# Surface water monitoring in small water bodies: potential and limits of multi-sensor Landsat time series

Andrew Ogilvie[1,2], Gilles Belaud[1], Sylvain Massuel[1], Mark Mulligan[2], Patrick Le Goulven[1], and Roger Calvez[1]

[1]G-EAU, AgroParisTech, Cirad, IRD, IRSTEA, Montpellier SupAgro, Univ Montpellier, Montpellier, France
[2]Department of Geography, King's College London, WC2R 2LS London, UK

*Correspondence to:* Andrew Ogilvie (andrew.ogilvie@ird.fr)

**Abstract.** Hydrometric monitoring of small water bodies (1-10 ha) remains rare, due to their limited size and large numbers, preventing accurate assessments of their agricultural potential or their cumulative influence in watershed hydrology. Landsat imagery has shown its potential to support mapping of small water bodies but the influence of their limited surface areas, vegetation growth and rapid flood dynamics on long term surface water monitoring remains unquantified. A semi-automated

5    method is developed here to assess and optimise the potential of multi-sensor Landsat time series to monitor surface water extent and mean water availability in these smallest water bodies. Extensive hydrometric field data (1999-2014) for 7 small reservoirs within the Merguellil catchment in Central Tunisia are used to calibrate the method and explore its limits. MNDWI is shown out of six commonly used water detection indices to provide high overall accuracy and threshold stability during high and low floods, leading to a mean surface area error below 15%. Applied to 546 Landsat 5, 7 and 8 images over 1999-

10    2014, the method reproduces surface water extent variations across small lakes with high skill ($R^2 = 0.9$) and mean RMSE of $9\,300\,\mathrm{m}^2$. Comparison with published global water data sets reveals a mean RMSE of 21 800 $\mathrm{m}^2$ (+ 134%) on the same lakes and highlights the value of a tailored MNDWI approach to improve hydrological monitoring in small lakes and reduce omission errors of flooded vegetation. The rise in relative errors due to the larger proportion and influence of mixed pixels restricts surface water monitoring below 3 ha with Landsat (NRMSE = 27%). Interferences from clouds & scan line corrector

15    failure on ETM+ after 2003 also decrease the number of operational images by 51%, reducing performance on lakes with rapid flood declines.

# 1 Introduction

## 1.1 Monitoring water resources in small reservoirs

Support from governmental and international projects combined with farmer initiatives have led to significant development of

20    small reservoirs (< 1 $\mathrm{Mm}^3$) across the globe, including India (Bouma et al., 2011; Venot and Krishnan, 2011; Massuel et al., 2014), Brazil (Burte et al., 2005) and parts of Northern and sub-Saharan Africa (Nyssen et al., 2010; Sawunyama et al., 2006; Talineau et al., 1994). These have been built to reduce soil loss and silting of downstream dams, as well as harvest scarce and unreliable water resources for local users (Habi and Morsli, 2011; Wisser et al., 2010).

Despite their growing importance worldwide, small reservoirs are rarely monitored in situ, except for research purposes (Albergel and Rejeb, 1997) due to their quantity, size and geographical dispersion. Instrumenting and maintaining a hydrometric observation network requires significant time, equipment and transport costs (Liebe et al., 2005) which are not compatible with the localised and modest importance in the water budget of small reservoirs. As a result, little information is available on their

5 water availability and their cumulative influence (Ogilvie et al., 2016b).

Water availability in small reservoirs is constrained by design capacities but determined by site-specific hydrological dynamics regulated by both natural (climate, topography, geology, geomorphology, pedology), and human factors (maintenance, leaks, withdrawals, releases). Studies based on hydrological measurements and geochemical analyses have identified the water balance of small reservoirs (Gay, 2004; Grunberger et al., 2004; Lacombe, 2007; Li and Gowing, 2005; Nyssen et al., 2010)

10 but highlighted multiple modelling uncertainties due to evaporation, groundwater flows, high spatial rainfall variability and human management (Li and Gowing, 2005; Grunberger et al., 2004; Lacombe, 2007; Leduc et al., 2007; Kingumbi et al., 2007). Rainfall-runoff modelling of gauged small reservoirs in Tunisia notably failed to exceed $R^2 = 0.5$ (Lacombe, 2007) and difficulties are greater in ungauged catchments (Cudennec et al., 2005).

## 1.2 Remote sensing in hydrology

15 Images from a variety of active and passive sensors have been used successfully in hydrology to inventory, to assess flooded areas, river stage levels and widths, as well as to investigate hydrological processes and water balance issues. Low spatial resolution sensors such as MODIS (Moderate Resolution Imaging Spectroradiometer, 250 m) and AVHRR (Advanced Very High Resolution Radiometer, 1.1 km) providing daily coverage of the globe, notably facilitated the assessment and monitoring of large wetlands (Guo et al., 2017) including the Niger Inner Delta (Mahé et al., 2011; Seiler et al., 2009; Ogilvie et al.,

20 2015; Bergé-Nguyen and Crétaux, 2015), the Okavango (Wolski and Murray-Hudson, 2008; Gumbricht et al., 2004), the Tana Delta (Leauthaud et al., 2013) or the Mekong (Kuenzer et al., 2015; Sakamoto et al., 2007), large rivers such as the Amazon (Martinez and Le Toan, 2007; Alsdorf et al., 2007) and large lakes notably in East Africa (Swenson and Wahr, 2009; Ouma and Tateishi, 2006) and China (Ma et al., 2007; Qi et al., 2009). These sensors have also been used for global assessments (Prigent et al., 2007; Papa et al., 2010; Klein et al., 2015) but their low spatial resolutions remain inadequate for small reservoirs, as a

25 single MODIS pixel corresponds to an area of 6.25 ha.

High spatial resolution imagery such as SPOT (Satellite Pour l'Observation de la Terre), Quickbird and Ikonos capture images under 5 m resolution but their reduced spatial coverage (10-20 km wide for the highest resolution) make them less suitable for medium-sized catchments and their costs are prohibitive for operational long-term monitoring. Recent sensors such as Sentinel-2 capture 10 m images of the entire globe every 5 days, providing enhanced opportunities but hydrological

30 investigations which require historical perspectives will continue to rely on previous sensors. Landsat which provides free multispectral images every 16 days since the 1970s at medium geometric resolution (30 m since Landsat 4 in 1982) therefore continues to provide the most potential to detect and monitor small water bodies.

### 1.3   Landsat imagery to map and monitor small water bodies

Smaller water bodies have been included in worldwide and local inventories, done with Landsat imagery (Feng et al., 2016; Verpoorter et al., 2014; Yamazaki et al., 2015). Local studies on small lakes from 1 ha to 25 ha in West Africa (Gardelle et al., 2010; Liebe et al., 2005; Annor et al., 2009; Soti et al., 2010; Jones et al., 2017), in Southern Africa (Sawunyama et al., 2006),
5 in South India (Mialhe et al., 2008), and in South America (Rodrigues et al., 2011) specifically investigated the potential of remote sensing methods to assess surface areas of small reservoirs. Field validation confirmed the potential of remote sensing (RS) methods as a cost-effective technique to identify water bodies but revealed problems of variable accuracy (Ran and Lu, 2012; Annor et al., 2009; Solander et al., 2016) on small lakes (under 100 ha). Importantly though small water bodies were detected using satellite imagery, the feasibility of monitoring surface water dynamics and providing long term water availability
10 assessments was not investigated (Baup et al., 2014; Pekel et al., 2016). Studies provided a snapshot of floods in reservoirs at a given time (Liebe et al., 2005; Feng et al., 2016), or of seasonality using two dates (Annor et al., 2009) but did not explore their dynamics over time.

Landsat has recently been used to undertake long term monitoring of surface water bodies at global (Pekel et al., 2016) and regional scales (Mueller et al., 2016; Tulbure et al., 2016; Avisse et al., 2017). The remarkable study by Pekel et al. (2016)
15 exploited 1 petabyte of Landsat archive imagery to study surface water dynamics in 30 m pixels globally over 1984-2015. Small water bodies are included, but Yamazaki and Trigg (2016) warns of limitations in Pekel et al. (2016) as "resolution issues prevent analysis of small water bodies". Similarly Mueller et al. (2016) focussing on Australian water bodies recognised the difficulties in the "presence of water and vegetation within the pixel" and that their "product may not be fit for [...] small farm dams". In their own inventory of water bodies, Yamazaki et al. (2015) had also noted the omission errors on water bodies
20 below 100 ha. These errors are influenced by the low spatial resolution but essentially by the increased presence of flooded vegetation and shallow waters which affect the reflectance signal (Sawunyama et al., 2006; Annor et al., 2009; Mueller et al., 2016; Yamazaki et al., 2015). These studies recognise the difficulties of using Landsat imagery, but adequate ground truth data remains sparse and extensive field validation has not quantified the limits of these approaches on the smallest lakes. Validation depends on higher resolution imagery or other published water maps which suffer from similar classification inaccuracies
25 (Yamazaki et al., 2015), non negligible on these small water bodies. Pekel et al. (2016) for example in their validation used sample pixels of seasonal water bodies within 1° tiles but these did not specifically account for small water bodies as shown in this paper.

Furthermore, hydrological monitoring of small water bodies imposes specific constraints on temporal resolution, due to the limited flood amplitude and the rapid flood declines, and "with the 16-day repeat cycle of Landsat [...] floods may be
30 missed" (Yamazaki and Trigg, 2016). Clouds and shadows in optical imagery as well as the scan line corrector failure (after 31.05.2003 on Landsat 7 ETM+ sensor) are known to have a detrimental influence in global water data sets (Pekel et al., 2016; Mueller et al., 2016). These have a proportionally larger influence on small water bodies where gap-fill methods are not suited to localised rapid changes in land cover (Zeng et al., 2013; Zhu and Woodcock, 2014). The influence of Landsat's limited temporal resolution and these interferences on representing flood dynamics in small lakes remains however unknown.

### 1.4 Improving surface water monitoring in small reservoirs

The objective of this research was to assess and optimise the ability of Landsat to provide sufficient, accurate observations of long term surface water dynamics in the smallest water bodies (1-10 ha). Exceptional long term hydrological time series of 7 small reservoirs over up to 15 years are used here to develop and validate a semi-automated method using Landsat imagery,

5   capable of reproducing surface water dynamics and long term water availability in small water bodies.

Numerous spectral water indices and classification methods exist but these are often tested and developed empirically on larger water bodies (Feyisa et al., 2014) and their suitability in different settings must be investigated (Palmer et al., 2015). Suitable automation of water detection over several reservoirs and over several images also imposes specific constraints on the methodology, in terms of detection accuracy, threshold stability, cloud presence, and computer processing when manipulating

10  satellite imagery (Mialhe et al., 2008; Ogilvie et al., 2015; Pekel et al., 2016). Widely used water detection indices were compared to assess their suitability to monitor floods on small reservoirs. The interferences from clouds, shadows and scan line corrector failure were investigated and reduced here by defining optimal filter thresholds based on long term field data. Results were compared with analysis performed using available global water data sets (Pekel et al., 2016) to quantify accuracy issues and argue for the benefits of specific approaches for small water bodies.

15  ## 2 Methods

### 2.1 Small reservoirs in the Merguellil upper catchment

The case study area is the Merguellil upper catchment located in central semi-arid Tunisia (Fig. 1). A total of 58 small reservoirs were built across this $1180 \ \mathrm{km^2}$ catchment since the 1960s, as a result of international projects and an ambitious nationwide water and soil conservation strategy. These were inventoried through the combination and cross referencing of records from

20  local authorities, literature (Kingumbi, 2006; Lacombe, 2007; CNEA, 2006), satellite imagery and field visits (Ogilvie et al., 2016a).

Initial design capacities of small reservoirs in the catchment range between $17\,000 \ \mathrm{m^3}$ and $1\,590\,000 \ \mathrm{m^3}$ though the median size only reaches $66\,000 \ \mathrm{m^3}$. Peak flooded surface areas vary between 0.5 ha and 18 ha. Lakes flood as a result of intense localised rainfall events (Ogilvie et al., 2016b), during autumn and spring mostly. High annual and interannual variability in

25  rainfall ($329 \ \mathrm{mm/year} \pm 131 \ \mathrm{mm}$) combined with significant heterogeneity in the infiltration, evaporation rates and occasional withdrawals lead to large disparities in the frequency and duration of floods in small lakes (Ogilvie, 2015; Lacombe, 2007).

Due to the importance of the downstream irrigated Kairouan plain for national agricultural production, the catchment has been the focus of numerous research projects to understand the changes and influences exerted by evolving climatic and human dynamics (dam and small reservoir construction, irrigation withdrawals, etc.). Small reservoirs in the Merguellil upper

30  catchment and along the Tunisian NE-SW mountain range ("Dorsale") consequently benefited from extensive hydrological monitoring thanks to ongoing research collaboration with the local authorities (Albergel and Rejeb, 1997).
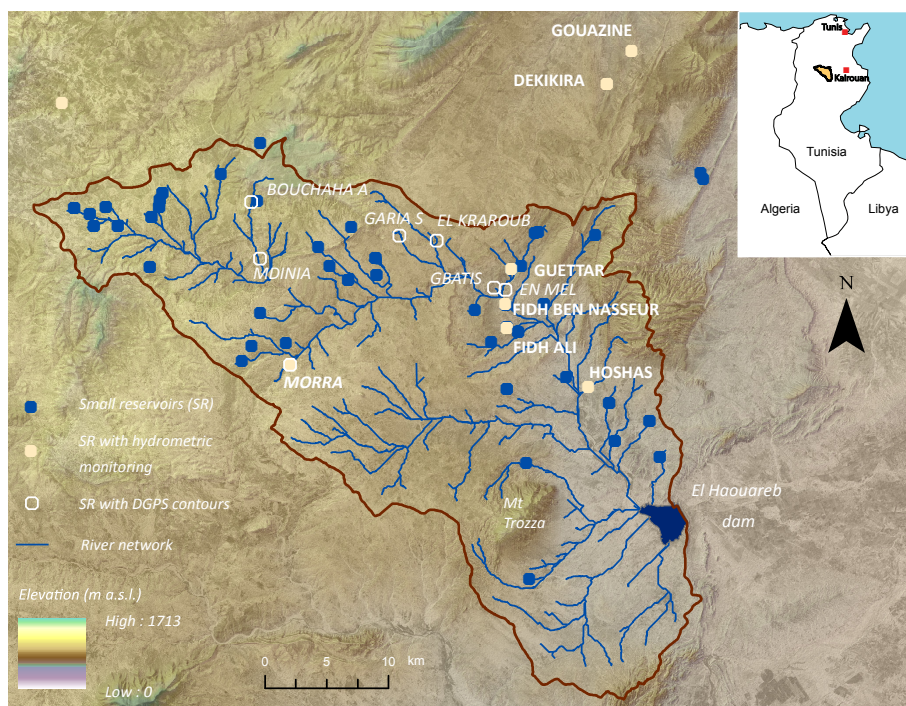
**Figure 1.** Merguellil upper catchment and location of instrumented small reservoirs

## 2.2 Field hydrometric monitoring

Seven reservoirs helped calibrate and validate the spectral water indices. Lakes were chosen amongst those identified as flooded on a classified 10 m SPOT image in spring 2013, and flooded surface areas were measured with real time kinematic (RTK) GPS, to reduce uncertainties from stage-height rating curves. These lakes provided a range of surface areas from 0.2 ha to 7.8

5   ha (Table 2) to test the performance of indices.

