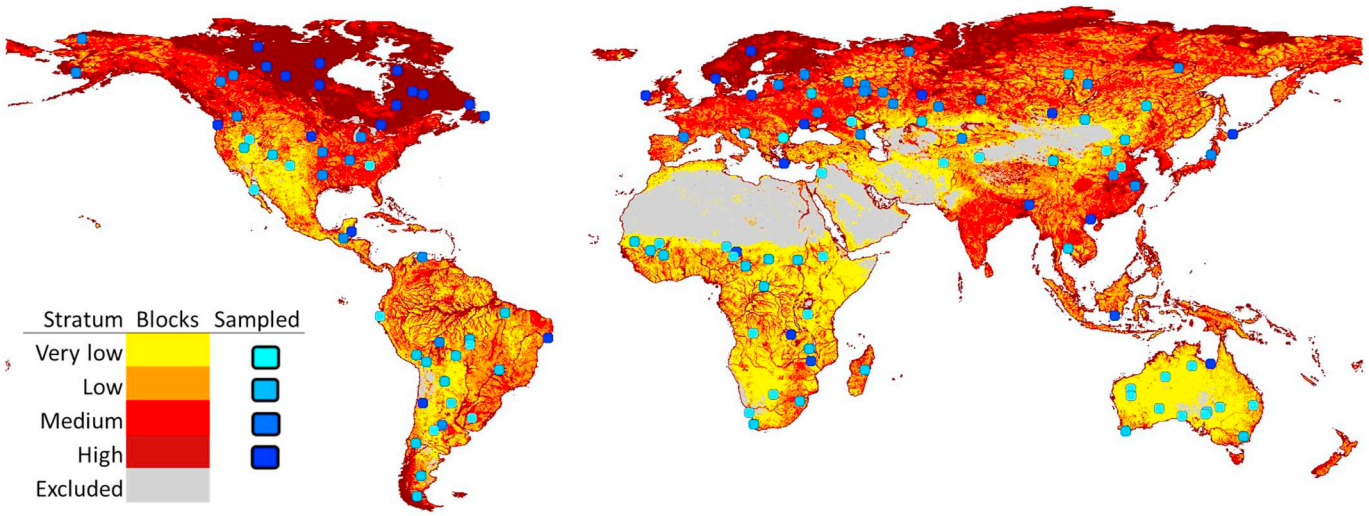


Fig. 4. Distribution of strata and sampled blocks for 5 m assessment. The thresholds of the percent of block area covered by surface water for the very low, low, medium, and high strata are 0%, 0–0.08%, 0.08%–2%, and > 2%.

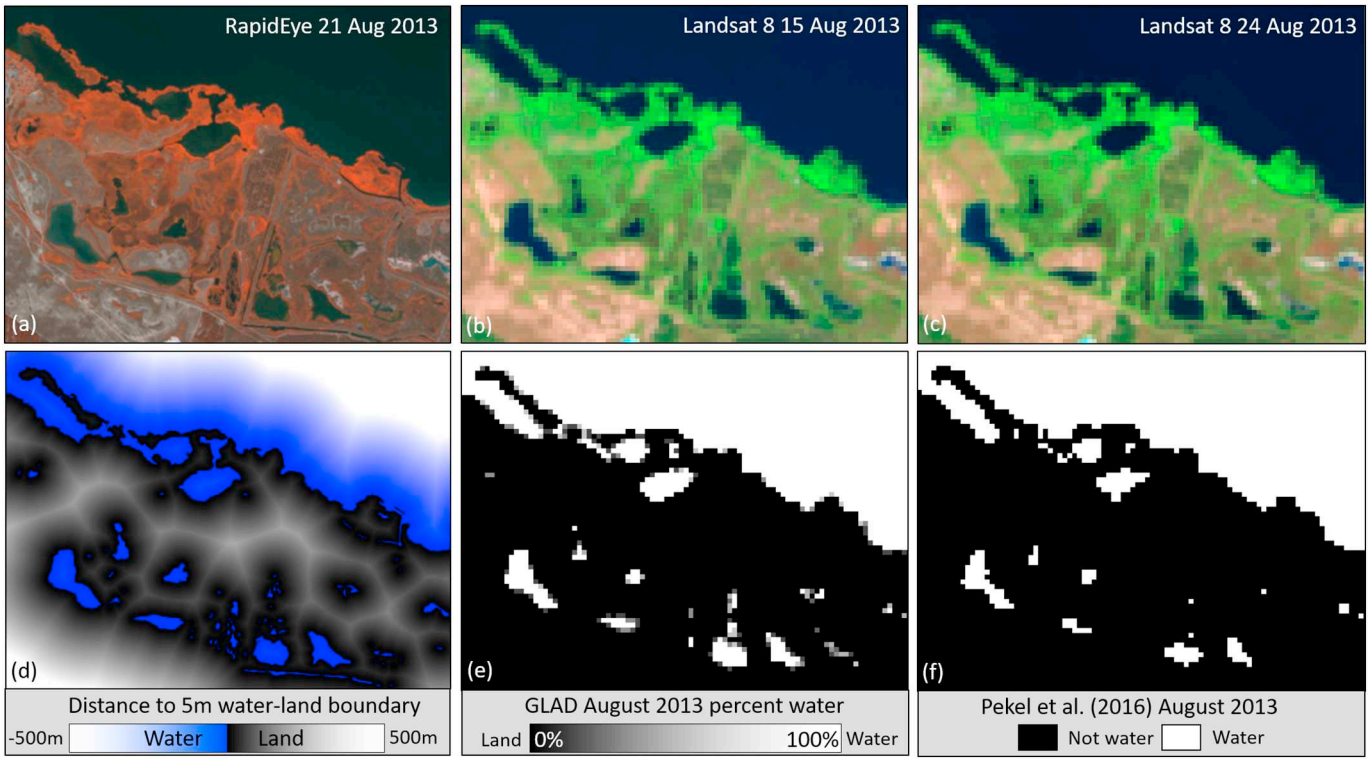


Fig. 5. Zoom within a sample block in the high water stratum. Top-left: RapidEye data (NIR-Red Edge-Red) from August 21, 2013. Top-middle: Landsat data (SWIR1-NIR-Red) from August 15, 2013. Top-right: Landsat data (SWIR1-NIR-Red) from August 24, 2013. Bottom-left: 5 m RapidEye classification of water and land with the distance to the closest water-land boundary, the blue gradients are classified as water and the gray gradients are classified as land. Bottom-middle: GLAD percent water for August 2013. Bottom-right: Pekel et al. (2016) not water and water classification for August 2013. Image centered at 46.52°N, 31.84°E on the Ukrainian coast of the Black Sea. (Please see web version of this article for color figures.)

and land pixels at 5 m. To explore the impact of these mixed pixels on the accuracy results as well as to determine the relationship between errors and the proximity to a land-water boundary, we calculated the distance to the nearest land-water boundary delineated in the 5 m RapidEye classification (Fig. 5). Distances were calculated in meters for every 5 m pixel within each sample block with each 5 m water pixel assigned the distance to the closest land pixel and each 5 m land pixel assigned the distance to the closest water pixel. The distances were binned in 5 m intervals and commission and omission rates were computed per bin using the ratio estimator (Eq. (1)). For commission, y_i is defined as the area classified as water in the monthly map and land in the reference data within the given distance bin and x_i as the reference land area within the distance bin from the water-land boundary. For omission, y_i is defined as the area classified as land in the monthly map and water in the reference data within the given distance bin and x_i as the reference water area within the distance bin from the water-land boundary. User's and producer's accuracies were also computed for the whole region within 30 m of a land-water boundary which represents the potential mixed pixel region at Landsat scale and for the whole region beyond 30 m from land-water boundary which represents the

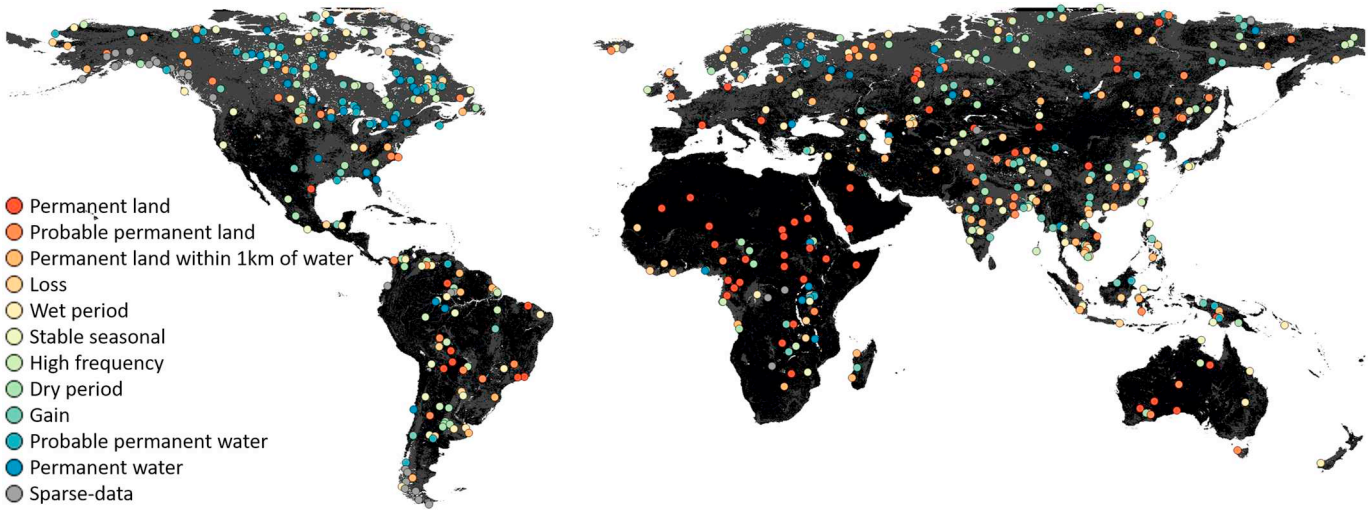


Fig. 6. Distribution of sampled pixels of the Landsat time series.

pure land or pure water pixel space at Landsat resolution. The spatial heterogeneity of surface water was further explored by estimating the percent of global inland water that is within mixed pixels at Landsat resolution. We used the ratio estimator (Eqs. (1) and (2)) with y_i representing the area of water within mixed pixels in block i and x_i representing the total area of water in block i.

2.3.2. Landsat time-series sample

A probability sample-based assessment was conducted to estimate areas and validate the inter-annual dynamics map from 1999 to 2018. The map was clipped using the Global Administrative Areas dataset (University of California, Berkley 2012) to exclude coastal and ocean waters, consistent with previous studies (Feng et al., 2016; Pekel et al., 2016). This assessment is for all areas with data within 56°S-75°N, excluding Greenland.

