Improved Remotely Sensed Satellite Products for Studying Lake Victoria’s Water Storage Changes

M. Khaki\textsuperscript{a,1}, J. Awange\textsuperscript{b}

\textsuperscript{a}School of Engineering, University of Newcastle, Callaghan, New South Wales, Australia.
\textsuperscript{b}School of Earth and Planetary Sciences, Discipline of Spatial Sciences, Curtin University, Perth, Australia.

Abstract

Lake Victoria (LV), the world’s second largest freshwater lake, supports a livelihood of more than 42 million people and modulates the regional climate. Studying its changes resulting from impacts of climate variation/change and anthropogenic is, therefore, vital for its sustainable use. Owing to its sheer size, however, it is a daunting task to undertake such study relying solely on in-situ measurements, which are sparse, either missing, inconsistent or restricted by governmental red tapes. Remotely sensed products provide a valuable alternative but come with a penalty of being mostly incoherent with each other as they originate from different sources, have different underlying assumptions and models. This study pioneers a procedure that uses a Simple Weighting approach to merge LV’s multi-mission satellite precipitation and evaporation data from various sources and then improves them through a Postprocessing Filtering (PF) scheme to provide coherent datasets of precipitation (\(p\)), evaporation (\(e\)), water storage changes (\(\Delta s\)), and discharge (\(q\)) that accounts for its water budget closure. Principal component analysis (PCA) is then applied to the merged-improved products to analyze LV’s spatio-temporal changes resulting from impacts of climate variation/change. Compared to the original unmerged data (0.62 and 0.37 average correlation for two samples), the merged-improved products are largely in agreement (0.91 average correlation). Furthermore, smaller imbalances between the merged-improved products are obtained with precipitation (37\%) and water storage changes (35\%) being the largest contributors to LV’s water budget. This data improvement scheme could be applicable to any inland lake of a size similar to LV.

Keywords: Remote sensing, Lake Victoria, Water balance, Water storage variation, Hydroclimate
1. Introduction

Lake Victoria, spanning an area of 68,800 km$^2$ with a basin size of more than 252,000 km$^2$ (e.g., Onganga and Awange, 2005), is the largest lake in the developing world, and the world’s second largest freshwater lake after Lake Superior in the US. The lake, which is shared by Kenya, Tanzania, and Uganda directly supports the livelihood of more than 42 million people, with the population projected to triple by 2050, (see e.g., Okungu et al., 2005; Bremner et al., 2013). Furthermore, being the source of the White Nile, i.e., one of the main streams of the Nile river, the lake supports the livelihood of Egypt, Sudan and South Sudan (e.g., Awange et al., 2014). Moreover, it is known to modulate the regional climate (e.g., Nicholson et al., 2003; Awange et al., 2013a). Any significant change in the lake water storage, triggered e.g., by climatic impacts (e.g., Yin and Nicholson, 1998; Conway, 2002; Omondi et al., 2014) or anthropogenic factors such as dam expansion (e.g., Consulate, 2012; Aman, 2014), therefore, is likely to affect millions of people who directly depend on it for livelihood plus others the world over who indirectly depend on it. Therefore, it is essential to continuously monitor its behaviour in terms of water storage changes and effective climate parameters as undertaken, e.g., by Awange et al. (2008) and Swenson and Wahr (2009) among others.

Its monitoring has often taken on various forms, e.g., use of ground-based in-situ measurements (e.g., Njuru, 2002; Mugidde et al., 2003), land surface models (LSM; e.g., Khan et al., 2011; Chamberlain et al., 2014), and satellite remote sensing products (see, e.g., Piper et al., 1986; Woodward and Warren, 2007; Song et al., 2015; Crataux et al., 2016; Sichangi and Makokha, 2017; Anyah et al., 2018). Due to its wide area and limited number of in-situ stations (e.g., rain and water level gauges), monitoring of the lake solely based on ground-based measurements, i.e., “boots on the ground”, becomes practically impossible. Moreover, there is no reliable regional land hydrological model in the area for undertaking such monitoring. Satellite remote sensing, on the other hand, due to their vast coverage, high spatio-temporal resolutions, and easier access provide better tools for analyzing the hydroclimate variations within Lake Victoria Basin (LVB).

A number of literature have studied Lake Victoria using various satellite remotely sensed products, e.g., satellite radar altimetry to observe the lake’s water level variations (e.g., Awange et al., 2013b; Uebbing et al., 2015; Sichangi and Makokha, 2017) and their importance for flood monitoring (Birkett et al., 1999), the Gravity Recovery and Climate Experiment (GRACE) for studying the lake’s total water storage (TWS) changes (e.g., Awange et al., 2008, 2014;
Hassan and Jin, 2004), satellite precipitation data for studying the lake’s rainfall (e.g., Kizza et al., 2009; Awange et al., 2013b), a combination of both ground-based and remotely sensed observations for studying the lake’s water balance (e.g., Yin and Nicholson, 1998; Swenson and Wahr, 2009). Despite this plethora of studies, a precise study of the hydrological processes of Lake Victoria using merged and improved coherent datasets from multiple sources, what would also benefit other inland lake waters the world over, is still missing. For example, although Khaki et al. (2018a) used a multi-mission satellite data to study various water storage including surface and sub-surface water components over the Nile basin, their study does not account for the discrepancy between different datasets from various sources. It is also important to further study the water storage changes within the water balance equation to analyze the interrelationship between the different water components.