Seven small reservoirs with extensive field data were then used to test the performance of the method over time, i.e. identify how spatial and temporal resolution affected monitoring of surface water extent dynamics and water availability. Lakes were selected based on the length and quality of their hydrometric time series and include a range of altitudes (Fig. 1), capacities (Table 3) and associated flood dynamics (amplitude and duration). With the exception of lake Morra, these 7 lakes were not

10   used in the calibration of the indices, giving further weight and confidence to the validation of their thresholds.

Reservoir levels were monitored with automatic stage pressure transducers and daily readings by local observers on limni-metric stage ladders were used to cross check and fill gaps in automatic readings. Transducer uncertainties are minimal, with a 0.2 cm resolution and theoretical ±0.25% precision. Stage values were converted to water volumes using rating curves available for each lake. These had been acquired and updated to account for silting as part of regular surveys (Albergel and

15   Rejeb, 1997). Two additional surveys were undertaken here with a RTK GPS and interpolated based on Delaunay Triangular Interpolation Networks (TIN), well suited to bathymetric applications.
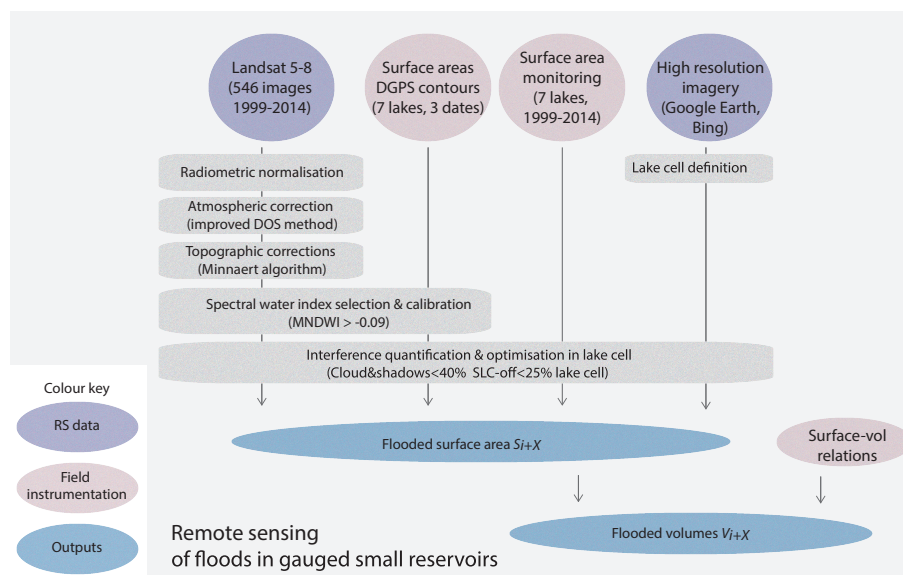
**Figure 2.** Chain of treatments applied to the 546 Landsat (TM, ETM+ and OLI) images

Lakes cells to automate extraction of the flooded area (Fig. 3) were defined using freely available high resolution imagery from Digital Globe and Astrium available in Google Earth and Bing Maps. These cells defined the maximal flooded area based on visual interpretation of geomorphological boundaries of the lake (with a 10% buffer) and helped constrain overestimations, i.e. pixels which do not belong to the lake but share similar reflectance profile, such as surrounding flooded vegetation, river
5  stretches and temporary ponds or waterlogging.

## 2.3   Landsat preprocessing

Landsat Level 1 Terrain Corrected (L1T) (orthorectified) products from USGS Earth Explorer were used here. Level 2A products corrected to land surface reflectance using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) or MACCS treatment chain were not available for Landsat 8 or before 2002 for our catchment at the time of research. The required
10  radiometric (including atmospheric & topographic) corrections were therefore derived for each sensor and band combination. These are detailed in the Appendix. Based on data availability of Landsat images and field limnimetric data for the seven lakes, 546 Landsat 5-8 images over the period 1999-2014 were acquired and treated (Fig. 2).

## 2.4   Spectral water indices for multi-image analysis of small water bodies

Six widely used water detection indices (Table 1) were compared to assess their suitability to monitor floods on small reservoirs
15  in terms of detection accuracy and threshold stability. NDPI is effectively the opposite of MNDWI and was therefore not investigated separately. NDVI was also considered as negative values of the index have been used to detect water bodies (Ma et al., 2007; Mohamed et al., 2004). In this rural, gently sloping watershed, the $AWEI_{nsh}$ variant was used, as $AWEI_{sh}$ is

**Table 1.** Spectral water indices compared in this paper

| | |
|---|---|
| Normalised Difference Water Index (McFeeters, 1996) | $NDWI = \frac{Green - NIR}{Green + NIR}$ |
| Modified NDWI (Xu, 2006) | $MNDWI = \frac{Green - SWIR}{Green + SWIR}$ |
| Normalised Difference Pond Index (Lacaux et al., 2007) | $NDPI = \frac{SWIR - Green}{SWIR + Green}$ |
| Normalised Difference Moisture Index (Gao, 1996) | $NDMI = \frac{NIR - SWIR}{NIR + SWIR}$ |
| Normalised Difference Turbidity Index (Lacaux et al., 2007) | $NDTI = \frac{Red - Green}{Red + Green}$ |
| Normalised Difference Vegetation Index (Rouse et al., 1973) | $NDVI = \frac{NIR - Red}{NIR + Red}$ |
| Automated Water Extraction Index (Feyisa et al., 2014) | $AWEI_{nsh} = 4 \times (Green - SWIR_1)$ $-(0.25 \times NIR_4 + 2.75 \times SWIR_2)$ |

optimised to remove urban and topographic shadow pixels. For NVDI and NDTI, values below the threshold were classified as water, and conversely for the four other indices. For SWIR, band 5 for TM/ETM+ and band 6 for OLI were used in accordance with other studies (Ji et al., 2009; Ouma and Tateishi, 2006).

## 2.5 Calibration and comparison of spectral water indices

The six water detection indices were calibrated on two Landsat 8 images to assess the performance and stability of each index in high waters (March 2013), and in low waters (June 2013) when greater proportions of vegetation growth, algae, turbidity can affect water detection. A third image from May 2013 when vegetation and algae growth was evident but in reduced proportions was used to validate thresholds. Timed to coincide with the 16-day Landsat 8 acquisition cycle (Table 2), the lag between satellite overpass and field observation was inferior to 1 week, similar or inferior to previous studies (Liebe et al., 2005; Annor et al., 2009; Sawunyama et al., 2006).

Georeferenced GPS contours for each lake and date (Fig. 3 and Table 2) were used to calculate confusion matrices for each index, lake and Landsat image. Classification thresholds for each index were varied in increments of 0.01, considering the observed influence of a second decimal in the thresholds (Ogilvie et al., 2015), and optimal thresholds were determined based on minimising overall accuracy rates. The median threshold from all calibration points (i.e. 7 lakes on 2 images) was retained as the optimal threshold for each index. The median was preferred over mean threshold to reduce the influence of single lakes (and dates) where the optimal threshold is significantly different due to mixed reflectance issues. Similarly, an optimal inter-lake and inter-date threshold was not determined based on aggregated pixels from all lakes, as we sought to give equal weighting to each lake rather than allowing large lakes to steer threshold values. This also ensured mixed pixels found across small reservoirs were better accounted for in the calibration (Jain, 2010). Spectral unmixing techniques (Van Der Meer,
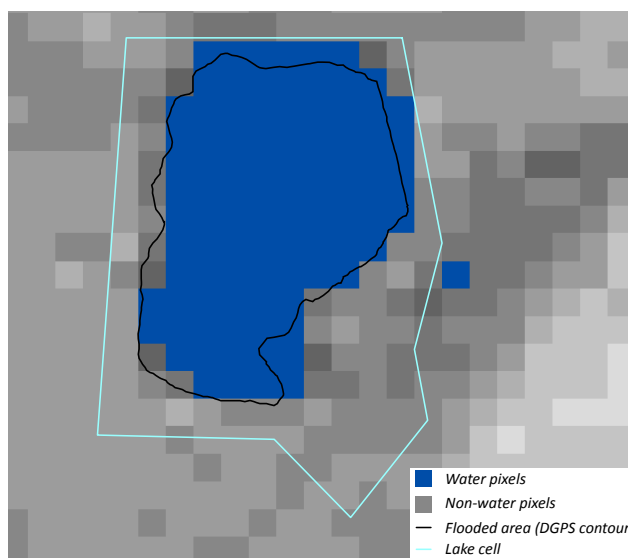
**Figure 3.** MNDWI classification on Landsat 8 29.03.2013 image for lake Morra

1995) were not used here considering the marginal benefit (5% increase in sub-pixel accuracy) observed in Sun et al. (2017) and the requirements to adapt the approach to deal with residual clouds and SLC gaps in multi-temporal Landsat time series.

$$PDAI = \frac{|S_{RS} - S_{GPS}|}{S_{GPS}} * 100 \qquad (1)$$

The Percentage Deviation Area Index (PDAI, Eq. (1)) (Sawunyama et al., 2006) was calculated to provide a relative measure
5  of the spectral indices' accuracy in estimating flooded surface area remotely ($S_{RS}$) compared to the field GPS measurements ($S_{GPS}$). PDAI was not used for calibration as it is not a pixel-based assessment and was here shown to disguise spatial errors, i.e. over-classified pixels compensating for under-classified areas. Feyisa et al. (2014) showed that additional statistics notably the Kappa coefficient and McNemar's test which focus on the frequency of pixels correctly and incorrectly classified did not provide added benefit when optimising water detection thresholds (Fisher and Danaher, 2013). Confusion matrices and PDAI
10  errors per lake on the third validation image were then used to discuss the performance of each index and determine the optimal water detection index for these small water bodies.

### 2.6 Influence of clouds, shadows and ETM+ scan-line corrector failure

Gap-fill methods designed to address the loss of pixels from SLC failure and interferences from clouds & shadows rely on nearby spatial or temporal spectral information. Suited for larger lakes and land use studies, spatial interpolation algorithms
15  and decision tree classifications (Maxwell, 2004; Bédard et al., 2008) assume that nearby pixels share similar reflectance profile which is not compatible with small scale analysis developed here (Zeng et al., 2013; Feng et al., 2016). Our objects of interest vary between 1 ha and 30 ha and are easily contained within a single SLC-off band, or individual clouds. SLC failure led to

the loss of around 22% of pixels (Zeng et al., 2013), distributed along oblique lines ranging from 14 pixels wide at the edges to 2 at the centre (i.e. $420\,\mathrm{m}$ to $60\,\mathrm{m}$) (Maxwell, 2004).

Temporal methods (Zeng et al., 2013; Zhu and Woodcock, 2014, 2012; Scaramuzza et al., 2004) as those employed by the USGS to provide SLC corrected images exploit additional spectral data from another image close in time. These rely on the
5   relative stability of pixel reflectance between images of nearby dates or at least assume "similar patterns of spectral differences between dates" (Chen et al., 2011; Zeng et al., 2013). These are again not suited to the detection of ephemeral localised changes (Zhu and Woodcock, 2014; Yin et al., 2016), such as small scale, rapid changes in water surface area, which do not follow the same gradual pattern as nearby land use (vegetation) between successive images.

The Fmask algorithm (Zhu and Woodcock, 2012) was used here to quantify the presence of clouds and shadows over each
10   lake on each Landsat image (details in Appendix). Errors from clouds, shadows and SLC-off pixels were then assessed and minimised in post-processing by removing images for each lake with interferences above a determined threshold. This optimal threshold was determined by minimising the root mean squared error (RMSE) on surface area aggregated over time (15 years) in order to give importance to both the number and quality of the Landsat observations over time, and therefore compromise between maintaining high temporal resolution and minimising spatial errors from clouds, shadows and SLC-off interferences.
15   Thresholds were varied in 5% increments and the quality of fit in terms of $R^2$ and RMSE between remotely sensed surface area and field surface area was calculated for decreasing subsets of images. Gouazine lake which possessed both the longest time series (over 15 years) and the most accurate data and rating curves (updated 6 times) was used to optimise the thresholds. The approach was repeated on four other lakes (Morra, Dekikira, Guettar, Fidh Ali) to confirm the suitability of the thresholds and explore the resulting availability of suitable Landsat imagery over all lakes.

20   ## 2.7   Remote monitoring of surface water dynamics and water availability

The complete method was applied to all 546 images and surface water extents were assessed for the 7 small reservoirs over 1999-2014. Values affected by excessive cloud & shadow & SLC presence over each lake were removed and remaining observations were linear interpolated over time into continuous time series of flood dynamics. Alternate smoothing and interpolation of discrete observations were shown not to improve the results (Ogilvie, 2015), considering the abrupt flood dynamics in small
25   water bodies. These were then converted to volumes using the site-specific updated surface-volume rating curves. The mean water availability per year was calculated to assess the suitability of the approach to derive long term water availability data for stakeholders. Outliers were further constrained by the known initial maximum capacity of lakes. The skill of the method in reproducing surface water extent dynamics and water availability assessments over the seven reservoirs were discussed in light of $R^2$ and RMSE values against the extensive field based data.