Twelve strata derived using the mapped classes were created from the time-series analysis, and fifty 0.00025° pixels were randomly selected per stratum, totaling 600 sampled pixels (Fig. 6). Mapped permanent water was separated into two strata: 1) the high confidence permanent water stratum which included all pixels whose R-G-B transformation resulted in each value ≥90%, and 2) the likely water stratum which included all pixels that did not meet the criteria defining the first stratum but that had mean annual open water present for ≥90% of the observable portion of the year and inter-annual variation ≤33%. Permanent land was separated into three strata to target possible omission errors. The land buffer stratum included all land pixels within 1 km of any water class, the high confidence land stratum included all remaining land pixels whose R-G-B transformation resulted in each value ≤10%, and the likely land stratum included all land pixels that did not meet that criteria but that had mean annual open water present for ≤10% of the observable portion of the year and inter-annual variation ≤33%. There was an additional sparse-data stratum which was defined as all pixels with < 10 years of observations or < 15 total observations. The sparse-data stratum was used for area estimation, but not for the accuracy assessment because it did not have a map dynamic type label.

Stratification was essential to ensure that sufficient sample sizes were allocated to each class given that inland surface water only covers 3–4% of the planet and of that only a small fraction belongs to each of the dynamic classes (Pekel et al., 2016). It was also necessary to have the strata take into account the entire time period because 40% of surface water area was found to be dynamic. The reference data must also cover the entire time period monitored and the only data source for such a task is the Landsat archive. The standard for reference data is that the reference classification must be of equal or higher quality than the map itself. Manual interpretation of the time-series of individual sample pixels provided this more accurate time-series characterization relative to the output of our global-scale algorithm (Olofsson et al., 2014). For the reference data, an observation from every month was selected as a compromise between exhaustive interpretation of every scene and feasibility that maintained sufficient temporal density to capture inter-annual water dynamics.

Because the data were processed in geographic (Lat/Long) at 0.00025° resolution and not in equal area projection, pixels differ in area depending on latitude. Pixels were sampled with inclusion probabilities proportional to the area of the pixel using the following method (Brewer and Hanif, 1982). All pixels were listed per stratum and the cumulative sum of pixel areas was computed for all pixels previous to and including the current pixel in the list. Fifty floating point numbers between zero and the total stratum area were randomly generated. For each of the randomly selected numbers, the first pixel that had cumulative area larger than this number was selected. This protocol results in a stratified random sample for which the inclusion probability of a pixel in each stratum is proportional to the area of the pixel.

For each sampled pixel, a Landsat observation was randomly selected from each month for each year, resulting in up to 240 scenes being visually interpreted for each pixel. If for a given scene the sampled pixel was flagged as cloud, a new scene would be randomly selected from the same month if available, otherwise the original scene would be retained for interpretation. For each sampled pixel, a html page was built with thumbnails of all selected Landsat scenes (Fig. 7). For each thumbnail, the pixel was labeled as land, water, or bad data through visual interpretation of the Landsat data and auxiliary high-resolution data from Google Earth was used to provide additional context. Since the data are resampled to 0.00025° via the nearest neighbor method, each 0.00025° pixel retains the spectral reflectance data of the nearest 30x30m pixel in the original Landsat data, which is in Universal Transverse Mercator (UTM) projection, and the footprint of the UTM pixel was utilized when considering the high-resolution data from Google Earth.

These land and water labels in the reference data were used to evaluate the inter-annual water dynamics map. A total of 87,926 scenes (600 sampled pixels × 12 months × 20 years minus months with no available scenes) were visually interpreted as land, water, or bad data, and out of these, 57,230 observations were labeled as land or water both through visual interpretation and via the map classification trees. While this is a large number of scenes, it was manageable because only

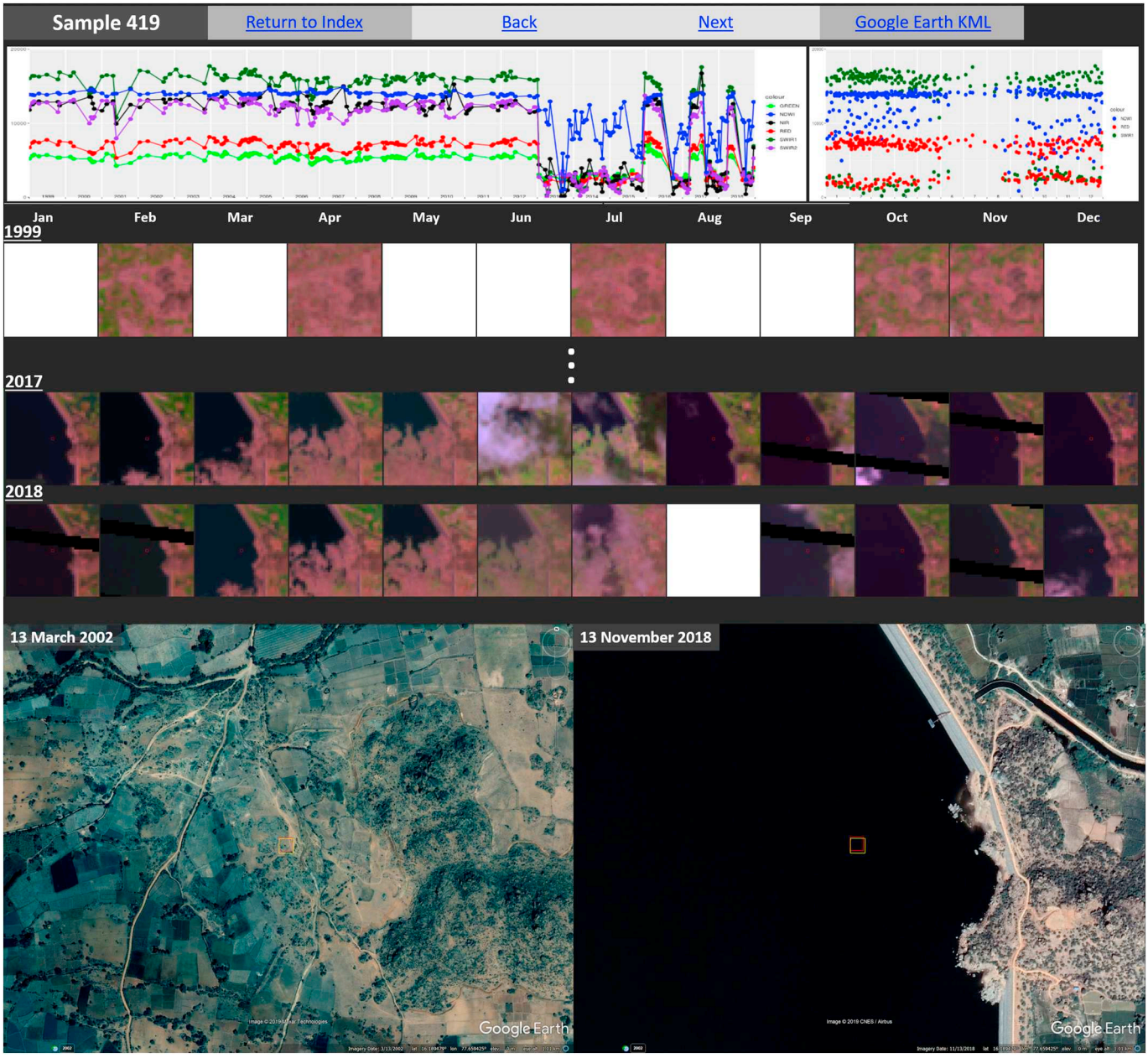


Fig. 7. Example sampled pixel from the wet period stratum, centered at 16.189375°N, 77.659375°E within the fluctuating footprint of Rayalumpad Reservoir in Telangana, India. The reservoir was constructed in 2009, first flooded the sampled pixel in 2013 and remained flooded through 2015, and then seasonally inundated the sampled pixel 2016–2018. The pixel is outlined in red in each of the Landsat thumbnails (SWIR1-NIR-red) and in the images from Google Earth, with an additional yellow outline for the source UTM pixel footprint. The graphs at the top are time series of the spectral reflectance of the pixel to aid interpretation (Green reflectance in light green, red reflectance in red, NIR in black, SWIR1 in dark green, SWIR2 in purple, and NDWI in blue). Each Landsat scene could be individually marked as land, water, or bad data or a month, year, or the whole collection could be labeled as one of these classes. The full reference html page includes all months and years 1999–2018 and can be viewed here: https://glad.geog.umd.edu/timeSeriesReference/pagesUTM/sample419.html. (Please see web version of this article for color figures.)

the 600 sample pixels needed to be interpreted, many of which were stable through time. The individual visual interpretations were aggregated to form an annual time series by calculating the percent of clear reference observations that were water per year and smoothed with a 3-year mean moving window as was done with the algorithm-generated time-series. This reference time-series was input to the water dynamics model (Section 2.2) to generate class labels for each sampled pixel.

These class labels of the reference data were used to estimate area of the water dynamic classes and aggregations of the classes. Additionally, the area that is inundated each year for > 25, 50, 75, and 90 and equal to 100% of the time was calculated from the annual percent values of each sampled pixel. Using this same stratified sample, we estimate user's and producer's accuracies for the trend classes. Details of area and accuracy estimation calculations can be found in Appendix A.

2.3.3. *Comparison with Pekel et al. (2016) dataset*

The most comprehensive previous global surface water dataset is that of Pekel et al. (2016) with monthly water/not water maps from 1984 to 2018 with considerable data gaps pre-1999. They evaluate their product using a sample-based assessment that only quantifies water/not water state. The area estimates of change they provide are

not produced using a probability-based sample of reference data but instead change is quantified from pixel counts, which ignores the biases inherent in the map products. Moreover, the multi-temporal maps made by Pekel et al. (2016) that identify change are principally bi-temporal, consisting of the change in percent of an epoch with open water present between two epochs, and a time-1 time-2 transition map. The transition map of Pekel et al. (2016) is derived from the first year with sufficient surface observations in the Landsat Thematic Mapper time-series, starting in 1984 through 2000, and a last year of 2018, for the map update released this year. The transition map represents the classes in a transition matrix generated from the annual labels of permanent, seasonal, and land in the first and last year, an approach that ignores all intervening years except for pixels where the first and last year both have land labels. Permanent water on an annual time-scale is defined as all months with data flagged as water, and seasonal water is defined as having at least one month flagged as water and at least one month flagged as land. In the case of land labels for both the first and last year, the annual time-series is checked whether or not there was ever labeled water between the start and end years. If there is such ephemeral water, it is determined whether permanent or seasonal water states were observed more frequently and the pixel is labeled ephemeral permanent or ephemeral seasonal according to a majority rule. If either the first or last year are water, the class label is assigned only using bi-temporal first and last year labels.

Pekel et al. (2016) performed an extensive validation on their water/not water classification, but did not include the entirety of the map in the sampling frame. Importantly, their omission sampling frame for south of 60°N was only within the SRTM Water Bodies Dataset (SWBD) and for north of 60°N within the water class of Feng et al. (2016) and did not validate change. In SWBD, all water bodies with a width smaller than 183 m are removed, and thus the omission accuracy only applies to bodies of water having width larger than 183 m that are within the SWBD, a data set derived from a period of 10 days in February 2000. Finally, they did not use the reference data to report map-based area summations or to estimate uncertainty bounds of reported areas. In this study, we extended the results of the Pekel et al. (2016) study by producing maps that report various change classes based on all the years of the study period, conducting an accuracy assessment that represented the entire map, and reporting sample-based estimates of area of change accompanied by standard errors to quantify uncertainty of these estimates.