30   ## 2.8   Comparison with published global water data sets

Results were compared to those obtained when using the global surface water data sets produced by Pekel et al. (2016) with $30\,\mathrm{m}$ Landsat imagery. These exploited original approaches based on a combination of expert systems, visual analytics and evidential reasoning. 200 monthly water history rasters produced by the European Commission Joint Research Centre (JRC)

for our region of interest were downloaded through Google Earth Engines. Monthly surface water extents areas over 1999-2014 were extracted from the raster files for the 13 lakes used previously. Confusion matrices between three JRC images and the DGPS contours on 7 lakes were first calculated to provide pixel-based comparisons of the method's ability to detect water in the smallest reservoirs. Monthly surface water values extracted from JRC datasets over 1999-2014 for 7 small reservoirs

5    were then compared with mean monthly surface areas based on in situ monitoring. RMSE and RSE values on monthly surface water dynamics as well as mean annual water availability were again calculated to compare with results from our approach. In parallel, the number of NA values in monthly water history rasters within each lake grid cell were collected for each image to reduce their detrimental influence on surface water area estimates. NA values result from clouds and SLC-off interferences which are not differentiated in the JRC global water data sets.

## 3    Results & discussion

### 3.1    Spatial accuracy of spectral water indices in small water bodies

#### 3.1.1    Water classification with variable thresholds (calibration)

All six spectral water indices performed well when calibrated for a single image (March or June), achieving consistently high overall accuracy rates (Fig. 4) over 80%. In high waters (March), associated surface area errors remain below 20% on 4 out

15    of 6 lakes. Even on the smallest lake (Gbatis) the PDAI values varied between 9% and 27% depending on the index, showing remarkable precision considering its size of 1 ha and the coarse 30 m image resolution.

In low waters (June), PDAI values degraded however for NDVI, NDWI and NDMI and exceeded 50% on the smallest lake. Only MNDWI and NDTI maintained PDAI rates below 25% for all lakes. In the dry season, lower water levels, greater algae and vegetation growth make accurate discrimination more complex (Ji et al., 2009). Smaller lakes intrinsically have a

20    higher proportion of mixed pixels and shallow waters (Mialhe et al., 2008) where the spectral reflectance is a combination of water, soil and vegetation. The coarse resolution exacerbates these errors due to the proportionally greater influence of a single misclassified pixel on the result (1 ha corresponds to a little more than 9 pixels).

NDVI which suffers from inability to correctly distinguish vegetation from flooded vegetation sees its performance diminish accordingly. NDMI on the other hand performed better in low waters, coherent with its objective to detect moisture and all

25    types of water. Thanks to the middle infrared band which penetrates water less and is less affected by turbidity & vegetation in the water, NDMI was able to detect the two smallest lakes but low user accuracy led to gross PDAI errors, as it overclassified many vegetated pixels. MNDWI and NDTI fared the best with these smaller reservoirs (eg. Gbatis) due to their ability to detect turbid waters, and distinguish flooded vegetation over wet vegetation on land.

AWEI performed well on the first image, but the index failed on the other images with overall accuracies falling below 60%

30    on four lakes. These were here related to inherent discrepancies in the radiometric corrections apparently amplified by the nature of the AWEI equation, rather than contained as with other normalised band ratio indices. Normalising reflectance values between the 546 images may help reduce these issues, but as a result AWEI was not considered for the semi-automation of this
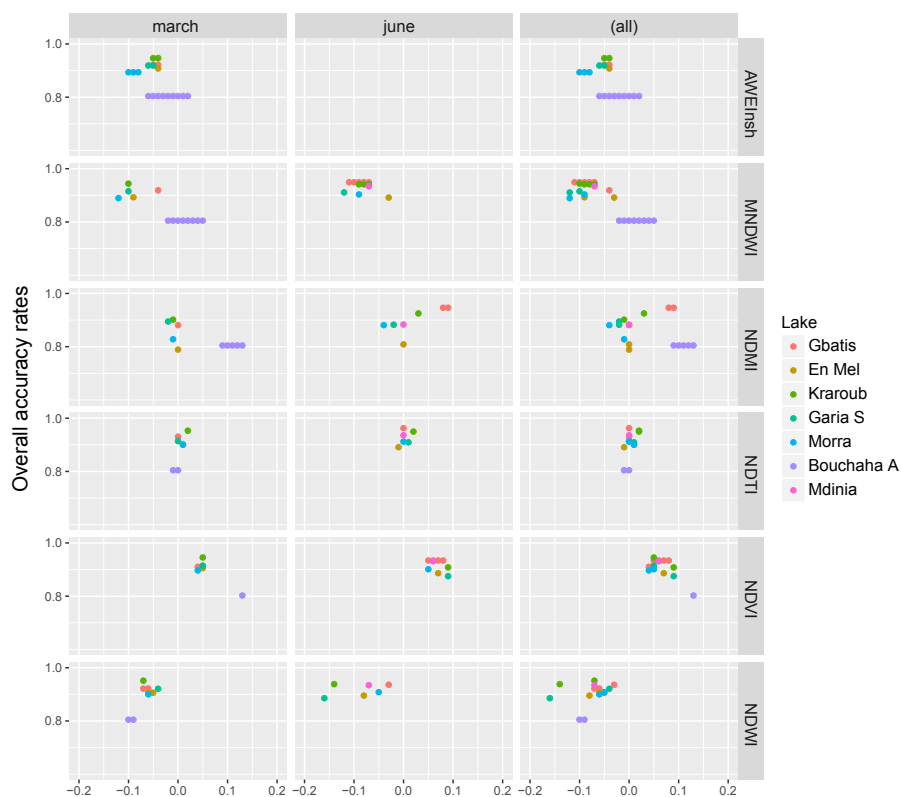
**Figure 4.** Optimal threshold (based on overall accuracy) for each lake & index on March and June calibration images

method. Furthermore, Jiang et al. (2014) showed that MNDWI outperformed AWEI when extracting narrow water bodies on Landsat.

### 3.1.2 Water classification with fixed thresholds (validation)

The optimal threshold determined from the March and June images on seven lakes was used in validation on a third image
5 (May). Accuracy rates remained high for MNDWI and NDTI but reduced significantly for the three other indices (Fig. 6).
Issues of shallow waters and vegetation were again prevalent here and even an image-specific threshold could only minimise
omission and commission errors to 50% with NDVI, resulting in PDAI of 46%.

The stability of the individual thresholds (Fig. 4) combines with the sensitivity to threshold change (Fig. 5) to determine the
amplitude of the errors when using fixed thresholds. A marginal shift in threshold due to shallow waters and abundant algae
10 will then introduce larger errors and omission rates with NDVI and NDWI reached 76% and 100% on the smallest lakes. The
significant rate of change observed highlighted that thresholds must be optimised up to 2 decimals. MNDWI achieved better
accuracies on these small lakes and achieved consistently the lowest PDAI errors, as a result of the lower rate of change of the
MNDWI index over NDTI (Fig. 5).
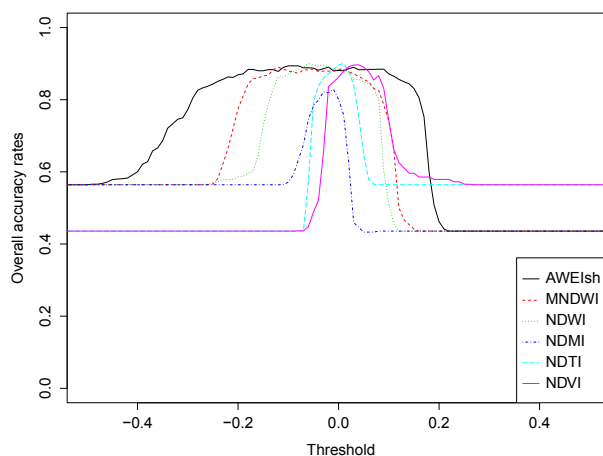
**Figure 5.** Rate of change of overall accuracy rates for varying thresholds of MNDWI on Landsat 8 29.03.2013 image for lake Morra
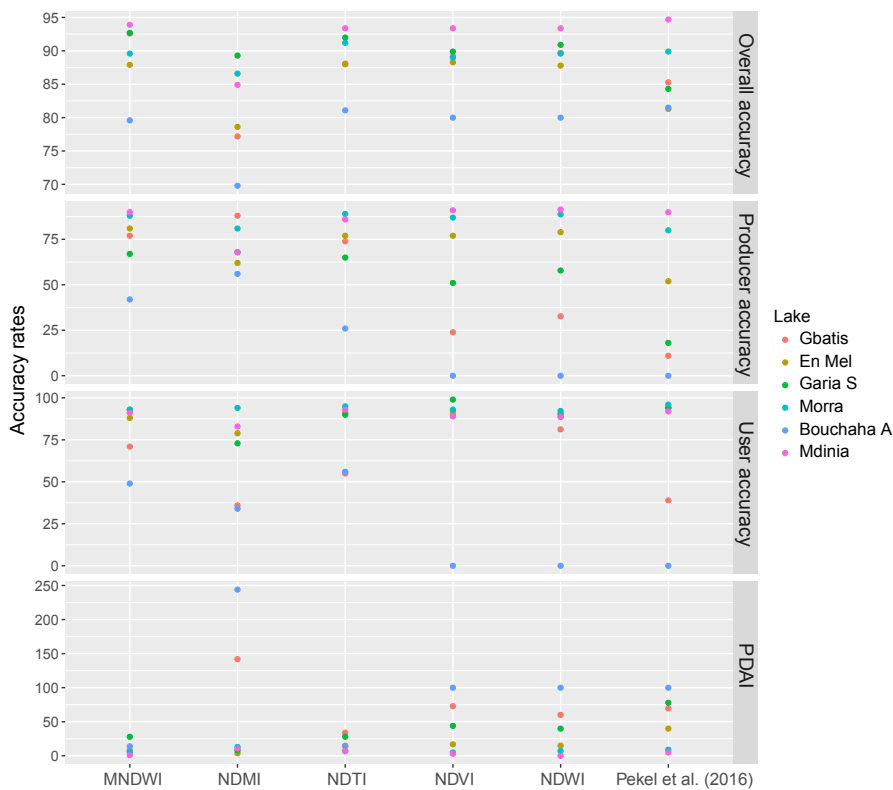
**Figure 6.** Accuracy and error values for 5 spectral water indices and the Pekel et al. (2016) expert systems approach on May validation image

**Table 2.** Landsat 8 MNDWI flooded surface area assessments compared to field measurements for 7 lakes on the calibration and validation images

|  | Lake name | Gbatis | En Mel | Kraroub | Garia S | Morra | Bouchaha A | Mdinia |
|---|---|---|---|---|---|---|---|---|
|  | Cell n° | 18 | 21 | 25 | 28 | 30 | 35 | 50 |
| Calibration | L8 surface area 29.03.13 (ha) | 1.08 | 5.94 | 8.01 | 2.16 | 8.19 | 0.18 | 7.56 |
|  | GPS surface area 26.03.13 (ha) | 0.99 | 5.60 | 7.78 | 2.51 | 7.67 | 0.18 | N/A |
|  | L8 surface area 09.06.13 (ha) | 0.45 | 4.41 | 4.86 | 1.44 | 6.75 | 0.09 | 6.39 |
|  | GPS surface area 11-14.06.13 (ha) | 0.37 | 4.42 | 5.29 | 2.07 | 7.07 | N/A | 6.30 |
| Validation | L8 surface area 24.05.13 (ha) | 0.72 | 4.59 | 17.37[*] | 1.62 | 7.20 | 0.18 | 6.12 |
|  | GPS surface area 21.05.13 (ha) | 0.67 | 4.85 | 5.92 | 2.25 | 7.33 | 0.21 | 6.20 |
|  | Overall accuracy (%) | 92.6 | 87.9 | 64.8 | 92.7 | 89.6 | 79.6 | 93.9 |
|  | Producer accuracy (%) | 77 | 81 | 86 | 67 | 88 | 42 | 90 |
|  | User accuracy (%) | 71 | 88 | 38 | 93 | 93 | 49 | 91 |
|  | PDAI (%) | 8 | 7 | 244 | 28 | 5 | 14 | 1 |

[*] 46% of clouds and 50% shadows were detected over this lake on this L8 image

These results confirm firstly its suitability over other band ratio indices on the smallest lakes, in line with results on larger water bodies (Feyisa et al., 2014; Ji et al., 2015; Jiang et al., 2014; Ogilvie et al., 2015). Applied across seven lakes and three images, the method yields a $R^2$ Pearson correlation coefficient of 0.99 and low RMSE errors, confirming secondly that a fixed-threshold MNDWI can be used to study over time lakes from 1 ha with 30 m Landsat images (Table 2). Surface area errors

5 ranged from 1 to 28% (mean 10.5%, s.d. 8.7%) with some underestimations observed due to undetected narrow inlets on certain lakes. On the smallest lakes (Bouchaha A and Gbatis) however errors reached only 14 and 8%. In in the dry season, Bouchaha A reduced to only 0.2 ha, therefore PDAI error reached 50% with all indices which is to be expected as it corresponds to just over 2 Landsat 30 m pixels.

Optimising thresholds for each lake and image individually can reduce PDAI errors below 5%, highlighting the potential for

10 further improvements by developing (unsupervised) classification methods capable of being used over numerous locations and images reliably and efficiently (in terms of CPU and user input). Optimising thresholds for each lake but not each image only improved performance on certain small lakes.

### 3.1.3 Comparison with previous approaches

Previous studies to map small reservoirs yielded PDAI errors of 10% to 50% on lakes in the range 1.5 - 5 ha and with image-

15 specific thresholds (Liebe et al., 2005; Sawunyama et al., 2006). Errors in this study remained minimal even on lakes of this size despite the same spatial resolution of the images, pointing to the improved ability of indices in detecting mixed pixels caused by shallow waters. Including mixed pixels from small reservoirs in the calibration also improved performance (Ji et al.,
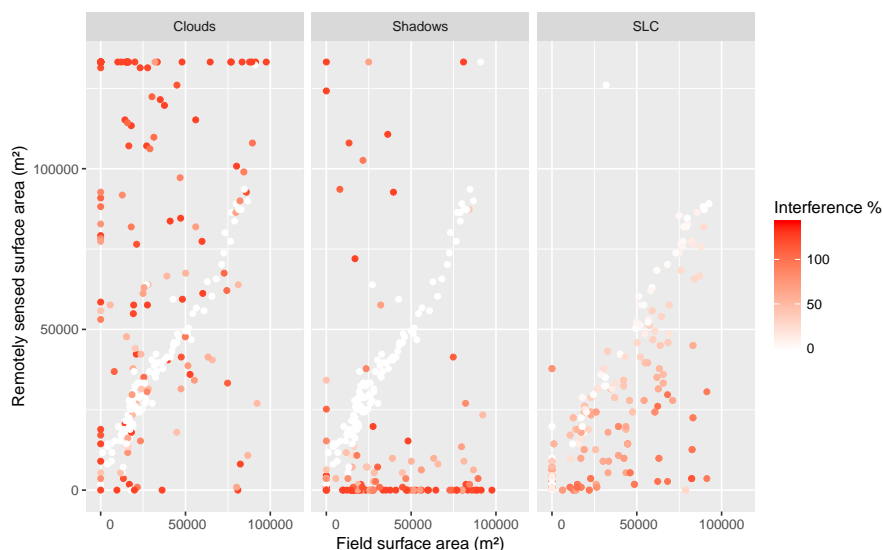
**Figure 7.** Errors due to interference from clouds, shadows and SLC-off percentage across Gouazine lake cell

2009; Jain, 2010) and led to a slightly lower threshold (-0.09). The MNDWI threshold for water detection is often near 0 but here a value close to that found in wetlands (Ogilvie et al., 2015) with similar mixed spectral profiles was determined.

Confusion matrices performed on the JRC monthly surface water rasters (Pekel et al., 2016) confirmed their excellent performance on larger lakes in high waters (e.g. Morra, Mdinia) with PDAI under 10% and overall accuracies near 90%. Overall accuracy values remained high ($> 80\%$) on other lakes, due to the large number of non water pixels correctly classified in each lake cell but water omission errors (i.e. producer accuracy) on the smaller lakes (under 5 ha) were substantial, rising in some cases over 50% (Fig. 6). Commission errors remained minimal (10% on average) indicating that the expert system approach suffered from the inability to detect all types of water, rather than from overdetection. Confusion matrices on the June images confirmed the progressive decline in accuracy as water reduces and the presence of flooded vegetation, algae and shallow waters increase. In June, overall accuracy rates declined and PDAI for instance rose from 9.1% to 18.7% on Morra and up to 74.7 % on Kraroub where significant algae and vegetation growth was observed. In comparison, MNDWI designed to detect mixed water reflectance led to a 7.1 % PDAI on this 5 ha lake. Across all 7 lakes, omission errors reached 32% in high waters and 41% in shallow waters. These are substantially higher than values provided in Pekel et al. (2016) (23% for seasonal water bodies in Landsat 8) and in Mueller et al. (2016) (12% in small water bodies and 26% where water and vegetation mix). These suggest that the validation approach which uses in Pekel et al. (2016) sample pixels with "high probability of water occurence" within 1° tiles introduced a bias towards larger, near permanent water bodies, which also contained greater proportions of clear water pixels.
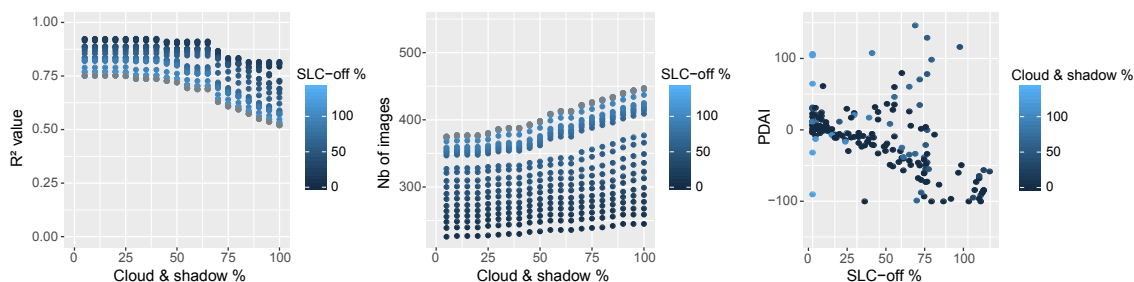
**Figure 8.** Change in $R^2$ (between field and remotely sensed surface areas), remaining number of Landsat scenes and PDAI for increasing percentages of SLC-off and cloud & shadows pixels tolerated (Lake Gouazine)

### 3.2 Influence of clouds, shadows & SLC-off pixels on multi-temporal analysis

#### 3.2.1 Interference from clouds, shadows & SLC-off

The influence of clouds, shadows and SLC failure on estimated flooded areas from Landsat imagery over 1999-2014 are illustrated in Fig. 7 and Fig. 8 based on field data from one small reservoir. Detected on 28% of the 546 Landsat images, clouds lead essentially to commission errors (false positives) and overestimation, as a result of visible wavelengths being reflected while much of the electromagnetic energy is absorbed by the droplets in clouds. Their effect was not systematic and proportional (Fig. 8), as shown by underestimations in 30% of cases where clouds were detected across the whole lake cell. The diversity in the nature and properties (temperature, thickness, water content...) of clouds are indeed firstly responsible for the heterogeneous reflectance observed and the resulting classification difficulties. The influence of shadows is more moderate and less frequent, responsible for overestimations in only 5% of images. This results partly from the greater difficulties in discriminating shadows and the overlap from clouds. Pixel loss from SLC failure varies across Landsat tiles and for a given lake, 30% of the 287 Landsat 7 SLC-off images suffered from minor pixel loss (< 10%). SLC-off pixel loss led to a systematic underestimation of flooded areas, on average by 35% (Fig. 8).

#### 3.2.2 Optimising image availability vs. interferences

Optimising on $R^2$ tends towards removing all cloud & SLC interferences, and therefore reduces the number of available observations. Considering the objective of hydrological monitoring and maintaining sufficient temporal repetitivity, images with up to 25% SLC-off pixels and 40% cloud & shadow pixels over the studied lake cell were retained here. These thresholds were found to minimise the mean squared error on surface area aggregated over 15 years which gives importance to both the number and quality of the Landsat observations over time.

This calibration removed a significant 51% of images, leaving on average 282 ±27 Landsat images suitable for observations across these 7 small reservoirs (Table 3). This equates to 1.5 images/month over the 1999-2014 period studied though these are unevenly distributed as seen in Fig. 9, due partly to low overall cloud cover between March 2013 - December 2014 over
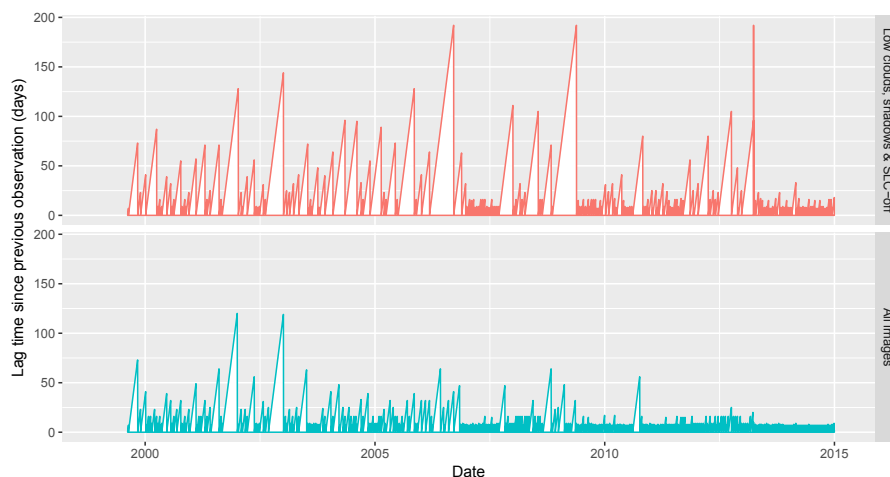
**Figure 9.** Lag time between successive Landsat images, based on all images available and after optimisation to reduce clouds, shadows and SLC-off interferences

our area of interest. The acquisition cycle of Landsat 7 is also staggered by 8 days compared to Landsat 5 and 8, leading to a revisit frequency up to 8 days. The catchment is located at the overlap of two paths visited at 7 days interval, further increasing availability.

Significantly when studying one lake, 59% of the Landsat 7 post-2003 images were discarded due to SLC-off pixels, and
5  31% of all Landsat images were removed due to excessive cloud presence. Variations between lakes in the number of images available were high when using the same thresholds, due to greater cloud presence on higher altitude lakes (Morra) as well as the irregular SLC interference detailed earlier. These confirm the value of lake-specific cloud detection to account for the scale of small reservoirs (i.e. a 0.1 $\mathrm{km}^2$ area within a 33 300 $\mathrm{km}^2$ Landsat 8 tile, i.e. 0.0003%). In other catchments, proximity to coastal areas and more pronounced orography may increase the proportion of images affected by clouds.

10  ### 3.3 Remote monitoring of surface water dynamics in small reservoirs

Results from this semi-automated method applied across 546 Landsat images (1999-2014) and 7 lakes yielded a strong correlation ($R^2 = 0.92$) between remotely sensed flood areas and field data (Fig. 10). Confirming its excellent performance in terms of accuracy and stability across reservoirs and over several years, results also substantiate the method's ability to be used across multiple sensors notably Landsat 8 OLI, with Landsat 7 ETM+ and Landsat 5 TM, in line with what Vogelmann et al. (2001)
15  showed on TM and ETM+. On larger lakes with long term and accurate time series such as Gouazine, $R^2$ reached 0.95 with an RMSE of 6800 $\mathrm{m}^2$ (Table 3). When accounting for the lake size, this equates to a median relative (PDAI) error of 11% (Fig. 11).

Landsat observations provided here accurate insights into the surface water extent dynamics, closely reproducing flood peaks and declines notably (Fig. 12). Reduced image availability due to clouds, shadows and SLC-off however prevented accurate
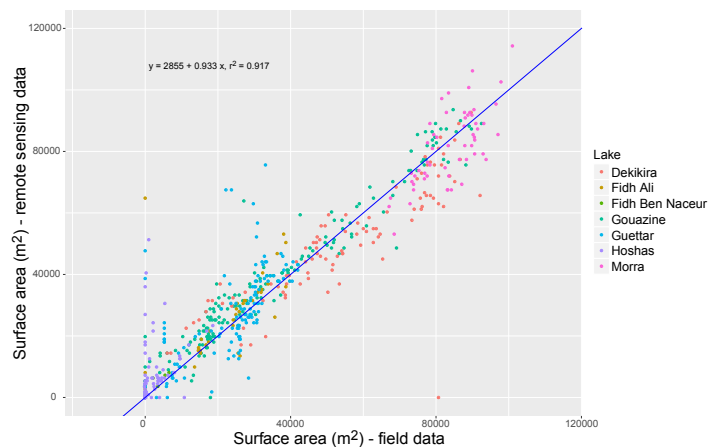
**Figure 10.** Remotely sensed surface areas with MNDWI on Landsat imagery versus field data over 1999-2014 across all 7 lakes
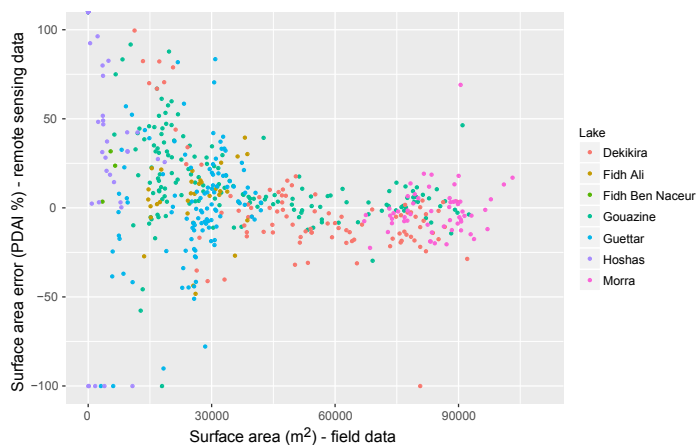
**Figure 11.** Surface area error (in PDAI %) from remote sensing data against field data for 7 lakes

representation (interpolation) of the steep flood rises associated with intense rapid floods as opposed to the more gradual declines, easier to interpolate. As seen in Fig. 7 and Fig. 8, increasing the allowance of clouds, shadows and SLC-off pixels to raise image availability leads to a decline in $R^2$ and a rise in PDAI. Conversely, single outliers remain, but reducing cloud, shadows and SLC thresholds was shown not to remove these larger classification errors. These outliers are due to the inherent