We evaluated Pekel et al.'s (2016) global water dataset in three different ways. (1) To evaluate the monthly water product of Pekel et al. (2016), we calculated accuracies for the monthly water/non-water labels of Pekel et al. (2016) using the same single-date 5 m reference data set described in Section 2.3.1 and performed the same steps of accuracy analysis as we did for our monthly product, which allows for accuracy estimates that apply to the whole map. Pekel et al.'s (2016) monthly data are labeled as water, land, or no observation and thus no threshold was needed for the single-date validation. (2) In order to compare the sensitivity of our monthly water layers with those of Pekel et al. (2016) to detect and correctly identify change, we applied the same water dynamics model we developed in this study (Section 2.2) to the monthly water history of Pekel et al. (2016). This task was performed using the same rules applied in the creation of our change categories with the input monthly water history being from Pekel et al. (2016) instead of GLAD and with their water labels converted to 100% and not-water labels to 0% on a per month basis. This enabled us to create an annual time-series from monthly data in an identical manner for Pekel et al. (2016), GLAD, and the reference data set. Water dynamics were therefore modeled consistently across all three data sets. Accuracies for Pekel et al. (2016) and GLAD were subsequently assessed using the same method. (3) To evaluate the transition map of Pekel et al. (2016), we compared their transition labels against the water dynamics class of the reference data. Since the readily available transition map of Pekel et al. (2016) is from the first representative year to 2018 rather than 1999–2018, we computed a 1999–2018 transition map from their yearly water history according to the rules outlined in their published study. A one-to-one correspondence does not exist between the transition map labels and the class definitions developed in this study. In this case, we did not compute accuracies, but instead we have provided the confusion matrix. In this way, we are able to quantitatively compare the Pekel et al. (2016) bi-temporal water dynamics to a reference data set that is stratified on change and that more completely characterizes surface water dynamics.

3. Results

3.1. Maps

Monthly, seasonal, and annual water percent layers were created and maps of inter-annual dynamics for 1999–2018 were created for the entire year as well as for each set of three consecutive months. Natural dynamics such as the meandering of the rivers in the Ganges-Brahmaputra basin (Coleman, 1969) are observed as well as direct human change such as the expansion of rice cultivation and aquaculture (Davidson, 2014; Tessler et al., 2016; Zhao et al., 2008) (Fig. 8). Large areas of water gain include the lakes of the Tibetan Plateau (Zhang et al., 2014), the prairie potholes of the USA and Canada (Zou et al., 2018), and the creation of reservoirs, particularly in Southeast Asia (Zarfl et al., 2015). Many of the large saline lakes of the world have all substantially diminished in area since 1999 (Wurtsbaugh et al., 2017). Much of the area with multiple transitions between open water and land occur within wetlands and floodplains. In addition to visualizing inter-annual dynamics, mean water percent was also calculated per month for the period of 1999–2018, enabling analysis of seasonal water presence (Fig. 9). Consistent annual seasonal open water can be seen in many of the floodplains around the world, for example, the Barotse floodplain in Zambia (Cai et al., 2017). Direct human seasonal dynamics such as single and double cropping of rice paddies can also be distinguished through the annual or biannual open water flooding regime. While much of the measured seasonal and inter-annual fluctuation is due to true variation in surface water presence, some of this fluctuation in open water presence, particularly in wetlands, is due to variation in vegetation levels that obscure the water surface. All layers are available for visualization or download at www.glad.umd.edu/dataset/global-surface-water-dynamics.

We evaluated the effects of the filtering and smoothing of the time-series on the output inter-annual surface water dynamics maps. The filtering of anomalous water detections caused 0.17% of the final area mapped as land to be classified as land rather than a dynamic class and the filtering of anomalous land detections caused 0.29% of final area mapped as permanent water to be classified as permanent water rather than a dynamic class. We evaluated the impact of using 30% as the magnitude threshold defining a transition and found that if instead the threshold was set at 10, 20, 40, or 50% that 0.16–0.33% of the entire continental area changes class or, equivalently, 5.1–8.7% of the area mapped as permanent or dynamic water classes. Using the single year annual time-series as input to the water dynamics model instead of the 3-year mean annual time-series has a much larger impact, with 1.8% of the entire continental area changing class, or as a percent of the area mapped as permanent or dynamic water classes 45% changes class and increases the total area mapped as dynamic by 4.2%.

3.2. Sample-based area estimates

We estimated 2,928,992 (± 93,027) km^2 of the continental area to be permanent open water from 1999 to 2018 (Table 1). Areas that were open water at any point during 1999–2018 totaled 4,815,478 (± 82,986) km^2. There was a total of 138,011 (± 28,163) km^2 that had unidirectional gain of open surface water and 53,154 (± 10,883) km^2 that had unidirectional loss of open water, whereas, areas that

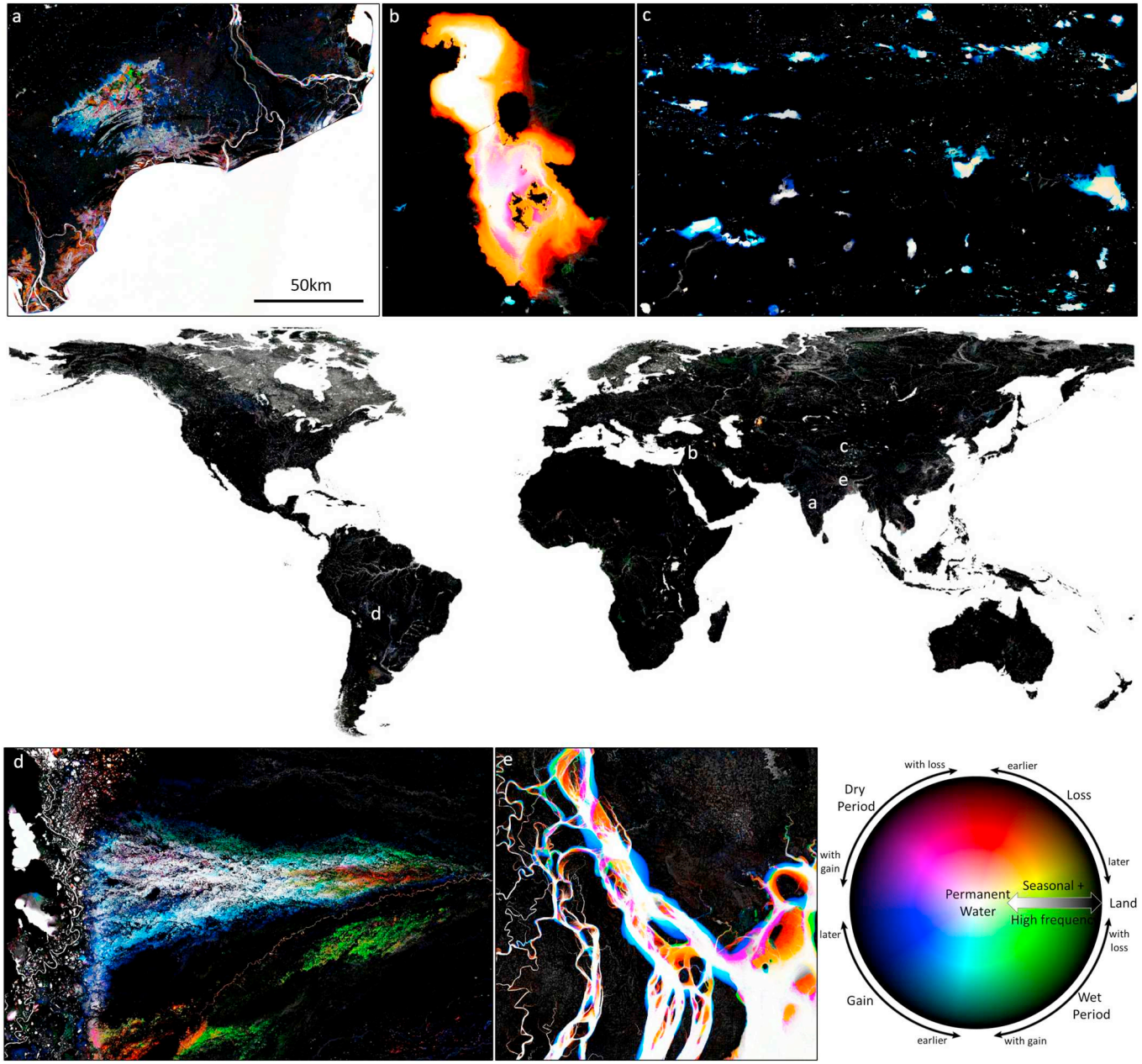


Fig. 8. 1999–2018 inter-annual water dynamics map examples for all 12 months with hues representing the type of change dynamic, saturation the intensity of the change, and value or brightness representing the maximum percent of a year that was inundated. All examples are shown at the same scale. (a) Expansion of aquaculture and shifting management practices on the coast of India. (b) Dramatic reduction of Lake Urmia in Iran, with slight recovery. (c) Increase of lakes across the Tibetan Plateau. (d) Shifting open water patterns in the Pantanal in Brazil, the largest wetland in the world. (e) Meandering of the Meghna River in Bangladesh by 10 km and the growth of new islands in the Bay of Bengal. (Please see web version of this article for color figures.)

changed between persistent water and persistent land two or more times through the period totaled 950,719 (± 104,034) km^2. The mean annual area with open water present for ≥50% of the observable portion of the year was 3,687,781 km^2 (standard deviation over all years of the annual area estimates = 46,440 km^2).