5  uncertainties in the method, partly incomplete detection of clouds and classification difficulties (e.g. overestimation as a result of detecting water flow in the tributaries or peripheral flooding due to cells being defined broadly to account for exceptional flood volumes).
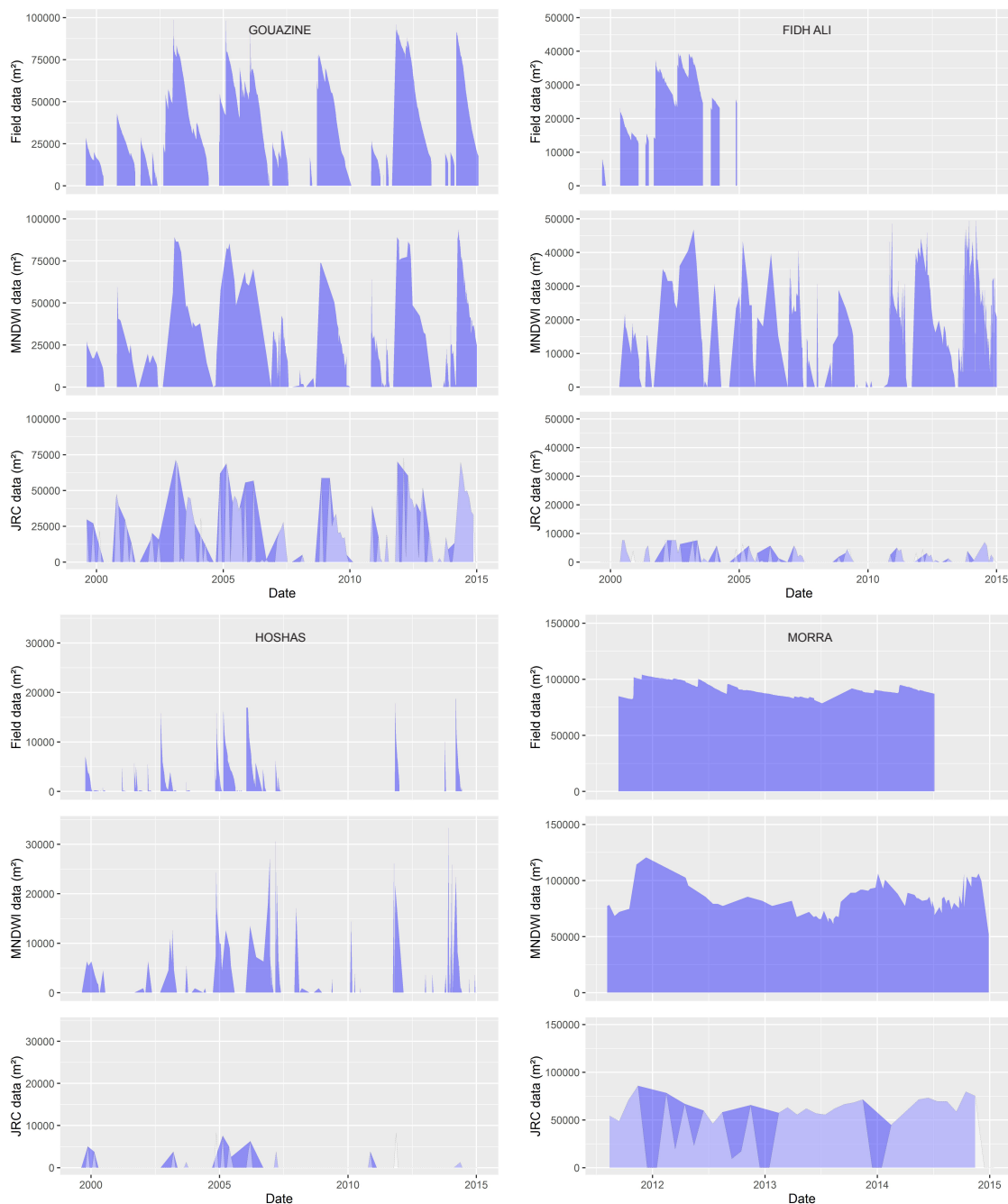
**Figure 12.** Surface water extent dynamics from field data and remote sensing over 1999-2015 for 4 lakes. JRC data (Pekel et al., 2016) is filtered to images with less than 25% NA over the lake cell (the complete JRC datasets are represented in pale blue)

### 3.3.1 Greater errors on small lakes

On smaller lakes, the influence of incorrectly detected pixels is proportionally greater due to the low spatial resolution and reduced pixel numbers. Mean PDAI rises from 13% to 27% on surface areas under 3 ha and scattering is reflected in the reduced NSE skill values (Fig. 11 and Table 3). On Hoshas, this is exacerbated by the numerous empty periods and the greater
5   mathematical influence on $R^2$ of minor classification noise. Errors are also proportionally greater at the lower part of the H-S rating curves. Over a reduced period (2011-2014) where the lake was regularly flooded and data confidence is greater, $R^2$ for Hoshas rises to 0.77. Smaller lakes also suffer from greater classification difficulties, discussed in calibration of indices, due to more heterogeneous pixels where soil, vegetation and water reflectances overlap. Threshold calibration of our MNDWI index led here to a minor overestimation for the lowest water levels, as seen in Fig. 11 and for Fidh Ali and low flood levels in
10  Gouazine in Fig. 10. In absolute terms, however, mean error remains contained to 9 300 m$^2$ for both small and larger lakes.

### 3.3.2 Dealing with rapid flood declines

On these smaller lakes, the method continues in many cases to be able to derive time series reproducing surface water dynamics, notably highlight drought periods and the minor flooded areas (e.g. Fidh Ali in Fig. 12). However, the reduced amplitude of the flood and crucially the rapid decline can lead to difficulties. Out of 7 events on the smallest lakes (Fidh Ben Nasseur and
15  Hoshas), 2 minor events (0.5 ha flood) were completely undetected and two were detected after surface area had declined by 50%. On these lakes, 1 ha floods were completely infiltrated and evaporated within 21 days. Figure 12 also illustrates the difficulty in interpolating points with variable and non negligible time lag into a hydrologically coherent daily time series. Available observations remained however relatively accurate on Fidh Ben Nasseur ($R^2 = 0.91$), showing the potential of the method even on the smallest lakes (< 1 ha), subject to sufficient temporal resolution.

20  ### 3.3.3 Reduced surface water amplitudes

On the largest lake studied here (Morra), GPS contours revealed the obsolescence of the stage-surface rating curve, which has not been updated over the past 20 years due to the lake remaining flooded through the year and the access difficulties to perform bathymetric surveys. This partly explains the low $R^2$ calculated but results also display difficulties in extracting consistent flood trends from the remote sensing data (Fig. 12). These can be explained by the morphology of this large deep lake, where the
25  uncertainties of the remote sensing method are significant compared to the amplitude of the surface area variations (80 000 - 100 000 m$^2$ on Morra against 0 - 90 000 m$^2$ on Gouazine). Accordingly, it is not only the spatial resolution and number of pixels but the amplitude of variations in the flooded surface area over time (e.g. 0 - 90 000) which determine how much remote sensing of surface water extent is affected by uncertainties and scattering.

### 3.4 Remote assessment of long term water availability in small reservoirs

30  Daily time series of surface area aggregated and converted into mean annual surface volumes are shown in Fig. 13. The fit with the mean volume assessed from stage data is again excellent for Gouazine ($R^2 = 0.91$) (Table 3). On other lakes, the absence of

**Table 3.** Landsat RS performance on daily surface areas and mean annual volumes

| Lakes (observation period) | Initial capacity ($10^3$m$^3$) | Mean surface area (m$^2$) | Nb images (1999-2014) | Daily surface area | | Mean annual volume | |
|---|---|---|---|---|---|---|---|
| | | | | NSE | RMSE (m$^2$) | NSE | RMSE (m$^3$) |
| Gouazine (1999-2014) | 237 | 28700 | 269 | 0.95 | 6800 | 0.91 | 14400 |
| Dekikira (1999-2008) | 219 | 53900 | 240 | 0.73 | 12200 | 0.59 | 36600 |
| Morra (2011-2014) | 705 | 84200 | 250 | 0.43 | 12200 | 0.29 | 87900 |
| Guettar (2011-2014) | 150 | 24000 | 302 | 0.69 | 9600 | 0.81 | 8500 |
| Fidh Ali (1999-2005) | 134 | 15800 | 308 | 0.81 | 6900 | 0.73 | 13100 |
| Hoshas (1999-2014) | 130 | 1300 | 310 | 0.18 | 7200 | 0.38 | 5000 |
| Fidh Ben Nasseur (1999-2001) | 47 | 800 | 298 | 0.91 | 900 | N/A | 600 |
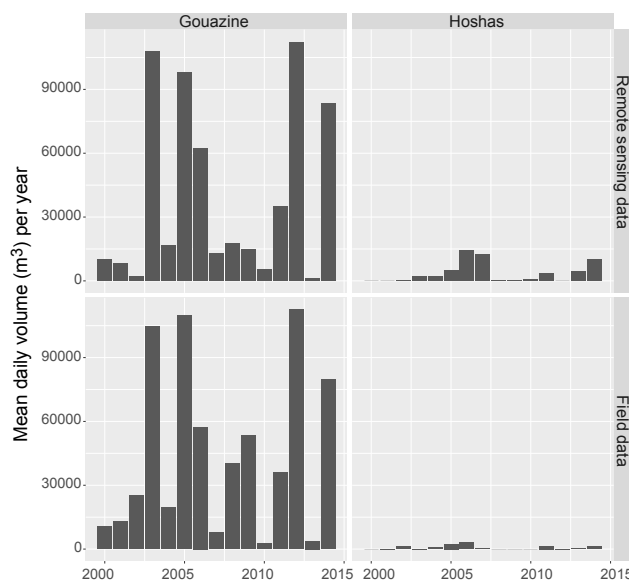
**Figure 13.** Annual mean volumes from field data and remote sensing data over 1999-2015 for 2 lakes

stage data over several years limited the number of values which could be compared, but for available years Fig. 14 indicates a good overall fit ($R^2$ = 0.82). The method therefore provides a valuable tool to assess and compare water resources between years and lakes. Results here highlight the significant variability in terms of inter-lake and inter-annual water availability that can be observed across all small reservoirs in the catchment (Ogilvie et al., 2016a).

5    $R^2$ values increased for certain lakes as a result of the smoothing of the method's uncertainties (e.g. Guettar and Hoshas), but conversely reduced on other lakes (e.g. Dekikira). Interpolation over time between discrete Landsat observations can lead single outliers to reduce overall performance if they are not rapidly corrected by a subsequent observation. The influence of outliers is
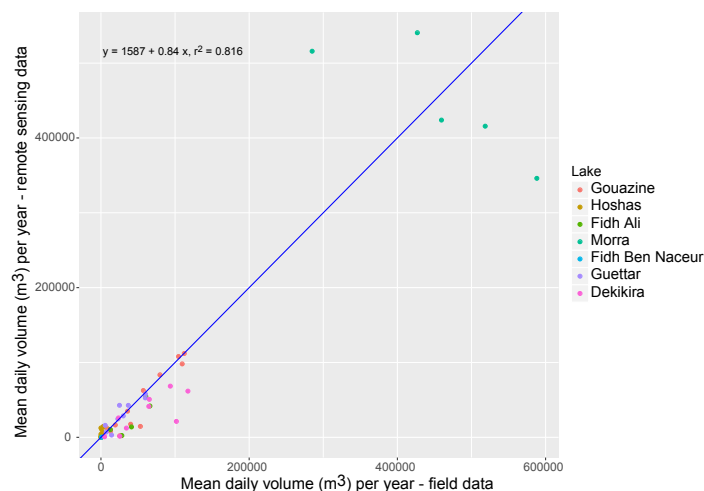
**Figure 14.** Mean daily volume per year with MNDWI on Landsat imagery versus field data over 1999-2014 across all 7 lakes

then variable and depends on the lag between successive correct observations. Similarly, all observations do not carry the same importance when assessing mean annual volumes. Landsat observations close to the peak of an event are accordingly more important than observations during dry periods or the (gradual) decline of flooded areas, which can be more easily interpolated.

### 3.5   Comparison with published global data sets

5  Analysis of available monthly surface water datasets (Pekel et al., 2016) against field data for 7 lakes over up to 15 years (1999-2014) highlighted firstly the incidence of NA pixels and the necessity to assess and remove NA pixels above individual lakes to reduce underestimations (Fig. 15) and reproduce coherent flood dynamics (Fig. 12). Filtering out images with over 25% NA values per lake cell led to optimal RMSE errors and is coherent with the SLC threshold determined with our approach. After NA removal, NSE across all 7 lakes rose from 0.39 to 0.78 but Fig. 16 confirms the residual difficulties, consistent with

10  the omission errors identified on single JRC water history rasters (Fig. 6). Mean RMSE reaches 22 500 $m^2$ and removing all NAs only improves RMSE further to 21 800 $m^2$ compared to 9 300 $m^2$ with the MNDWI approach. On small lakes under 5 ha which present greater shallow waters and flooded vegetation, PDAI errors are proportionally more important and reach 61% on average. On Fidh Ali (Fig. 12), difficulties due to the significant vegetation in the lake lead to vast underestimations and fail to detect certain floods completely. On surface areas over 5 ha, PDAI errors remain high around 35% and interestingly continue

15  to underestimate surface water areas. These result from omission errors during summer months due to vegetation growth in permanent lakes as well as undetected vegetated areas on the fringes inundated during large floods. Figure 12 highlights the valuable insights into the general flood dynamics that can be extracted from the JRC datasets on larger lakes such as Gouazine and Morra, though omission errors remain apparent in the underestimated flood peaks.

These classification inaccuracies are further exacerbated by the limited number of observations. After NA removal, JRC

20  monthly water history sets provide for instance 99 exploitable images over 1999-2014 on Gouazine, compared to 269 when
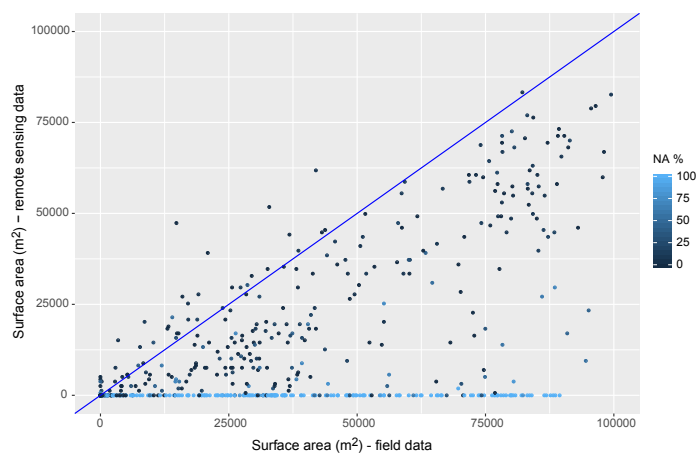
**Figure 15.** Remotely sensed surface areas with global surface water data sets (Pekel et al., 2016) versus field data over 1999-2014 across all 7 lakes. Percentage of NA pixels was calculated over each lake cell
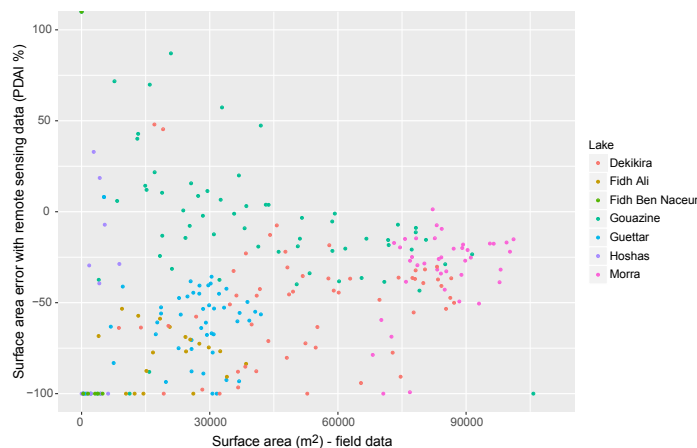
**Figure 16.** Surface area error (in PDAI %) from global surface water data sets (Pekel et al., 2016) against field data for 7 lakes

using the complete L5, L7 and L8 archives filtered for clouds, shadows and SLC failure (Fig. 8 and Fig. 9). The greater temporal resolution allows better detection of dry spells, multiple flood peaks, timing of flood peaks and more coherent flood rises and declines as seen on Gouazine, while on the small Hoshas lake two floods in 2013-2014 are otherwise undetected with JRC datasets (Fig. 12). These difficulties also have implications on the accuracy of maximum extent estimates, and though median underestimation is 41%, the spread in the error (Fig. 16) even for larger lakes, prevents systematic corrections. Surface area inaccuracies on small lakes in JRC datasets also degrade water resource assessments for farmers or water balances for hydrologists, with for instance Gouazine RMSE on mean water availability increasing by 24% to $17\,800\ \mathrm{m}^3$.

## 4   Conclusions

Validated against significant hydrometric field data, this paper demonstrates the potential and limitations of Landsat imagery to monitor surface water extent dynamics and water availability within the smallest reservoirs (1-10 ha). Results confirmed the superiority of MNDWI out of six widely used band ratio indices to detect flooded areas in seven reservoirs of varying sizes and

5    over three flood stages. Overall accuracy rates were high despite difficulties of shallow water and flooded vegetation and the MNDWI threshold remained sufficiently stable to allow automation across several lakes from 1-12 ha and over Landsat 5, 7 and Landsat 8 images. The performance of other water detection indices when used in semi automation (fixed thresholds) reduced significantly due to their greater sensitivity to minor changes in thresholds. With MNDWI, surface area errors remained low, under 15% on 5 lakes including the smallest ones, remarkable considering the demanding objective of detecting areas under 1

10   ha with 30 m spatial imagery.

Applied over 546 images, results highlight the impact of spatial inaccuracies as well as cloud & shadow interferences over small lakes. Residual errors from cloud uncertainties, due to undetected cirrus clouds and shadows, as well as greater vegetation, shallow water and associated mixed pixels led to a number of outliers which can be detrimental notably on smaller lakes. Filtering images with excessive cloud and SLC presence reduced temporal resolution to 1.5 images/month which can

15   lead to minor floods and those subject to high infiltration to be overlooked by Landsat time series. With a mean RMSE of 9 300 m$^2$ and $R^2$ reaching 0.9, the method however confirmed the operational potential of Landsat imagery to study surface water dynamics of floods over 3 ha, as well as long term water availability ($R^2 = 0.82$).

Comparisons with global water data sets (Pekel et al., 2016) notably confirmed the benefits and relevance of specific, complementary approaches to improve classification in smaller water bodies. Designed and calibrated to account for the specificities

20   of small reservoirs, this semi-automated MNDWI approach reduced RMSE by 57% in comparison, by minimising omission errors from flooded vegetation and shallow waters. Results also argue for the importance of exploiting all Landsat archive imagery to maximise temporal resolution, capture flood peaks, reproduce coherent flood dynamics and improve water resource assessments considering the rapid flood declines in small lakes. Clearer distinction of clouds and SLC-off in the JRC datasets would have allowed finer filtering of images due to the non systematic cloud error observed here.

25   Accurate long term water availability assessments are notably required to inform farmers of water resources in the millions of small reservoirs worldwide (Wisser et al., 2010; Lehner et al., 2011) and improve hydrological modelling. Using free medium resolution Landsat imagery, this method may be replicated regionally or globally to provide long term data sets across small water bodies. Future research will help clarify whether the apparent stability of the MNDWI threshold across these 7 lakes and in other regions (Ogilvie et al., 2015) allows the approach to be transposed without additional field calibration and whether

30   automatic thresholding approaches may be complementary (Vala and Baxi, 2013; Coltin et al., 2016).

This research provides evidence of the opportunities from remote sensing in small-scale hydrology, secured by the next generation of optical sensors providing free images every 5 days (e.g. Sentinel-2, or future SWOT) with high spatial resolution (< 10 m). Monitoring flood dynamics remains however dependent on low cloud cover and results here point to the value of assessing their presence at the lake and not image level. They also highlight the need for continued research in improving cirrus

cloud & shadows detection and notably specific methods capable of addressing the incremental change even on small land use objects (Zhu and Woodcock, 2014). Alternate optimisation of clouds & SLC-off or concomitant imagery sources including active sensors (Eilander et al., 2014) could also be used to maintain more observations at critical stages (e.g. flood peaks). Remotely sensed observations may also be combined with field data and modelling to overcome image availability issues and

5 problems of incoherent hydrological dynamics.

*Code and data availability.* Landsat L1T data are available from USGS Earth Explorer. Monthly water history rasters (Pekel et al., 2016) are available through Google Earth Engines, image collection ID: JRC/GSW1_0/MonthlyHistory. Fmask is available at https://github.com/prs021/fmask. Details on the additional scripts and R packages used to process the Landsat imagery are provided in Appendix A

## Appendix A: Remote sensing preprocessing

10 The following preprocessing chain was applied to the 546 Landsat images. Functions belonging to the R *rgdal*, *raster*, *sp*, and *landsat* packages were used but their code was adapted to account for the differences in Landsat 8 bands and metadata files.

### A1  Radiometric normalisation

Quantized calibrated (QCAL) digital numbers (DN) for each pixel were converted to at sensor (a.k.a. top of atmosphere, TOA) spectral radiance $L$ by reversing the calibration process used by USGS (Chander et al., 2007). The linear transformation

15 used (Eq. (A1)) can be rewritten as Eq. (A2) as $DN_{min}$ is 0 and $DN_{max}$ is 255. $G_{rescale}$ and $B_{rescale}$ are the "radiance multiplicative" and "radiance additive" rescaling factors, sometimes called gain and bias (or offset). These sensor and band specific factors were extracted automatically for each band and image from the Landsat metadata files.

$$(DN - DN_{min}) \cdot (L_{max} - L_{min}) = (DN_{max} - DN_{min}) \cdot (L - L_{min}) \tag{A1}$$

$$L = G_{rescale} \cdot DN + B_{rescale} \tag{A2}$$

20    where

$$G_{rescale} = \frac{L_{max} - L_{min}}{DN_{max} - DN_{min}} = \frac{L_{max} - L_{min}}{255} \tag{A3}$$

$$B_{rescale} = L_{min} - \frac{L_{max} - L_{min}}{DN_{max} - DN_{min}} DN_{min} = L_{min} \tag{A4}$$

### A2  Atmospheric corrections

At sensor radiance were then converted to at canopy (at surface) reflectance using the Dark Object Subtraction method (Chavez,

25 1996). The DOS approach encompasses conversion to at sensor reflectance to account for intra-annual differences in earth to

sun distance and solar illumination (Paolini et al., 2006). It also incorporates atmospheric corrections to estimate top of canopy (TOC) reflectance (Eq. (A5)) and account for the complex distorting effect of air molecules, aerosol particles, water vapour, carbon dioxide, methane, etc. which absorb and scatter electromagnetic waves (Hagolle et al., 2015). Absolute corrections were preferred over relative atmospheric correction methods such as the pseudo invariant features (PIF) and histogram matching as

5 these do not account for the band specificity of atmospheric disturbances and are therefore not recommended with band ratio classifications (Lu et al., 2002). Furthermore, key inputs for physically based models of (part of the) atmospheric effects such as aerosol optical thickness were not available from nearby Aeronet sites for historical Landsat 5 & 7 data. DOS methods provide a simple, widely used method for atmospheric corrections, proven to perform well and as reliably as other algorithms and methods (Paolini et al., 2006; Lu et al., 2002; Song et al., 2001).

10 $$\rho_{TOC.DOS} = \frac{(L_{sat} - L_{haze}) \cdot \pi \cdot E_{sun}^2}{T_{auv} \cdot (Ex_{sun} \cdot cos(90 - sun_{elev}) \cdot T_{auz} + E_{down})} \tag{A5}$$

$$L_{haze} = SHV - L_{sat1\%} \tag{A6}$$

$$L_{sat1\%} = 0.01 \cdot \frac{E_{sun} \cdot cos(90 - sun_{elev})}{\pi \cdot E_{sun}^2} \tag{A7}$$

$$L_{haze} = G_{rescale} \cdot DN_{min} + B_{rescale} - 0.01 \cdot \frac{E_{sun} \cdot cos(90 - sun_{elev})}{\pi \cdot E_{sun}^2} \tag{A8}$$

The DOS method supposes that certain objects within an image are dark (in complete shadow) and that radiance values mea-

15 sured for these objects are due to atmospheric scattering (path radiance, $L_{haze}$) (Chavez, 1996). These dark objects are selected from the lowest DN value across at least 1000 pixels. This defines the starting haze value (SHV) which is converted from DN to TOA radiance according to Eq. (A2). As per Eq. (A8), an allowance is made considering that pixels are rarely completely dark and 1% natural reflectance is subtracted. $L_{haze}$ converted to satellite reflectance is then subtracted to reflectance values across the image (Eq. (A5)). The improved DOS employed here calculates $L_{haze}$ for all bands using only one band (blue band)

20 to "correlate haze values and maintain the spectral relationship between bands" (Goslee, 2011; Chavez, 1996). In the SWIR bands, where minimal scattering occurs, SHV is defined as 0 (Goslee, 2011).

$T_{auv}$ is the atmospheric transmittance from the target to the sensor, $T_{auz}$ is the atmospheric transmittance from the sun to the target and $E_{down}$ is the downwelling diffuse spectral irradiance at the surface. DOS1 level of correction which assumes no atmospheric transmission losses in either direction ($T_{auv}$ & $T_{auz}$=1) and no diffuse downward radiation at the ground

25 ($E_{down}$=0), i.e. corrects only for additive scattering was used here. More complex corrections (for haze, aerosols) are less adapted to normalising successive images as required here and do not improve overall accuracies (Song et al., 2001; Goslee, 2011).

$E_{sun}$ is the earth to sun distance in astronomical units (AU) provided in the metadata files for OLI. For other sensors (ETM+, TM), this was calculated based on the date of acquisition indicated in the metadata (and function ESDIST in the R Landsat

30 (Goslee, 2011) package).

**Table A1.** Exoatmospheric solar constant per band for ETM+ and TM sensors (Chander et al., 2009)

|   | Landsat 7 ETM+ | Landsat 5 TM |
|---|---|---|
| 1 | 1997 | |
| 2 | 1812 | 1796 |
| 3 | 1533 | |
| 4 | 1039 | |
| 5 | 230.8 | 220 |
| 6 | | |
| 7 | 84.9 | |
| 8 | 1362 | |

$sun_{elev}$ is the solar elevation angle in degrees for the scene centre provided by the metadata file and again based on known changes in earth-sun geometry over the months. The complementary angle $(90-sun_{elev})$ provides the local solar zenith angle.

$Ex_{sun}$ is the band-specific exoatmospheric solar irradiance constant in $\mathrm{Wm}^{-2} \times \mathrm{\mu m}^{-1}$ calculated based on Eq. (A9). For OLI, $L_{max}$ and $REF_{max}$ are provided for each band. For ETM+ and TM sensors, $Ex_{sun}$ for each band were taken from the literature (Table A1).

$$Ex_{sun} = \frac{L_{max} \cdot (\pi \cdot E_{sun}{}^2)}{REF_{max}} \tag{A9}$$

**A3   Topographic normalisation**

Finally, topographic corrections to account for the effect on surface reflectance values of reduced illumination from inclined surfaces were implemented. The Minnaert algorithm (Minnaert, 1941), shown to perform as well or better than C correction, Cosine correction and Gamma correction (Vanonckelen et al., 2013; Hantson and Chuvieco, 2011) was used. Minnaert adds a band specific constant $K$ to the cosine correction, a trigonometric correction defined in Eq. (A10) where $\rho_t$ is the reflectance value generated by the inclined surface, and $\rho_h$ is the value on a theoretical flat surface. Illumination is modelled as per Eq. (A11) where $\gamma_i$ is the incident angle, slope angle is $\theta_p$, solar zenith angle is $\theta_z$, solar azimuth angle $\phi_a$, and aspect angle $\phi_o$. Slope and aspect were derived in R from the SRTM version 3 $30\,\mathrm{m}$ digital elevation model, and information on solar azimuth and zenith angle were extracted from satellite and solar geometry in the metadata files.

$$\rho_h = \rho_t \cdot \left(\frac{cos\theta_z}{IL}\right)^K \tag{A10}$$
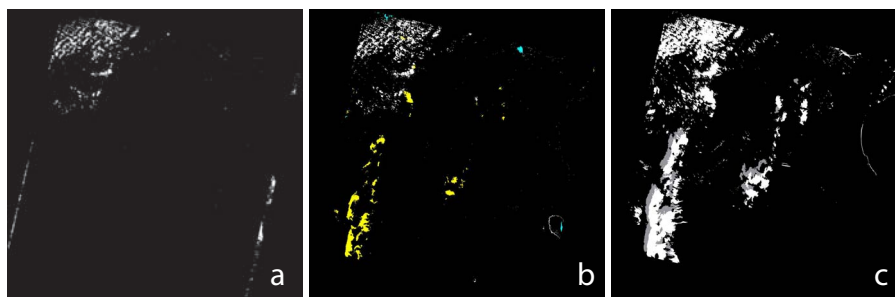
**Figure A1.** Comparing cloud detection methods on Landsat 8 29.03.2013 image: (a) Clouds detected over 1.7% of the image with ACCA(Irish et al., 2006) and Goslee (2011) methods; (b) Clouds incl. cirrus detected over 8.2% of the image with Landsat 8 1380nm band; (c) Clouds incl. cirrus and cloud shadows detected over 11.5% of the image with Fmask (Zhu and Woodcock, 2014) method.

$$IL = cos\gamma_i = cos\theta_p \, cos\theta_z + sin\theta_p \, sin\theta_z \, cos(\phi_a - \phi_o) \tag{A11}$$

**A4    Cloud and shadow detection per lake**

Fmask (Zhu and Woodcock, 2012) which uses top of atmosphere reflectances for all visible and IR bands, and brightness temperature for band 6 (Eq. (A12)) as inputs was used to assess clouds and shadow presence on each image and over each lake cell. It reports a 96.41% overall accuracy vs 84.8% for the Automated Cloud Cover Assessment (ACCA) algorithm (Irish et al., 2006) used originally by the USGS. DN values in Landsat thermal bands were converted to TOA radiance using Eq. (A2) and then to (brightness) temperature (Kelvin) using Eq. (A12).

$$brightness \; temperature = \frac{K_2}{log(\frac{K_1 \cdot \varepsilon}{L} + 1)} \tag{A12}$$

$K1$ and $K2$ can be found in the Landsat 8 metadata and in the literature for Landsat 5 and 7 (Chander et al., 2009). $\varepsilon$ is emissivity, and can be estimated as 0.95 to account for atmospheric interferences. The method then combines probability masks with a series of rule based tests to identify potential cloud pixels based on the cloud spectral behaviour. Cloud shadows are modelled based upon estimating cloud height based on temperature, view angle and solar illumination geometry reported in the metadata files (Huang et al., 2010; Zhu and Woodcock, 2012). In addition to cloud shadows, Fmask version (3.2.1) used here also exploits the Landsat 8 1380 nm wavelength for the detection of the cirrus clouds (Fig. A1). This wavelength present on other lower spatial resolution sensors is strongly absorbed by the water vapour in the (lower) atmosphere therefore only high cirrus clouds where water vapour is low will reflect solar radiation in this wavelength. As discussed, associated methods combining the information from several (15 for Tmask) successive images to improve cloud shadow detection (Zhu and Woodcock, 2014; Goodwin et al., 2013) were not suitable, due to the localised and rapid changes investigated here.