3.3. Accuracy

3.3.1. Accuracy of monthly water and land classification

The user's and producer's accuracies (corresponding to commission and omission rates) at 30 m resolution of the monthly mapped water class defined by a threshold of 50% applied to the percent water are 93.7 (± 1.5)% and 96.0 (± 1.2)% respectively. The accuracies of the monthly mapped water class of Pekel et al. (2016) are 95.2 (± 1.2)% and 90.3 (± 2.3)% using the same reference sample data. All of the accuracies presented in this subsection refer to these water classes for different resolutions and regions. At 5 m resolution, the user's and producer's accuracies in our study are 97.5 (± 0.7)% and 97.7 (± 0.7)% and Pekel et al.'s are 98.0 (± 0.5)% and 95.1 (± 1.3)%. These numbers are heavily impacted by the region that is within 30 m of the water body edge, both within the water body and in adjacent land, as can be seen when the water class commission and omission rates are plotted as a function of distance away from the land-water boundary as delineated in the 5 m reference data (Fig. 10). At a distance > 30 m, the mapped water class of our study has a user's accuracy of 96.9 (± 0.9)% and a producer's accuracy of 99.0 (± 0.3)%, and the mapped water class of Pekel et al. (2016) has user's and producer's accuracies of 99.4 (± 0.2)% and 99.5 (± 0.1)%. For the area 0-30 m

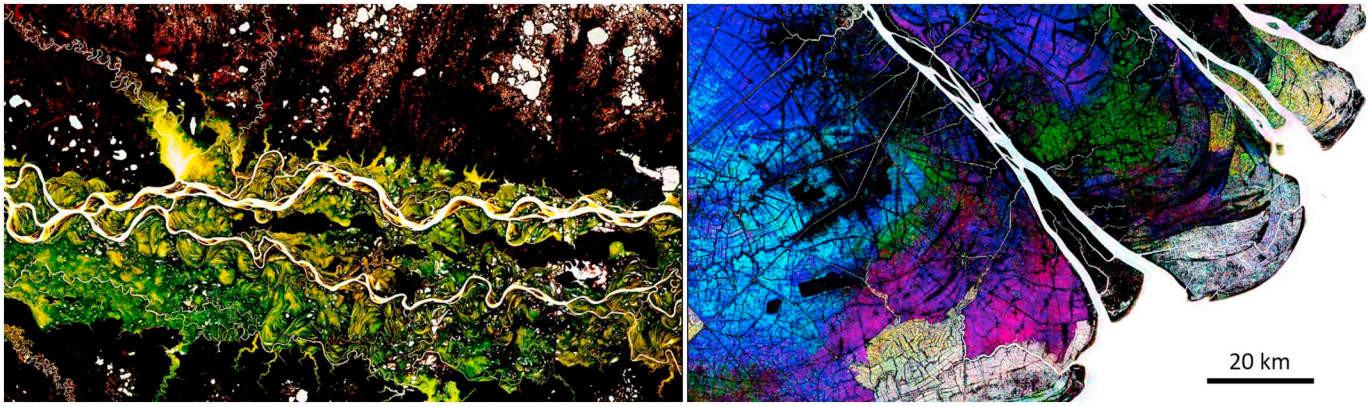


Fig. 9. 1999–2018 seasonal water examples. Both examples are shown at the same scale. Left: R-G-B of monthly mean water percent of April-June-August showing the seasonal floodplains of the Ob River, Russia. Right: R-G-B of monthly mean water percent of June-August-October showing the various rice and aquaculture flooding cycles in the Mekong Delta, Vietnam with the magenta areas experiencing two separate open water flooding periods and the green areas experiencing three separate open water flooding periods. (Please see web version of this article for color figures.)

from the closest land-water boundary, our study's user's and producer's accuracies are 70.9 (± 1.6)% and 66.0 (± 3.7)3% and Pekel et al.'s user's and producer's accuracies are 75.4 (± 2.0)% and 51.6 (± 3.3)%. Thus both datasets map water with high accuracy beyond 30 m from shore lines and river banks, but both have difficulty distinguishing land and water when they are blurred together in mixed pixels. However, considering only the edge and mixed pixels, our map has less bias and correctly identifies as water 28% more of the 5x5m footprints with water cover, and for this region within 30 m of the land-water boundary two thirds of 5 m space is correctly assigned. These contrasting commission and omission rates of the static water classification associated with mixed pixels carry over to the following change detection accuracy results because of the large area of mixed pixels. We found that 10.9 (± 1.9)% of global inland water is within mixed pixels at Landsat resolution, defined as having both land and water at 5 m resolution within the pixel. The most common sources of commission error aside from mixed pixels are over lava flows, urban centers with tall buildings and shadowed streets, glaciers (particularly when mixed with debris), and cloud shadow over dark, dense forests with persistent cloud cover such as in Gabon.

3.3.2. Accuracy of classes of water dynamics derived from time-series

The accuracy of each class of water dynamics derived from our dataset for 1999–2018 varied greatly by the number of changes represented by each class. The same was true for the accuracies of the classes derived from the monthly layers of the study by Pekel et al. (2016) when the same water dynamics model developed in our study was applied to the Pekel et al. (2016) data. The resulting map and labels derived from the water dynamics model were not produced by Pekel et al. (2016) but it is instructive to examine the accuracy of their product when their data are translated through this model. Accuracies were highest for land with the user's and producer's accuracies of our study at 99.9 (± 0.0)% and 99.7 (± 0.1)%, followed by permanent water with user's and producer's accuracies of 97.8 (± 1.8)% and 85.8 (± 2.4)%, both of which had similar accuracies for Pekel et al.'s (2016) dataset. These are the largest classes, accounting for 98.6% (± 0.1%) of the global continental area.

The accuracies of each of the classes where pixels sometimes are land and sometimes have water, hereafter referred to as 'all dynamic types', are much lower and for the GLAD dataset have higher rates of commission than omission (Table 2). For each of the dynamic types the data from our study have lower rates of omission than the results from using Pekel et al.'s (2016) monthly water history, and for three out of the six dynamic types our study has lower rates of commission. However, many of the misclassifications are between the change and seasonal classes themselves, rather than missing the transitory nature of the surface water entirely, as can be seen by the accuracies when all dynamic types are aggregated into a single class. Our study resulted in user's and producer's accuracies for the class "all dynamic types" of 68.3(± 3.9)% and 93.4(± 0.6)%, respectively. Pekel et al. (2016) yielded user's and producer's accuracies for "all dynamic types" of 46.1(± 19.3.)% and 60.2(± 4.9)%. However, one of the sampled pixels in the land buffer stratum was labeled stable seasonal using the Pekel et al. (2016) dataset, and if this one sample unit is removed from the analysis, the user's accuracy of "all dynamic types" increases to 78.5(± 5.0)% and the producer's accuracy remains the same. Thus the

Table 1
Area estimates of trends with associated uncertainty of open water within the continental area between 75°N and 56°S as delineated by the Global Administrative Areas dataset (University of California, Berkeley 2012) and excluding Greenland. The last four categories are aggregates of the classes above.

	Area (± SE) km^2	Percent of total area	Percent of all water	Class definition
Land	126,971,335 (± 82,868)	96.34 (± 0.06)	–	Mean water percent ≤10% and inter-annual variability ≤33%
Permanent water	2,928,992 (± 93,027)	2.22 (± 0.07)	60.82 (± 1.93)	Mean water percent ≥90% and inter-annual variability ≤33%
Stable seasonal	735,347 (± 99,792)	0.57 (± 0.08)	15.69 (± 2.07)	Intra-annual variability with inter-annual variability < 50%
Gain	138,011 (± 28,163)	0.10 (± 0.02)	2.87 (± 0.58)	Land-dominant to water-dominant
Loss	53,154 (± 10,883)	0.04 (± 0.01)	1.10 (± 0.23)	Water-dominant to land-dominant
Dry period	47,344 (± 8982)	0.04 (± 0.01)	0.98 (± 0.19)	Water-dominant to land-dominant to water-dominant
Wet period	120,543 (± 49,543)	0.09 (± 0.04)	2.50 (± 1.03)	Land-dominant to water-dominant to land-dominant
High frequency	784,417 (± 93,460)	0.60 (± 0.07)	16.29 (± 1.94)	3+ transitions between water-dominant and land-dominant
Multiple transitions	950,719 (± 104,034)	0.72 (± 0.08)	19.74 (± 2.16)	Dry period, wet period, and high frequency (2+ transitions)
All change types	1,141,884 (± 106,120)	0.87 (± 0.08)	23.71 (± 2.20)	Gain, loss, dry period, wet period, high frequency
All dynamic types	1,895,159 (± 114,006)	1.44 (± 0.09)	39.36 (± 2.37)	Gain, loss, dry period, wet period, high frequency, stable seasonal
All with water	4,815,478 (± 82,986)	3.65 (± 0.06)	100.00	Permanent water, stable seasonal, gain, loss, dry period, wet period, high frequency

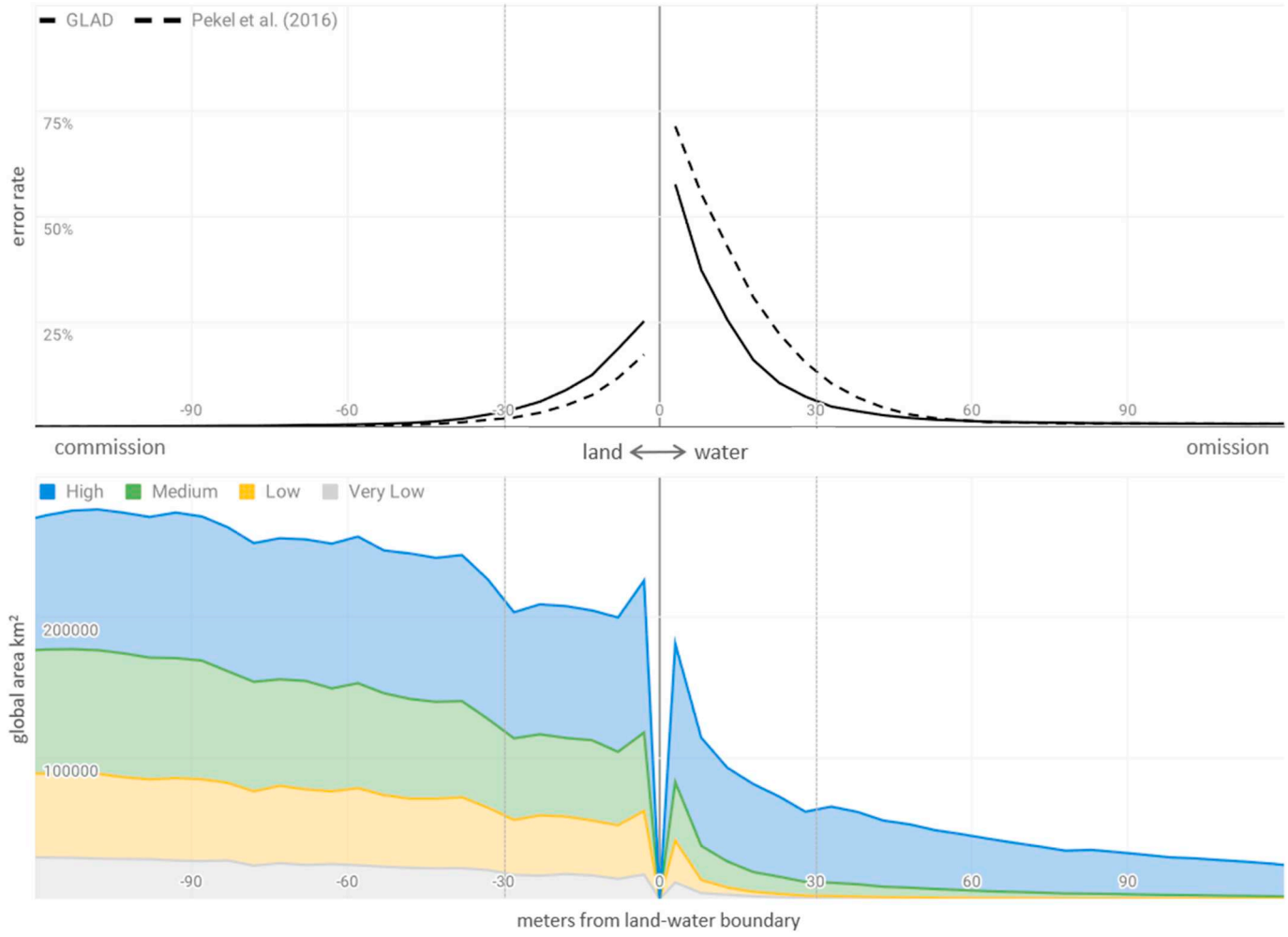


Fig. 10. Error rates at 5 m resolution of the monthly Landsat water detection as a function of distance from the land-water boundary as defined in the 5 m reference data. Top: The left half of the figure has distances extending further into land and displays commission error rates of water for both this study and Pekel et al. (2016) and the right half has distances extending further into water bodies and displays omission error rates of water. Each vertical line represents 30 m from the land-water boundary. The space within 30 m on either side of the land-water boundary represents the mixed pixel space at Landsat resolution. Bottom: The area within each stratum at the given distance from the land-water boundary. The area of water that the low and the very low water strata contribute drops off very quickly due to small water body size. There is no area at 0 m because this is the boundary line between land and water.