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* We gratefully acknowledge the collaboration from the Direction Générale de l'Aménagement et de la Conservation des Terres Agricoles (DG ACTA) and Direction Générale des Ressources en Eau (DGRE) in Tunis and local representatives at the Kairouan and Siliana Commissariat Régional au Développement Agricole (CRDA). These works were partly financed by the ANR AMETHYST and
5    SICMED DYSHYME projects. SPOT 5 10 m imagery was provided through CNES ISIS programme, request 703.

# References

Albergel, J. and Rejeb, N.: Les lacs collinaires en Tunisie : Enjeux, Contraintes et Perspectives, Comptes rendus de l'Académie d'agriculture de France, 83, 77–88, 1997.

Alsdorf, D., Bates, P., and Melack, J.: Spatial and temporal complexity of the Amazon flood measured from space, Geophysical Research Letters, 34, 6, https://doi.org/10.1029/2007GL029447, 2007.

Annor, F. O., van de Giesen, N., Liebe, J., van de Zaag, P., Tilmant, A., and Odai, S.: Delineation of small reservoirs using radar imagery in a semi-arid environment: A case study in the upper east region of Ghana, Physics and Chemistry of the Earth, Parts A/B/C, 34, 309–315, https://doi.org/10.1016/j.pce.2008.08.005, 2009.

Avisse, N., Tilmant, A., Müller, M. F., and Zhang, H.: Monitoring small reservoirs' storage with satellite remote sensing in inaccessible areas, Hydrology and Earth System Sciences, 21, 6445–6459, https://doi.org/10.5194/hess-21-6445-2017, 2017.

Baup, F., Frappart, F., and Maubant, J.: Combining high-resolution satellite images and altimetry to estimate the volume of small lakes, Hydrology and Earth System Sciences, 18, 2007–2020, https://doi.org/10.5194/hess-18-2007-2014, 2014.

Bédard, F., Reichert, G., Dobbins, R., and Trépanier, I.: Evaluation of segment-based gap-filled Landsat ETM+ SLC-off satellite data for land cover classification in southern Saskatchewan, Canada, International Journal of Remote Sensing, 29, 2041–2054, https://doi.org/10.1080/01431160701281064, 2008.

Bergé-Nguyen, M. and Crétaux, J.-F.: Inundations in the Inner Niger Delta: Monitoring and Analysis Using MODIS and Global Precipitation Datasets, Remote Sensing, 7, 2127–2151, https://doi.org/10.3390/rs70202127, 2015.

Bouma, J. A., Biggs, T. W., and Bouwer, L. M.: The downstream externalities of harvesting rainwater in semi-arid watersheds: An Indian case study, Agricultural Water Management, 98, 1162–1170, https://doi.org/10.1016/j.agwat.2011.02.010, 2011.

Burte, J., Coudrain, A., Frischkorn, H., Chaffaut, I., and Kosuth, P.: Impacts anthropiques sur les termes du bilan hydrologique d'un aquifère alluvial dans le Nordeste semi-aride, Brésil, Hydrological Sciences Journal, 50, 95–110, https://doi.org/10.1623/hysj.50.1.95.56337, 2005.

Chander, G., Markham, B. L., and Barsi, J. A.: Revised landsat-5 thematic mapper radiometric calibration, IEEE Geoscience and Remote Sensing Letters, 4, 490–494, https://doi.org/10.1109/LGRS.2007.898285, 2007.

Chander, G., Markham, B. L., and Helder, D. L.: Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors, Remote Sensing of Environment, 113, 893–903, https://doi.org/10.1016/j.rse.2009.01.007, 2009.

Chavez, P. S. J.: Image-based atmospheric corrections- revisited and improved, Photogrammetric Engineering and Remote Sensing, 62, 1025–1035, https://doi.org/0099-1112/96/6209-1025, 1996.

Chen, J., Zhu, X., Vogelmann, J. E., Gao, F., and Jin, S.: A simple and effective method for filling gaps in Landsat ETM+ SLC-off images, Remote Sensing of Environment, 115, 1053–1064, https://doi.org/10.1016/j.rse.2010.12.010, 2011.

CNEA: Etude d'impact des travaux de conservation des eaux et du sol dans le gouvernorat de Kairouan, Tech. rep., Centre National des Etudes Agricoles (CNEA), Tunisie, 2006.

Coltin, B., McMichael, S., Smith, T., and Fong, T.: Automatic boosted flood mapping from satellite data, International Journal of Remote Sensing, 37, 993–1015, https://doi.org/10.1080/01431161.2016.1145366, 2016.

Cudennec, C., Slimani, M., and Le Goulven, P.: Accounting for sparsely observed rainfall space-time variability in a rainfall-runoff model of a semiarid Tunisian basin, Hydrological Sciences Journal, 50, 617–630, https://doi.org/10.1623/hysj.2005.50.4.617, 2005.

Eilander, D., Annor, F., Iannini, L., and van de Giesen, N.: Remotely Sensed Monitoring of Small Reservoir Dynamics: A Bayesian Approach, Remote Sensing, 6, 1191–1210, https://doi.org/10.3390/rs6021191, 2014.

Feng, M., Sexton, J. O., Channan, S., and Townshend, J. R.: 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, 2016.

Feyisa, G. L., Meilby, H., Fensholt, R., and Proud, S. R.: Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery, Remote Sensing of Environment, 140, 23–35, https://doi.org/10.1016/j.rse.2013.08.029, 2014.

Fisher, A. and Danaher, T.: A water index for SPOT5 HRG satellite imagery, New South Wales, Australia, determined by linear discriminant analysis, Remote Sensing, 5, 5907–5925, https://doi.org/10.3390/rs5115907, 2013.

Gao, B.-C.: NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space, Remote Sensing of Environment, 58, 257–266, https://doi.org/10.1016/S0034-4257(96)00067-3, 1996.

Gardelle, J., Hiernaux, P., Kergoat, L., and Grippa, M.: Less rain, more water in ponds: a remote sensing study of the dynamics of surface waters from 1950 to present in pastoral Sahel (Gourma region, Mali), Hydrology and Earth System Sciences, 14, 309–324, https://doi.org/10.5194/hess-14-309-2010, 2010.

Gay, D.: Fonctionnement et bilan de retenues artificielles en Tunisie: approche hydrochimique et isotopique, Ph.D. thesis, Université Paris XI, France, 2004.

Goodwin, N. R., Collett, L. J., Denham, R. J., Flood, N., and Tindall, D.: Cloud and cloud shadow screening across Queensland, Australia: An automated method for Landsat TM/ETM+ time series, Remote Sensing of Environment, 134, 50–65, https://doi.org/10.1016/j.rse.2013.02.019, 2013.

Goslee, S. C.: Analyzing Remote Sensing Data in R : The landsat Package, Journal of Statistical Software, 43, 1–25, https://doi.org/10.18637/jss.v043.i04, 2011.

Grunberger, O., Montoroi, J., and Nasri, S.: Quantification of water exchange between a hill reservoir and groundwater using hydrological and isotopic modelling (El Gouazine, Tunisia), Comptes Rendus Geosciences, 336, 1453–1462, https://doi.org/10.1016/j.crte.2004.08.006, 2004.

Gumbricht, T., Wolski, P., Frost, P., and McCarthy, T.: Forecasting the spatial extent of the annual flood in the Okavango Delta, Botswana, Journal of Hydrology, 290, 178–191, https://doi.org/10.1016/j.jhydrol.2003.11.010, 2004.

Guo, M., Li, J., Sheng, C., Xu, J., and Wu, L.: A Review of Wetland Remote Sensing, Sensors, 17, 777, https://doi.org/10.3390/s17040777, 2017.

Habi, M. and Morsli, B.: Contraintes et perspectives des retenues collinaires dans le Nord-Ouest algérien, Sécheresse, 22, 49–56, https://doi.org/10.1684/sec.2011.0293, 2011.

Hagolle, O., Huc, M., Pascual, D., and Dedieu, G.: A Multi-Temporal and Multi-Spectral Method to Estimate Aerosol Optical Thickness over Land, for the Atmospheric Correction of FormoSat-2, LandSat, VENS and Sentinel-2 Images, Remote Sensing, 7, 2668–2691, https://doi.org/10.3390/rs70302668, 2015.

Hantson, S. and Chuvieco, E.: Evaluation of different topographic correction methods for Landsat imagery, International Journal of Applied Earth Observation and Geoinformation, 13, 691–700, https://doi.org/10.1016/j.jag.2011.05.001, 2011.

Huang, C., Thomas, N., Goward, S. N., Masek, J. G., Zhu, Z., Townshend, J. R. G., and Vogelmann, J. E.: Automated masking of cloud and cloud shadow for forest change analysis using Landsat images, International Journal of Remote Sensing, 31, 5449–5464, https://doi.org/10.1080/01431160903369642, 2010.

Irish, R. R., Barker, J. L., Goward, S. N., and Arvidson, T.: Characterization of the Landsat-7 ETM+ Automated Cloud-Cover Assessment (ACCA) Algorithm, Photogrammetric Engineering and Remote Sensing, 72, 1179–1188, https://doi.org/10.14358/PERS.72.10.1179, 2006.

Jain, A. K.: Data clustering: 50 years beyond K-means, Pattern Recognition Letters, 31, 651–666, https://doi.org/10.1016/j.patrec.2009.09.011, 2010.

Ji, L., Zhang, L., and Wylie, B.: Analysis of Dynamic Thresholds for the Normalized Difference Water Index, Photogrammetric Engineering and Remote Sensing, 75, 1307–1317, https://doi.org/10.14358/PERS.75.11.1307, 2009.

Ji, L., Geng, X., Sun, K., Zhao, Y., and Gong, P.: Target Detection Method for Water Mapping Using Landsat 8 OLI/TIRS Imagery, Water, 7, 794–817, https://doi.org/10.3390/w7020794, 2015.

Jiang, H., Feng, M., Zhu, Y., Lu, N., Huang, J., and Xiao, T.: An Automated Method for Extracting Rivers and Lakes from Landsat Imagery, Remote Sensing, 6, 5067–5089, https://doi.org/10.3390/rs6065067, 2014.

Jones, S., Fremier, A., DeClerck, F., Smedley, D., Ortega Pieck, A., and Mulligan, M.: Big Data and Multiple Methods for Mapping Small Reservoirs: Comparing Accuracies for Applications in Agricultural Landscapes, Remote Sensing, 9, 1307, https://doi.org/10.3390/rs9121307, 2017.

Kingumbi, A.: Modélisation hydrologique d'un bassin affecté par des changements d'occupation. Cas du Merguellil en Tunisie Centrale, Ph.D. thesis, Université de Tunis El Manar, Ecole Nationale d'Ingénieurs de Tunis, 2006.

Kingumbi, A., Bargaoui, Z., Ledoux, E., Besbes, M., and Hubert, P.: Modélisation hydrologique stochastique d'un bassin affecté par des changements d'occupation: cas du Merguellil en Tunisie centrale, Hydrological Sciences Journal, 52, 1232–1252, https://doi.org/10.1623/hysj.52.6.1232, 2007.

Klein, I., Dietz, A., Gessner, U., Dech, S., and Kuenzer, C.: Results of the Global WaterPack: a novel product to assess inland water body dynamics on a daily basis, Remote Sensing Letters, 6, 78–87, https://doi.org/10.1080/2150704X.2014.1002945, 2015.

Kuenzer, C., Klein, I., Ullmann, T., Georgiou, E., Baumhauer, R., and Dech, S.: Remote Sensing of River Delta Inundation: Exploiting the Potential of Coarse Spatial Resolution, Temporally-Dense MODIS Time Series, Remote Sensing, 7, 8516–8542, https://doi.org/10.3390/rs70708516, 2015.

Lacaux, J., Tourre, Y., Vignolles, C., Ndione, J., and Lafaye, M.: Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal, Remote Sensing of Environment, 106, 66–74, https://doi.org/10.1016/j.rse.2006.07.012, 2007.

Lacombe, G.: Evolution et usages de la ressource en eau dans un bassin versant amenagé semi-aride. Le cas du Merguellil en Tunisie Centrale, Ph.D. thesis, Université Montpellier II, 2007.

Leauthaud, C., Belaud, G., Duvail, S., Moussa, R., Grünberger, O., and Albergel, J.: Characterizing floods in the poorly gauged wetlands of the Tana River Delta, Kenya, using a water balance model and satellite data, Hydrology and Earth System Sciences, 17, 3059–3075, https://doi.org/10.5194/hess-17-3059-2013, 2013.