user's accuracy result is strongly impacted by one influential "outlier" sample pixel. Through comparison between the producer's accuracies of 'all dynamic types' and of 'all change types', 93.4(± 2.7) % and 70.6(± 6.1)% respectively, we can see that many of the omissions of change occur when the sample pixel is classified in the map as stable seasonal, since that is the only additional class in 'all dynamic types'.

Table 2
The user's and producer's accuracies of the classes of water dynamics from the time-series data of this study and of Pekel et al. (2016) when the water dynamics model is applied to each monthly water history dataset and to the reference data. The last four categories are aggregates of the classes above and a sample pixel is considered correct for the aggregate if it is in one of the contributing classes. See Table 1 for class definitions.

	GLAD accuracy		Accuracy of GLAD water dynamics model applied to Pekel et al. (2016) monthly water history	
	User's	Producer's	User's	Producer's
Land	99.9 (± 0.0)	99.7 (± 0.1)	99.5 (± 0.1)	99.1 (± 0.8)
Permanent water	97.8 (± 1.8)	85.8 (± 2.4)	95.1 (± 1.8)	89.9 (± 2.6)
Stable seasonal	44.0 (± 7.1)	73.0 (± 5.6)	17.4 (± 12.1)	36.3 (± 8.3)
Gain	59.6 (± 7.2)	74.8 (± 13.9)	48.0 (± 12.8)	45.4 (± 10.9)
Loss	30.0 (± 6.5)	86.2 (± 7.4)	49.8 (± 19.3)	65.5 (± 11.4)
Dry period	46.0 (± 7.1)	81.1 (± 11.8)	17.7 (± 8.0)	31.7 (± 9.5)
Wet period	34.0 (± 6.8)	39.8 (± 16.7)	62.5 (± 15.1)	37.5 (± 17.5)
High frequency	54.3 (± 7.4)	54.9 (± 6.2)	50.4 (± 9.3)	35.4 (± 7.1)
Multiple transitions	58.2 (± 5.8)	62.2 (± 6.6)	54.1 (± 7.9)	40.3 (± 6.4)
All change types	60.0 (± 4.6)	70.6 (± 6.1)	62.8 (± 6.9)	50.0 (± 5.8)
All dynamic types	68.3 (± 3.9)	93.4 (± 2.7)	46.1 (± 19.3)	60.2 (± 4.9)
All with water	92.1 (± 1.6)	98.6 (± 0.6)	79.1 (± 15.5)	85.9 (± 2.3)

Table 3
Confusion matrix of GLAD water dynamics map with the reference classification. In row *i* and column *j*, the left half of the cell represents in a gray gradient the estimated percent of pixels labeled *i* in the reference data and that were mapped *j* in GLAD, and the right half of the cell represents in a green gradient the estimated percent of pixels mapped *j* in GLAD that were labeled *i* in the reference data. The grays not along the center diagonal of correct classification show the distribution of omission and the greens show the distribution of commission. The percent estimates were derived from the estimated area of each (*i*,*j*) cell. (Please see web version of this article for color figures.)

GLAD water dynamics classes map 1999-2018

Reference	Land		Permanent water		Stable seasonal		Gain		Loss		Dry period		Wet period		High frequency	
Land	99.7	99.9	-	-	0.2	18.0	0.0	8.5	0.0	28.0	0.0	8.0	0.0	20.0	0.1	10.9
Permanent water	-	-	86.1	97.8	6.8	16.0	0.5	8.5	0.9	18.0	0.2	8.0	0.5	10.0	4.9	19.6
Stable seasonal	7.4	0.0	0.8	0.2	73.0	44.0	1.4	6.4	2.5	12.0	0.7	6.0	0.4	2.0	12.8	13.0
Gain	-	-	-	-	17.9	2.0	74.8	59.6	-	-	1.2	2.0	6.4	6.0	-	-
Loss	-	-	-	-	-	-	-	-	86.2	30.0	3.1	2.0	10.6	4.0	-	-
Dry Period	-	-	12.7	0.2	-	-	-	-	-	-	81.3	46.0	6.0	2.0	-	-
Wet period	-	-	37.4	1.8	-	-	6.1	4.3	2.6	2.0	-	-	39.8	34.0	13.2	2.2
High frequency	1.4	0.0	-	-	31.8	20.0	2.8	12.8	2.1	10.0	3.0	28.0	4.2	22.0	54.9	54.3

The amplitude threshold of the time-series for a pixel to be labeled as change rather than seasonal is 50%, meaning a pixel must gain or lose the equivalent of half a year of open water one or more times through the 20 years of the study. Thus, in this case, the amplitude of the inter-annual changes in the algorithm generated time-series was below the threshold of 50% to be labeled as change and was thus labeled stable seasonal, but in the reference data time-series was greater than the threshold of 50% and was labeled with a change type. The distribution of omission and commission errors can be seen in Table 3.

We also compared the reference data time-series classification to the transition map of Pekel et al. (2016) (Table 4). The transition map underestimates water presence, as a number of water extent and dynamics classes are labeled as permanent land in Pekel et al. (2016). Specifically, 6% of the reference area of permanent water is labeled as land, 43% of seasonal water area is labeled as land, and 41% of the area of all the change classes is labeled as land. Change was also underestimated with an additional 11% of reference change area classified as permanent water in Pekel et al.'s (2016) map. While there is not a one-to-one correspondence between the reference class definitions and Pekel et al.'s (2016) definitions, the nature of the change is also misidentified; for example, 45% of what was classified by Pekel et al. (2016) as new permanent water from land and 52% of lost permanent water to land actually experienced multiple transitions between land-dominated and water-dominated states, the 'high frequency' reference class. The three classes of Pekel et al. (2016) which had the majority of the area correctly mapped within the corresponding reference classes are land with 99%, permanent water with 93%, and ephemeral permanent with 100% within the corresponding reference classes. The remaining classes of Pekel et al. (2016) have the majority of their area in a reference class with a definition that does not overlap. In summary, the Pekel et al. (2016) map omits considerable areas of surface water extent and change and mischaracterizes change types.

4. Discussion

This study provides the first sample-based area estimates of global surface water extent and change. The method employs mapped surface water change to stratify the global land surface and uses a probability sample of reference data to produce unbiased estimators of area of surface water extent and change, accompanied by standard errors to quantify the associated uncertainties of the area estimates. Good practice methodology establishes that areas should be estimated via a sample-based analysis rather than by merely summing the area of class pixels in the map (Olofsson et al., 2014; Stehman, 2013). Given that there are errors in all maps, pixel counting will generally result in an over or under estimation of the true area with unknown magnitude. In

Table 4
Confusion matrix of Pekel et al. (2016) transition map 1999–2018 with the reference classification. In row *i* and column *j*, the left half of the cell represents in a gray gradient the estimated percent of pixels labeled *i* in the reference data and that were mapped *j* by Pekel et al. (2016), and the right half of the cell represents in an orange gradient the estimated percent of pixels mapped *j* by Pekel et al. (2016) that were labeled *i* in the reference data. The percent estimates were derived from the estimated area of each (*i*,*j*) cell. (Please see web version of this article for color figures.)

Pekel et al. (2016) transition map 1999-2018

Reference	Land		Permanent water		Seasonal		New permanent		New seasonal		Seasonal to permanent		Lost permanent		Lost seasonal		Permanent to seasonal		Ephemeral permanent		Ephemeral seasonal	
Land	99.1	99.3	0.0	0.6	0.8	77.3	-	-	-	-	-	-	-	-	0.0	16.7	-	-	-	-	0.0	36.8
Permanent water	6.4	0.1	88.6	93.0	1.4	2.9	0.9	24.6	0.7	12.1	1.1	52.6	0.1	3.4	-	-	0.9	34.3	-	-	0.1	1.3
Stable seasonal	43.1	0.3	6.9	1.9	24.2	12.8	-	-	7.6	31.0	-	-	0.4	3.6	10.5	52.2	1.2	12.6	-	-	6.0	35.2
Gain	28.4	0.0	6.5	0.3	23.5	2.2	21.0	27.3	15.3	11.2	5.3	11.9	-	-	-	-	-	-	-	-	-	-
Loss	5.7	0.0	-	-	11.5	0.4	-	-	-	-	-	-	65.5	41.4	17.2	6.2	-	-	-	-	-	-
Dry period	6.0	0.0	26.8	0.5	21.2	0.7	-	-	7.1	1.9	7.1	5.7	-	-	7.1	2.3	24.8	16.2	-	-	-	-
Wet period	48.0	0.0	2.4	0.1	4.7	0.4	2.4	2.8	25.6	16.9	-	-	-	-	4.9	4.0	-	-	4.7	26.4	7.1	6.8
High frequency	46.5	0.3	13.7	3.7	6.1	3.2	6.2	45.3	6.5	27.0	2.3	29.8	5.9	51.6	3.7	18.7	3.6	36.9	2.1	73.6	3.3	19.9

contrast, a probability sample-based analysis allows for estimation of area bounded by quantified uncertainties that can be appropriately used in science and policy applications (Eggleston et al., 2006; Olofsson et al., 2013; Penman et al., 2016).

Two separate sample-based accuracy assessments were conducted to better evaluate both the spatial and temporal dimensions of the surface water maps and to estimate area of surface water extent and change. The sample assessment of monthly water detection was conducted using reference data derived from 5 m RapidEye imagery. This assessment provides accuracy results for the base monthly layers, describes spatial distribution of error as a function of proximity to water-land transitions, and offers insights into the limitations of using Landsat resolution (30 m) data for mapping surface water. The time-series sample provides class accuracies of the inter-annual dynamics map, but more importantly, area estimators of surface water extent and change accompanied by standard errors to quantify uncertainty of these estimates. This analysis enables proper use of the inter-annual dynamics map as well as highlights the challenges of time-series change mapping. These two analyses extend the validation of Pekel et al. (2016) in that (1) accuracy estimates represent the entire map, (2) the reference sample data explicitly evaluate change and were used to calculate unbiased estimators of areas, and (3) standard errors were reported to provide an assessment of uncertainty of the area estimates.

Results illustrate that while pure water pixels are usually easily discriminated from land pixels, change between the two categories is very difficult to characterize in mixed pixels and mixed pixels are prevalent as 22.8% of all 30 m pixels with water present also had land present. We found that the amount of area that has fluctuating rates of open water presence far exceeds the area that has unidirectional trends of gain and loss (Table 1). Of the pixels that experienced change in surface water, 83% did not experience unidirectional change. These results demonstrate how difficult it is to map water dynamics accurately at a 30 m spatial resolution and highlights the need for time-series of higher spatial resolution data, such as Sentinel-2 (10-20 m with 5 day revisit), or commercial data such as Planet (3 m with daily revisit) to improve change area estimation. Radar data sources, such as Sentinel-1, also offer a path forward for mapping open water, with advantages in both the spatial and temporal domains compared to Landsat, particularly in rainy seasons and flood events due to radar's ability to penetrate clouds. However, since the first Sentinel-2 satellite was launched in 2015, the first of Planet in 2014, and the first of Sentinel-1 in 2014, historical analysis before these dates will still rely on Landsat.

It is unknown how much surface water is left undetected due to being under forest cover or other vegetation obscuring the surface of the water from above. As a result, it is unknown how much variation in open water presence is due to variation in vegetation cover rather than surface water presence. Much of the area of stable seasonal and multiple transitions is within natural wetlands. These are areas that fluctuate greatly in both surface water extent as well as vegetation extent and density as part of a natural cycle. There have been various efforts to quantify wetland extent at regional and global extents, but as of yet, there is no consensus on global wetland area with some estimates double the area of others (Davidson et al., 2018), reflecting the current lack of consistency among map products of wetland extent (Hu et al., 2017). While coarser resolution products of global mapped surface water inundation exist (Fluet-Chouinard et al., 2015; Papa et al., 2010), no global inundation maps exist at medium spatial and temporal resolution. Medium resolution synthetic aperture radar has been used at a regional scale to map inundation during the wet and dry seasons, including in the forest (Hess et al., 2015).

4.1. Area comparison

Having reference data for every year enabled estimation of the mean annual area with various frequencies of open water presence (Fig. 11) and the standard deviation (SD) over all years of the annual area estimates. For example, 3.13 million km^2 (SD = 0.06 million km^2) has open water ≥90% of the year, 3.69 (SD = 0.04) million km^2 has open water > 50% of the year, and 4.12 (SD = 0.05) million km^2 is has open water ≥25% of the year. These estimated areas represent the continental area within 56°S and 75°N delineated by the Global Administrative Areas dataset (GADM) (University of California, Berkley 2012) and exclude Greenland. Existing Landsat-based published studies fall within the range of annual areas with associated SDs for different rates of open water presence (Chen et al., 2015; Feng et al., 2016; Pekel et al., 2016; Yamazaki et al., 2015) (Fig. 9). All but Pekel et al. (2016) use the Global Land Survey collection, which is comprised of single Landsat scenes from isolated years, with some scenes selected from surrounding years instead, rendering analysis of seasonality impossible (Gutman et al., 2008; Tucker et al., 2004). Two other Landsat based products map sub-categories of global open water with an estimate of the global areas of lakes of 4.76 million km^2 by Verpoorter et al. (2014) and an estimate of the global area of rivers of 0.773 (± 0.079) million km^2 by Allen and Pavelsky (2018).

Pekel et al. (2016) report 2.78 million km^2 of permanent water and 0.81 million km^2 of seasonal water October 2014–October 2015, where seasonal is defined as at least one month being labeled as land and at least one month being labeled as water. Pekel et al. (2016) use an upper bound of 78°N and the coast is also delineated by GADM. Pekel et al. (2016) additionally provide two time-series aggregate areas, 4.46 million km^2 with open water at any point between 1984 and 2015 and 2.4 million km^2 with permanent water from 1984 to 2015. For 1999–2018, we estimate 4.82 (± 0.08) million km^2 to be one of any of our water classes and 2.93 (± 0.09) million km^2 to be permanent open water as defined in this study (Table 1) and 2.43 (± 0.13) million km^2 to be permanent open water if we apply the strictest definition of 100% water detections. Although the studies cover two different time intervals, this strictest estimate from our analysis (2.43 million km^2) nearly matches Pekel et al.'s (2016) permanent water estimate of 2.4 million km^2. Our estimate for the area in any of our water classes is 0.36 million km^2 larger than Pekel et al.'s (2016) area of water at any time, and if we broaden our definition to include water at any time, our estimate is 9.48 (± 1.97) million km^2 which is over twice as large as Pekel et al.'s (2016) reported area for 1984–2015. This larger estimate includes all sample units that had a mean annual open water presence percent ≤10% and inter-annual variation ≤33% and also had at least one of the observations labeled water in the reference data, which is considered land in the smaller estimate. This estimate may be much larger due to the filtering of Pekel et al. (2016), which is not reproducibly described but is intended to remove cloud shadows, as well as, the bias of omission found for the Pekel et al. (2016) monthly water layers. Since our estimates are calculated only from the reference data and the strata areas, the GLAD map bias does not contribute to the reported difference. While not enough to account for the magnitude of the difference, our analysis also includes three additional years beyond the scope of Pekel et al.'s (2016) areas.

Our study reports change areas based on the entire time-series rather than bi-temporally. Our sample-based estimates for 1999–2018 are 138,011 (± 28,163) km^2 of gained persistent water and 53,154 (± 10,883) km^2 of lost persistent water. Much of the increase is due to the creation of reservoirs (Zarfl et al., 2015) as well as climate impacts such as has caused the increase of lakes in the Tibetan plateau (Zhang et al., 2014) and in the prairie pothole region of North America (Zou et al.,2018) and much of the decrease comes from desiccation of many of the large saline lakes in the Middle East (Wurtsbaugh et al., 2017).

Given that the area that experienced multiple transitions between water and land on an annual time-scale was over 400% larger than the area with only unidirectional change (i.e. either loss or gain), it is imperative to look at the whole time-series to quantify trends of gain or loss. In contrast, previous studies only reported change areas bi-temporally from map pixel counts. Taking loss for example, Pekel et al.

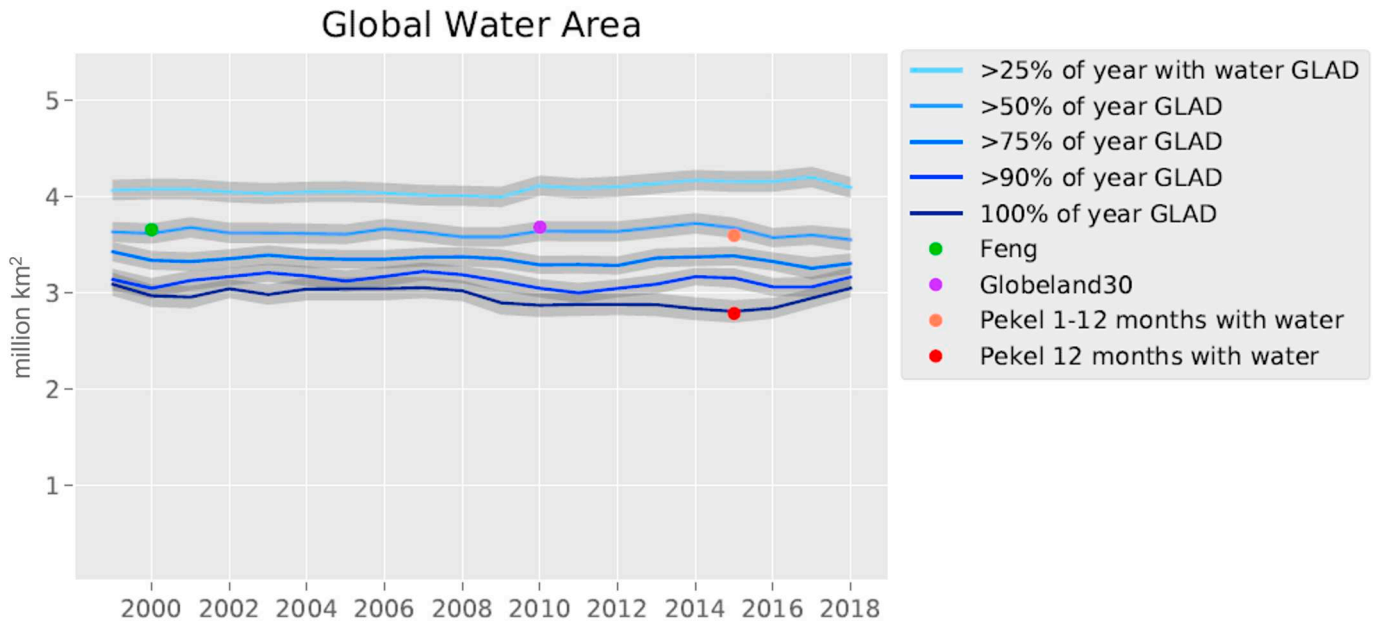


Fig. 11. Global area estimates from this study and other previously published studies based on Landsat data. Each estimate from this study is the area with open water for the given percent of the year, and the gray area bounding it corresponds to the standard error of the estimate. Confidence intervals were not provided for the other studies.

(2016) reported 162,000 km^2 of permanent water loss, 90,000 km^2 of which changed to land and 72,000 km^2 to seasonal water. These results cannot be appropriately compared to those of our study, because of differences in both their definition of loss and their time period. Pekel et al. (2016) defined loss as a transition from having all months labeled as water to having any months not labeled as water and this definition of loss was based on comparing only the first representative year (defined per pixel with a range of 1985–2000) and 2015, thereby ignoring all intervening years. The comparison of the Pekel et al. (2016) transition labels we generated for 1999–2018 versus the reference class for the time-series sample highlights this, showing that only 41% of the transition label lost permanent, defined as permanent water to land, is actually loss when the intervening years are taken into account and 52% changes back and forth between land-dominant and water-dominant. The goal of the definitions of change types used in our study is to characterize the dominant behavior of the pixel through all 20 years, but change areas are inherently dependent on the definition and the complexity of defining change types increases for a land cover that is as dynamic as open surface water.