Leduc, C., Ben Ammar, S., Favreau, G., Beji, R., Virrion, R., Lacombe, G., Tarhouni, J., Aouadi, C., Zenati Chelli, B., Jebnoun, N., Oi, M., Michelot, J., and Zouari, K.: Impacts of hydrological changes in the Mediterranean zone: environmental modifications and rural development in the Merguellil catchment, central Tunisia, Hydrological Sciences Journal, 52, 1162–1178, https://doi.org/10.1623/hysj.52.6.1162, 2007.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, Frontiers in Ecology and the Environment, 9, 494–502, https://doi.org/10.1890/100125, 2011.

Li, Q. and Gowing, J.: A Daily Water Balance Modelling Approach for Simulating Performance of Tank-Based Irrigation Systems, Water
Resources Management, 19, 211–231, https://doi.org/10.1007/s11269-005-2702-9, 2005.

Liebe, J., van de Giesen, N., and Andreini, M.: Estimation of small reservoir storage capacities in a semi-arid environment, Physics and Chemistry of the Earth, Parts A/B/C, 30, 448–454, https://doi.org/10.1016/j.pce.2005.06.011, 2005.

Lu, D., Mausel, P., Brondizio, E., and Moran, E.: Assessment of atmospheric correction methods for Landsat TM data applicable to Amazon basin LBA research, International Journal of Remote Sensing, 23, 2651–2671, https://doi.org/10.1080/01431160110109642, 2002.

Ma, M., Wang, X., Veroustraete, F., and Dong, L.: Change in area of Ebinur Lake during the 1998-2005 period, International Journal of Remote Sensing, 28, 5523–5533, https://doi.org/10.1080/01431160601009698, 2007.

Mahé, G., Orange, D., Mariko, A., and Bricquet, J. P.: Estimation of the flooded area of the Inner Delta of the River Niger in Mali by hydrological balance and satellite data, IAHS Publication, 344, 138–143, 2011.

Martinez, J.-M. and Le Toan, T.: Mapping of flood dynamics and spatial distribution of vegetation in the Amazon floodplain using multitemporal SAR data, Remote Sensing of Environment, 108, 209–223, https://doi.org/10.1016/j.rse.2006.11.012, 2007.

Massuel, S., Perrin, J., Mascre, C., Mohamed, W., Boisson, A., and Ahmed, S.: Managed aquifer recharge in South India: What to expect from small percolation tanks in hard rock?, Journal of Hydrology, 512, 157–167, https://doi.org/10.1016/j.jhydrol.2014.02.062, 2014.

Maxwell, S.: Filling Landsat ETM+ SLC-off Gaps Using a Segmentation Model Approach, Photogrammetric Engineering and Remote Sensing, 70, 1109–1111, 2004.

McFeeters, S. K.: The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, International Journal of Remote Sensing, 17, 1425–1432, https://doi.org/10.1080/01431169608948714, 1996.

Mialhe, F., Gunnell, Y., and Mering, C.: Synoptic assessment of water resource variability in reservoirs by remote sensing: General approach and application to the runoff harvesting systems of south India, Water Resources Research, 44, 14, https://doi.org/10.1029/2007WR006065, 2008.

Minnaert, M.: The reciprocity principle in lunar photometry, The Astrophysical Journal, 93, 403–410, https://doi.org/10.1086/144279, 1941.

Mohamed, Y., Bastiaanssen, W. G. M., and Savenije, H. H. G.: Spatial variability of evaporation and moisture storage in the swamps of the upper Nile studied by remote sensing techniques, Journal of Hydrology, 289, 145–164, https://doi.org/10.1016/j.jhydrol.2003.11.038, 2004.

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

Nyssen, J., Clymans, W., Descheemaeker, K., Poesen, J., Vandecasteele, I., Vanmaercke, M., Zenebe, A., Van Camp, M., Haile, M., Haregeweyn, N., Moeyersons, J., Martens, K., Gebreyohannes, T., Deckers, J., and Walraevens, K.: Impact of soil and water conservation measures on catchment hydrological response-a case in north Ethiopia, Hydrological Processes, 24, 1880–1895, https://doi.org/10.1002/hyp.7628, 2010.

Ogilvie, A.: Upscaling water availability and water use assessments in hydro-sociosystems: the small reservoirs of the Merguellil catchment (Central Tunisia), Ph.D. thesis, Université de Montpellier and King's College London, 2015.

Ogilvie, A., Belaud, G., Delenne, C., Bailly, J.-S., Bader, J.-C., Oleksiak, A., Ferry, L., and Martin, D.: Decadal monitoring of the Niger Inner Delta flood dynamics using MODIS optical data, Journal of Hydrology, 523, 368–383, https://doi.org/10.1016/j.jhydrol.2015.01.036, 2015.

Ogilvie, A., Belaud, G., Massuel, S., Mulligan, M., Le Goulven, P., Calvez, R., Goulven, P. L., and Calvez, R.: Assessing Floods and Droughts in Ungauged Small Reservoirs with Long-Term Landsat Imagery, Geosciences, 6, 42, https://doi.org/10.3390/GEOSCIENCES6040042, 2016a.

Ogilvie, A., Le Goulven, P., Leduc, C., Calvez, R., and Mulligan, M.: Réponse hydrologique d'un bassin semi-aride aux événements pluviométriques et aménagements de versant (bassin du Merguellil, Tunisie centrale), Hydrological Sciences Journal, 61, 441–453, https://doi.org/10.1080/02626667.2014.934249, 2016b.

Ouma, Y. O. and Tateishi, R.: A water index for rapid mapping of shoreline changes of five East African Rift Valley lakes: an empirical analysis using Landsat TM and ETM+ data, International Journal of Remote Sensing, 27, 3153–3181, https://doi.org/10.1080/01431160500309934, 2006.

Palmer, S. C. J., Kutser, T., and Hunter, P. D.: Remote sensing of inland waters: Challenges, progress and future directions, Remote Sensing of Environment, 157, 1–8, https://doi.org/10.1016/j.rse.2014.09.021, 2015.

Paolini, L., Grings, F., Sobrino, J. A., Jiménez Muñoz, J. C., and Karszenbaum, H.: Radiometric correction effects in Landsat multi-date/multi-sensor change detection studies, International Journal of Remote Sensing, 27, 685–704, https://doi.org/10.1080/01431160500183057, 2006.

Papa, F., Prigent, C., Aires, F., Jimenez, C., Rossow, W. B., and Matthews, E.: Interannual variability of surface water extent at the global scale, 1993–2004, Journal of Geophysical Research, 115, D12 111, https://doi.org/10.1029/2009JD012674, 2010.

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

Prigent, C., Papa, F., Aires, F., Rossow, W. B., and Matthews, E.: Global inundation dynamics inferred from multiple satellite observations, 1993–2000, Journal of Geophysical Research, 112, D12 107, https://doi.org/10.1029/2006JD007847, 2007.

Qi, S., Brown, D. G., Tian, Q., Jiang, L., Zhao, T., and Bergen, K. M.: Inundation Extent and Flood Frequency Mapping Using LANDSAT Imagery and Digital Elevation Models, GIScience and Remote Sensing, 46, 101–127, https://doi.org/10.2747/1548-1603.46.1.101, 2009.

Ran, L. and Lu, X. X.: Delineation of reservoirs using remote sensing and their storage estimate: an example of the Yellow River basin, China, Hydrological Processes, 26, 1215–1229, https://doi.org/10.1002/hyp.8224, 2012.

Rodrigues, L. N., Sano, E. E., Steenhuis, T. S., and Passo, D. P.: Estimation of Small Reservoir Storage Capacities with Remote Sensing in the Brazilian Savannah Region, Water Resources Management, 26, 873–882, https://doi.org/10.1007/s11269-011-9941-8, 2011.

Rouse, J., Haas, J., Schell, J., and Deering, D.: Monitoring vegetation systems in the Great Plains with ERTS, in: Proceedings 3rd ERTS Symposium, pp. 309–317, NASA SP353, Washington DC, USA, 1973.

Sakamoto, T., Van Nguyen, N., Kotera, A., Ohno, H., Ishitsuka, N., and Yokozawa, M.: Detecting temporal changes in the extent of annual flooding within the Cambodia and the Vietnamese Mekong Delta from MODIS time-series imagery, Remote Sensing of Environment, 109, 295–313, https://doi.org/10.1016/j.rse.2007.01.011, 2007.

Sawunyama, T., Senzanje, A., and Mhizha, A.: Estimation of small reservoir storage capacities in Limpopo River Basin using geographical information systems (GIS) and remotely sensed surface areas: Case of Mzingwane catchment, Physics and Chemistry of the Earth, Parts A/B/C, 31, 935–943, https://doi.org/10.1016/j.pce.2006.08.008, 2006.

Scaramuzza, P., Micijevic, E., and Chander, G.: SLC gap-filled products phase one methodology, Landsat Technical Notes, https://landsat. usgs.gov/documents/SLC_Gap_Fill_Methodology.pdf, 2004.

Seiler, R., Schmidt, J., Diallo, O., and Csaplovics, E.: Flood monitoring in a semi-arid environment using spatially high resolution radar and optical data, Journal of Environmental Management, 90, 2121–9, https://doi.org/10.1016/j.jenvman.2007.07.035, 2009.

5 Solander, K. C., Reager, J. T., and Famiglietti, J. S.: How well will the Surface Water and Ocean Topography (SWOT) mission observe global reservoirs?, Water Resources Research, 52, 2123–2140, https://doi.org/10.1002/2015WR017952, 2016.

Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., and Macomber, S. A.: Classification and Change Detection Using Landsat TM Data, Remote Sensing of Environment, 75, 230–244, https://doi.org/10.1016/S0034-4257(00)00169-3, 2001.

Soti, V., Puech, C., Lo Seen, D., Bertran, A., Vignolles, C., Mondet, B., Dessay, N., and Tran, A.: The potential for remote sensing and
10 hydrologic modelling to assess the spatio-temporal dynamics of ponds in the Ferlo Region (Senegal), Hydrology and Earth System Sciences, 14, 1449–1464, https://doi.org/10.5194/hess-14-1449-2010, 2010.

Sun, W., Du, B., and Xiong, S.: Quantifying Sub-Pixel Surface Water Coverage in Urban Environments Using Low-Albedo Fraction from Landsat Imagery, Remote Sensing, 9, 428, https://doi.org/10.3390/rs9050428, 2017.

Swenson, S. and Wahr, J.: Monitoring the water balance of Lake Victoria, East Africa, from space, Journal of Hydrology, 370, 163–176,
15 https://doi.org/10.1016/j.jhydrol.2009.03.008, 2009.

Talineau, J., Selmi, S., and Alaya, K.: Lacs collinaires en Tunisie semi-aride, Sécheresse, 5, 251–256, 1994.

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

20 Vala, H. J. and Baxi, A.: A review on Otsu image segmentation algorithm, International Journal of Advanced Research in Computer Engi-neering and Technology, 2, 387–389, 2013.

Van Der Meer, F.: Spectral unmixing of Landsat Thematic Mapper data, International Journal of Remote Sensing, 16, 3189–3194, https://doi.org/10.1080/01431169508954622, 1995.

Vanonckelen, S., Lhermitte, S., and Van Rompaey, A.: The effect of atmospheric and topographic correction methods
25 on land cover classification accuracy, International Journal of Applied Earth Observation and Geoinformation, 24, 9–21, https://doi.org/10.1016/j.jag.2013.02.003, 2013.

Venot, J. P. and Krishnan, J.: Discursive framing: Debates over small reservoirs in the Rural South, Water Alternatives, 4, 316–324, 2011.

Verpoorter, C., Kutser, T., Seekell, D. A., and Tranvik, L. J.: A Global Inventory of Lakes Based on High-Resolution Satellite Imagery, Geophysical Research Letters, 41, 6396–6402, https://doi.org/10.1002/2014GL060641, 2014.

30 Vogelmann, J. E., Helder, D., Morfitt, R., Choate, M. J., Merchant, J. W., and Bulley, H.: Effects of Landsat 5 Thematic Mapper and Landsat 7 Enhanced Thematic Mapper plus radiometric and geometric calibrations and corrections on landscape characterization, Remote Sensing of Environment, 78, 55–70, https://doi.org/10.1016/S0034-4257(01)00249-8, 2001.

Wisser, D., Frolking, S., Douglas, E. M., Fekete, B. M., Schumann, A. H., and Vörösmarty, C. J.: The significance of local wa-ter resources captured in small reservoirs for crop production - A global-scale analysis, Journal of Hydrology, 384, 264–275,
35 https://doi.org/10.1016/j.jhydrol.2009.07.032, 2010.

Wolski, P. and Murray-Hudson, M.: An investigation of permanent and transient changes in flood distribution and outflows in the Okavango Delta, Botswana, Physics and Chemistry of the Earth, Parts A/B/C, 33, 157–164, https://doi.org/10.1016/j.pce.2007.04.008, 2008.

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

Yamazaki, D. and Trigg, M. A.: Hydrology: The dynamics of Earth's surface water, Nature, 540, 348–349, https://doi.org/10.1038/nature21100, 2016.

5   Yamazaki, D., Trigg, M. A., and Ikeshima, D.: Development of a global ∼ 90 m water body map using multi-temporal Landsat images, Remote Sensing of Environment, 171, 337–351, https://doi.org/10.1016/j.rse.2015.10.014, 2015.

Yin, G., Mariethoz, G., and McCabe, M.: Gap-Filling of Landsat 7 Imagery Using the Direct Sampling Method, Remote Sensing, 9, 12, https://doi.org/10.3390/rs9010012, 2016.

Zeng, C., Shen, H., and Zhang, L.: Recovering missing pixels for Landsat ETM+ SLC-off imagery using multi-temporal regression analysis
10     and a regularization method, Remote Sensing of Environment, 131, 182–194, https://doi.org/10.1016/j.rse.2012.12.012, 2013.

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

Zhu, Z. and Woodcock, C. E.: Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change, Remote Sensing of Environment, 152, 217–234, https://doi.org/10.1016/j.rse.2014.06.012,
15     2014.