The areas reported in our study are the first to be estimated according to good practice guidelines. The areas reported by Pekel et al. (2016) were generated from summing the mapped area, or pixel counts. Doing so carries the classification bias into the estimated area. The validation performed by Pekel et al. (2016) within their study found a bias such that the number of seasonal water detections was 21–25% smaller than detected in their reference data set, which means there will be considerable bias in the areas computed from pixel counts. This bias is present even though their omission sampling frame itself omitted many smaller water bodies. Pekel et al. (2016) had separate sampling frames for omission and commission, and their omission sampling frame did not include many smaller water bodies due to only sampling within the SRTM Water Body Data (SWBD) for $< 60°$N. The SWBD only represents lakes that are at least the equivalent of 20 $\times$ 6 Landsat pixels, and only represents rivers that have a segment that is at least that same size (*SRTM Data Editing Rules,* 2003). Bodies of water with area < 0.1 km^2, which is a threshold roughly equal to the minimum size of SWBD water bodies, contribute a large fraction of total surface area of inland waters with estimates ranging between 12% and 17% of total inland water area (Downing et al., 2006; Verpoorter et al., 2014). Smaller water bodies have a higher proportion of mixed pixels and are more likely to be transient, both of which make them harder to map. Since only pixels within the SWBD were eligible to be sampled (all other pixels had an inclusion probability of 0), the omission error estimates only apply to the area within that mask. The stratified random sample of reference data that we selected from the whole map and which targets each dynamic class yields unbiased estimators of area that are representative of the whole map, and the standard errors quantify the uncertainty of these area estimates. These uncertainty bounds inform the appropriate use of the estimates in further research as well as policy.

Geolocation error is an issue that could affect the area of change, particularly of multiple transitions. However, the impact is likely small in the tropical and temperate zones, based on visual inspection of the time-series probability sample and because the Landsat 8 OLI shift between images of the same path-row was reported by Storey et al. (2014) to be less than three meters for 90% of image pairs. There can be greater shift in regions where data are collected from path-rows in multiple UTM zones due to nearest neighbor resampling into geographic projection, causing some output pixels to receive data from different source pixel footprints. This issue is more prominent in the high latitudes because of the high proportion of zone overlap. The use of the entire Landsat archive with several steps of outlier removal mitigates, but does not remove, the risk of geolocation error impacting results.

4.2. Maps

Water is unique compared to many other land cover types because it can be highly variable, literally ebbing and flowing over time, sometimes at regular annual rates and sometimes in long-term trends. Our approach to analyzing the time-series was to model high-confidence transitions for unidirectional and oscillating change dynamics. Monthly, seasonal, and annual percent water layers were generated and the annual time-series used for deriving change categories and to calculate a three-point model of inter-annual dynamics. The resulting R-G-B inter-annual dynamics map enables the viewing of mapped changes and their intensity (Fig. 6). This map extends beyond the results of previous studies in that it characterizes eight different stable and

dynamic classes in a continuum and from all years rather than just bi-temporally based on only the start and end dates. Given that the area that experienced multiple transitions between persistent water and persistent land was over four times larger than the area with monotonic gain or loss, taking all years into account when mapping changes is critical for derivative studies that will use the water dynamic maps and/or area estimates. The R-G-B inter-annual dynamics map, the map of stable and dynamic classes, and the monthly and annual water percent maps are available for visualization and download at www.glad.umd.edu/dataset/global-surface-water-dynamics.

The validation employing 5 m RapidEye data shows that the classification algorithm performed well in discriminating pure land from pure water pixels. Example error sources included ice cover, lava, dark vegetation cover, high turbidity or very shallow water, issues of timing, and, more frequently, the combination of shadow or haze with one of these covers, as well as, mixed pixels. To investigate the impact of mixed pixels, we assessed accuracy as a function of distance from the class transition. We found that water > 30 m from the edge of the water body (i.e. the width of a Landsat pixel) had very high accuracy in both the maps produced in our study and in the study by Pekel et al. (2016). However, we must be concerned about the classification accuracy of mixed pixels given that 10.9 (± 1.9)% of inland water area is within mixed pixels. In this study, we achieved user's and producer's accuracies of the water class in the mixed pixel region of 0-30 m from the water-land boundary of 70.9 (± 1.6)% and 66.0 (± 3.7)3% for the single month layers at a scale of 5 m. While theoretically a 30 m map could have 100% accuracy at 30 m, it could never have 100% accuracy at 5 m. Thus, though mixed pixels have much lower accuracies, we correctly map two thirds of the 5 m space within 30 m of the land-water boundary. Furthermore, in our study we correctly identified 28% more of the 5x5m footprints with water cover compared to the only previous dense surface water record at 30 m (Pekel et al., 2016). This increased accuracy translates to better quantification and monitoring of small water bodies. The higher classification error rate of mixed pixels extends to the inter-annual dynamics map and users should be aware that while core change pixels should be viewed with high confidence, edge and isolated pixels are more likely to be mapped inaccurately. Additionally, the dynamic classes have much lower accuracies than the permanent land and water classes.

5. Conclusion

This study presents the first set of area estimates for global open surface water extent and change that follow good practice guidance for area reporting (Eggleston et al., 2006; Olofsson et al., 2014; Penman et al., 2016; Stehman, 2013) as well as that use all Landsat imagery for the entire monitoring period to classify seven different temporal dynamics. We establish the necessity to evaluate time-series data through the entire period given that the area that transitioned multiple times between land and water inter-annually was four times larger than the area of unidirectional loss or gain. The maps produced in this study provide detailed visualizations of inter-annual surface water dynamics using the entire Landsat archive 1999–2018 that enable assessment of changes through the past 20 years. This analysis could be extended back to 1984 for regions that were regularly imaged from 1984 to 1998 such as the United States and Australia. Map accuracy is high for permanent land, permanent water, and water change as a single theme, with considerably lower accuracies for individual water dynamics. Given that 10.9 (± 1.9)% of global water is mixed with land at a 30 m resolution and that many change dynamics occur within mixed pixels, there is a clear need for improving spatial detail in tracking surface water changes. Fortunately, newer high spatial resolution data from the Sentinel-1 and Sentinel-2 series of satellites offer a ready input for advancing open surface water monitoring. Planet data, while likely not a feasible alternative for global mapping due to its high cost, may be used as reference data in assessing map accuracies and providing area estimates for various dynamics. In addition to higher spatial resolution, incorporating Sentinel-1 and Sentinel-2 together with Landsat will provide higher temporal resolution, enabling better characterization of ephemeral surface water. The presented Landsat-based method will continue to be updated, with the current map products available through 2019, and is available for visualization and download at www.glad.umd.edu/dataset/global-surface-water-dynamics.

CRediT authorship contribution statement

Amy H. Pickens: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization.**Matthew C. Hansen:**Conceptualization, Methodology, Resources, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.**Matthew Hancher:**Software, Writing - review & editing.**Stephen V. Stehman:**Methodology, Writing - review & editing.**Alexandra Tyukavina:**Methodology, Writing - review & editing.**Peter Potapov:**Methodology, Software, Writing - review & editing.**Byron Marroquin:**Investigation.**Zainab Sherani:** Investigation, Writing - review & editing.

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.

Acknowledgments

The research was supported by World Resources Institute through Norway's International Climate and Forest Initiative grant 16123561, by NASA grant 80NSSC18M0012 for applying earth observations for the United Nations Sustainable Development Goals, by USGS-NASA Landsat Science Team grant 140G0118C0013, and by the Skoll Global Threats Fund grant 0155748.

Appendix A

A.1. Area estimation of water dynamics classes

Area was estimated for each of the eight classes of water dynamics as well as for aggregations of these classes with these reference data. For a given class, the global area W of that class is estimated by:

$$\widehat{W} = \sum_{h=1}^{H} \widehat{W}_h \tag{3}$$

where H is the number of strata. The estimated area of the class within stratum h is $\widehat{W}_h = A_h \widehat{p}_h$ where A_h is the area of stratum h and $\widehat{p}_h$ is the sample proportion of pixels of that class within stratum h. The stratum area A_h is calculated by summing the areas of all pixels within the stratum. The estimated variance of a stratum-specific area estimate is:

$$\widehat{V}(\widehat{W}_h) = s^2_{h,u}/n_h \tag{4}$$

where n_h is the number of sampled pixels in stratum h, $s_{h,u}{}^2$ is the sample variance for the n_h values of u, where $u = A_h$ if the sample pixel is of the class being estimated and $u = 0$ if the sample pixel is not of that class. The standard error for the global estimated area is the square root of the sum of the variances over all strata:

$$SE(\widehat{W}) = \sqrt{\sum_{h=1}^{H} \widehat{V}(\widehat{W}_h)}. \tag{5}$$

For area estimates of aggregations of classes, the aggregation was defined as a new class and the above calculations were performed for each. Additionally, the area that is inundated each year for > 25, 50, 75, and 90 and equal to 100% of the time was calculated from the annual percent values of each sampled pixel.

A.2. Accuracy estimation of water dynamics classes

Using the stratified sample of Landsat time-series data, we estimate user's and producer's accuracies for the trend classes. Each pixel within the global continental area had a non-zero inclusion probability, so the accuracy estimates are representative of the entire map. To estimate the accuracies for a given class c, one can estimate with Eq. (3) the four areas of intersection of map class c and non-c and reference class c and non-c within an error matrix and calculate the derived accuracies. Standard error estimates require the per pixel inclusion probabilities and pairwise inclusion probabilities, designated by π_u and π_{uv} where u and v denote pixels. The pairwise inclusion probability is the probability that pixels u and v will both be included in the sample. The estimated variance for an estimated ratio (either user's or producer's accuracy) is:

$$\widehat{V}(\widehat{R}) = \frac{1}{\widehat{Z}^2} \sum \sum \left(1 - \frac{\pi_u \pi_v}{\pi_{uv}}\right) \frac{(y_u - \widehat{R} z_u)}{\pi_u} \frac{(y_v - \widehat{R} z_v)}{\pi_v} \tag{6}$$

where the double summation is over all possible pairs of sample pixels, and where the estimated ratio is $\widehat{R} = \frac{\widehat{Y}}{\widehat{Z}}$ and $\widehat{Y} = \sum_s {}^{y_u}/_{\pi_u}$ and $\widehat{Z} = \sum_s {}^{z_u}/_{\pi_u}$ (Särndal et al., 1992). The inclusion probability for a pixel u in stratum h is $\pi_u = n_h a_u / A_h$ where a_u is the area of pixel u, n_h is the sample size from stratum h, and A_h is the total area of all pixels in stratum h. The pairwise inclusion probability between two sample pixels u and v depends on whether the two pixels are from the same stratum. If from different strata, the pairwise inclusion probability is simply the product of the inclusion probabilities, $\pi_{uv} = \pi_u \pi_v$, which means $\left(1 - \frac{\pi_u \pi_v}{\pi_{uv}}\right) = \left(1 - \frac{\pi_u \pi_v}{\pi_u \pi_v}\right) = 0$ in Eq. (5). If the two pixels are from the same stratum,

$$\pi_{h,uv} = \frac{(n_h - 1)\pi_{h,u}\pi_{h,v}}{[n_h - \pi_{h,u} - \pi_{h,v} + K]} \tag{7}$$

where $K = \sum_{U_h} \pi_u^2 / n_h$ and U_h denotes all pixels in stratum h, including those outside the sample (Hartley and Rao, 1962). Lastly, if u and v are the same pixel, $\pi_{uu} = \pi_u$. For both accuracy estimates, $\widehat{Y}$ is the estimated total area correctly mapped as class c and y_u = area of pixel u if pixel u is both mapped as class c and has reference class c, otherwise, $y_u = 0$. For user's accuracy, $\widehat{Z}$ is the estimated area mapped as class c and z_u = area of pixel u if pixel u is mapped as class c, otherwise $z_u = 0$. For producer's accuracy, $\widehat{Z}$ is the estimated area of reference class c and z_u = area of pixel u if pixel u has reference class c, otherwise $z_u = 0$. For each estimated accuracy $\widehat{R}$, $SE(\widehat{R}) = \sqrt{\widehat{V}(\widehat{R})}$.

References

Allen, G.H., Pavelsky, T.M., 2018. Global extent of rivers and streams. Science, eaat0636. https://doi.org/10.1126/science.aat0636.

Arvidson, T., Gasch, J., Goward, S.N., 2001. Landsat 7's long-term acquisition plan — an innovative approach to building a global imagery archive. Remote Sensing of Environment, Landsat 7 (78), 13–26. https://doi.org/10.1016/S0034-4257(01)00263-2.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140. https://doi.org/10.1007/BF00058655.

Breiman, L., Friedman, J., Ohlsen, R., Stone, C., 1984. Classification and Regression Trees. Wadsworth International Group, Monterey, California.

Brewer, K.R.W., Hanif, M., 1982. Sampling With Unequal Probabilities. Springer, New York.

Cai, X., Haile, A.T., Magidi, J., Mapedza, E., Nhamo, L., 2017. Living with floods – household perception and satellite observations in the Barotse floodplain, Zambia. Physics and Chemistry of the Earth, Parts A/B/C, Infrastructural Planning for Water Security in Eastern and Southern Africa 100, 278–286. https://doi.org/10.1016/j.pce.2016.10.011.

Chander, G., Markham, B.L., Helder, D.L., 2009. Summary of Current Radiometric Calibration Coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI Sensors.

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

Chevallier, P., Pouyaud, B., Suarez, W., Condom, T., 2011. Climate change threats to environment in the tropical Andes: glaciers and water resources. Reg. Environ. Chang. 11, 179–187. https://doi.org/10.1007/s10113-010-0177-6.

Coleman, J.M., 1969. Brahmaputra river: channel processes and sedimentation. Sediment. Geol. 3, 129–239. https://doi.org/10.1016/0037-0738(69)90010-4. Brahmaputra river: Channel processes and sedimentation.

Davidson, N.C., 2014. How much wetland has the world lost? Long-term and recent trends in global wetland area. Mar. Freshw. Res. 65, 934–941. https://doi.org/10.1071/MF14173.

Davidson, N.C., Fluet-Chouinard, E., Finlayson, C.M., 2018. Global extent and distribution of wetlands: trends and issues. Mar. Freshw. Res. 69, 620–627. https://doi.org/10.1071/MF17019.

Downing, J.A., Prairie, Y.T., Cole, J.J., Duarte, C.M., Tranvik, L.J., Striegl, R.G., McDowell, W.H., Kortelainen, P., Caraco, N.F., Melack, J.M., Middelburg, J.J., 2006. The global abundance and size distribution of lakes, ponds, and impoundments. Limnol. Oceanogr. 51, 2388–2397. https://doi.org/10.4319/lo.2006.51.5.2388.

Eggleston, H.S., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., 2006. 2006 IPCC Guidelines for National Greenhouse Gas Inventories.

Feng, M., Sexton, J.O., Channan, S., Townshend, J.R., 2016. A global, high-resolution (30-m) inland water body dataset for 2000: first results of a topographic–spectral classification algorithm. International Journal of Digital Earth 9, 113–133. https://doi.org/10.1080/17538947.2015.1026420.

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

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, Big Remotely Sensed Data: tools, applications and experiences 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Grill, G., Lehner, B., Lumsdon, A.E., MacDonald, G.K., Zarfl, C., Liermann, C.R., 2015. An index-based framework for assessing patterns and trends in river fragmentation and flow regulation by global dams at multiple scales. Environ. Res. Lett. 10, 015001. https://doi.org/10.1088/1748-9326/10/1/015001.

Gutman, G., Byrnes, R.A., Masek, J., Covington, S., Justice, C., Franks, S., Headley, R., 2008. Towards monitoring land-cover and land-use changes at a global scale: the global land survey 2005. Photogramm. Eng. Remote. Sens. 74, 5.

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

science.1244693.
Hartley, H.O., Rao, J.N.K., 1962. Sampling with unequal probability and without replacement. Ann. Math. Stat. 33, 350–374.
Hess, L.L., Melack, J.M., Affonso, A.G., Barbosa, C., Gastil-Buhl, M., Novo, E.M.L.M., 2015. Wetlands of the lowland Amazon Basin: extent, vegetative cover, and dual-season inundated area as mapped with JERS-1 synthetic aperture radar. Wetlands 35, 745–756. https://doi.org/10.1007/s13157-015-0666-y.
Hu, S., Niu, Z., Chen, Y., 2017. Global wetland datasets: a review. Wetlands 37, 807–817. https://doi.org/10.1007/s13157-017-0927-z.
Lutz, A.F., Immerzeel, W.W., Shrestha, A.B., Bierkens, M.F.P., 2014. Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. Nature Clim. Change 4, 587–592. https://doi.org/10.1038/nclimate2237.
McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432. https://doi.org/10.1080/01431169608948714.
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 25years of Landsat imagery across Australia. Remote Sens. Environ. 174, 341–352. https://doi.org/10.1016/j.rse.2015.11.003.
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.
Papa, F., Prigent, C., Aires, F., Jimenez, C., Rossow, W.B., Matthews, E., 2010. Interannual variability of surface water extent at the global scale, 1993–2004. Journal of Geophysical Research: Atmospheres 115. https://doi.org/10.1029/2009JD012674.
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.
Penman, J., Green, C., Olofsson, P., Raison, J., Woodcock, C., Balzter, H., Baltuck, M., Foody, G.M., 2016. Integration of Remote-sensing and Ground-based Observations for Estimation of Emissions and Removals of Greenhouse Gases in Forests: Methods and Guidance From the Global Forest Observations Initiative.
Potapov, P.V., Turubanova, S.A., Tyukavina, A., Krylov, A.M., McCarty, J.L., Radeloff, V.C., Hansen, M.C., 2015. Eastern Europe's forest cover dynamics from 1985 to 2012 quantified from the full Landsat archive. Remote Sens. Environ. 159, 28–43. https://doi.org/10.1016/j.rse.2014.11.027.
Prigent, C., Papa, F., Aires, F., Jimenez, C., Rossow, W.B., Matthews, E., 2012. Changes in land surface water dynamics since the 1990s and relation to population pressure. Geophys. Res. Lett. 39. https://doi.org/10.1029/2012GL051276.
Särndal, C.E., Swensson, B., Wretman, J., 1992. Model-Assisted Survey Sampling. Springer-Verlag, New York.
SRTM, 2017. 90m Digital Elevation Database v4.1. CGIAR-CSI URL. https://cgiarcsi.community/data/srtm-90m-digital-elevation-database-v4-1/, Accessed date: 5 April 2019.
SRTM Data Editing Rules. USGS URL. https://dds.cr.usgs.gov/srtm/version2_1/Documentation/SRTM_edit_rules.pdf, Accessed date: 5 April 2019.
Stehman, S.V., 2013. Estimating area from an accuracy assessment error matrix. Remote Sens. Environ. 132, 202–211. https://doi.org/10.1016/j.rse.2013.01.016.
Storey, J., Choate, M., Lee, K., 2014. Landsat 8 Operational Land Imager On-Orbit Geometric Calibration and Performance. Remote Sens. 6, 11127–11152. https://doi.org/10.3390/rs61111127.
Tessler, Z.D., Vörösmarty, C.J., Grossberg, M., Gladkova, I., Aizenman, H., 2016. A global empirical typology of anthropogenic drivers of environmental change in deltas. Sustain. Sci. 11, 525–537. https://doi.org/10.1007/s11625-016-0357-5.
Tockner, K., Stanford, J.A., 2002. Riverine flood plains: present state and future trends. Environ. Conserv. 29, 308–330. https://doi.org/10.1017/S037689290200022X.
Tucker, C.J., Grant, D.M., Dykstra, J.D., 2004. NASA's Global Orthorectified Landsat Data Set. https://doi.org/10.14358/pers.70.3.313.
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 sub-continental scale in a semi-arid region. Remote Sens. Environ. 178, 142–157. https://doi.org/10.1016/j.rse.2016.02.034.
USGS Open-File Report 2011–1073: Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) [WWW Document], n.d. URL https://pubs.usgs.gov/of/2011/1073/ (accessed 4.5.19).
Verpoorter, C., Kutser, T., Seekell, D.A., Tranvik, L.J., 2014. A global inventory of lakes based on high-resolution satellite imagery. Geophys. Res. Lett. 41, 6396–6402. https://doi.org/10.1002/2014GL060641.
Wulder, M.A., White, J.C., Loveland, T.R., Woodcock, C.E., Belward, A.S., Cohen, W.B., Fosnight, E.A., Shaw, J., Masek, J.G., Roy, D.P., 2016. The global Landsat archive: status, consolidation, and direction. Remote Sens. Environ. 185, 271–283. https://doi.org/10.1016/j.rse.2015.11.032. Landsat 8 Science Results.
Wurtsbaugh, W.A., Miller, C., Null, S.E., DeRose, R.J., Wilcock, P., Hahnenberger, M., Howe, F., Moore, J., 2017. Decline of the world's saline lakes. Nat. Geosci. 10, 816. https://doi.org/10.1038/ngeo3052.
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.
Yamazaki, D., Trigg, M.A., Ikeshima, D., 2015. Development of a global ~90m water body map using multi-temporal Landsat images. Remote Sens. Environ. 171, 337–351. https://doi.org/10.1016/j.rse.2015.10.014.
Zarfl, C., Lumsdon, A.E., Berlekamp, J., Tydecks, L., Tockner, K., 2015. A global boom in hydropower dam construction. Aquat. Sci. 77, 161–170. https://doi.org/10.1007/s00027-014-0377-0.
Zhang, G., Yao, T., Xie, H., Zhang, K., Zhu, F., 2014. Lakes' state and abundance across the Tibetan Plateau. Chin. Sci. Bull. 59, 3010–3021. https://doi.org/10.1007/s11434-014-0258-x.
Zhao, X., Huang, Y., Jia, Z., Liu, H., Song, T., Wang, Yuesi, Shi, L., Song, C., Wang, Yiyong, 2008. Effects of the conversion of marshland to cropland on water and energy exchanges in northeastern China. J. Hydrol. 355, 181–191. https://doi.org/10.1016/j.jhydrol.2008.03.019.
Zou, Z., Xiao, X., Dong, J., Qin, Y., Doughty, R.B., 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. PNAS, 201719275. https://doi.org/10.1073/pnas.1719275115.