Due to the fast emerging satellite platforms, especially in the last two decades, there are different data sources for various data types, making the extraction of the most reliable datasets from the available products, e.g., a merged and improved rainfall data from various precipitation sources, a necessity for providing improved datasets. Moreover, the balance between different data types (i.e., precipitation, evaporation, water storage changes, and water discharge) that is normally addressed using the water balance equation stands to benefit from using such merged and improved datasets. Traditionally, hydrological model and data assimilation are used to establish the balance between different components (e.g., Pan and Wood, 2006; Sahoo et al., 2011; Pan et al., 2012; Khaki et al., 2017a, 2018b). Here, however, in the absence of an accurate model over the LVB, use is made of a data combination strategy to obtain a coherent data set of four water cycle components, i.e., precipitation ($p$), evaporation ($e$), water storage changes ($\Delta s$), and discharge ($q$). This could enable one to accurately analyze the lake’s hydrology and the associated climatic variation/change impacts. The main objectives of the present contribution, therefore, are (i) generate improved coherent water cycle components ($p$, $e$, $\Delta s$ and $q$) from different sources over lake Victoria, (ii) explore the changes in the lake’s water storage and its water level using these improved coherent datasets, and (iii) investigate climatic impacts on the lake’s water storage changes based on the improved datasets in (i). To achieve these goals, this study aims at using a proposed two-step filtering step by Aires (2014). The filter applies a Simple Weighted (SW) approach to merge remotely sensed products over Lake Victoria from multi-mission satellites and filter them employing a Postprocessing Filter (PF) to generate improved coherent products of precipitation ($p$), evaporation ($e$), water storage
changes ($\Delta s$), and discharge ($q$) by accounting for balance between them. These improved products are useful in not only analysing changes of the lake as a consequence of climate and anthropogenic impacts but also in correcting for the imbalance in the components of the water balance model ($\Delta s = p - e - q$). This procedure for improving remotely sensed data could potentially be applied to any inland lake of basin scale around the world.

Multi-mission satellite and ground-based products used include three precipitation products of Tropical Rainfall Measuring Mission Project (TRMM; TRMM, 2011), Global Precipitation Climatology Centre (GPCC; Schneider et al., 2008), and Climate Prediction Center (CPC) unified gauge dataset (Chen et al., 2002). For evaporation, three data products of MODIS (Moderate Resolution Imaging Spectroradiometer) Global Evapotranspiration Project (MOD16; Mu et al., 2007), Global Land Evaporation Amsterdam Model (GLEAM; Miralles et al., 2011), and ERA-interim (Simmons et al., 2007) are employed. Water storage changes from GRACE and water discharge time series from two ground stations (Jinja and Entebbe) are also used. In addition, satellite altimetry data from TOPEX/Poseidon (T/P) and its follow-on missions Jason-1 and -2, as well as ENVISAT (Environmental Satellite) are applied for the analysis of surface water variations. The altimetry data are used to build virtual stations covering the period from 1992 to 2016 over Lake Victoria (Figure 1). Lake level variations at virtual stations, and associated precipitation and TWS time series are used to analyze the Lake Victoria behaviors during the study period. To improve on the satellite altimetry range estimations, which are erroneous when used over an inland body of waters and rivers (Birkett et al., 1999; Calmant et al., 2008; Khaki et al., 2015), the Extrema Retracking (ExtR) algorithm of Khaki et al. (2014) is employed to retrack satellite waveform data. Furthermore, principal component analysis (PCA Lorenz, 1956; Preisendorfer, 1988) is used to better investigate spatio-temporal variations of Lake Victoria water storage and its relationship to climatic impacts.

The remainder of the study is organised as follows. In Section 2 datasets are presented while Section 3 provides the method. The results and discussion are presented in Section 4, and the study concluded in Section 5.

2. Data set

2.1. GRACE

Monthly GRACE level 2 (L2) potential coefficients products up to degree and order (d/o) 90 are obtained for the period 2002-2016 from the ITSG-Grace2014 gravity field model.
(Mayer-Gürr et al., 2014) to estimate TWS changes (see also Khaki et al., 2017b,c). Lower spherical harmonic degrees components are replaced with more accurate estimates of Swenson et al. (2008) (degree 1 coefficients) and Cheng and Tapley (2004) (Degree 2 and order 0 coefficients). The L2 gravity fields are then converted into $1^\circ \times 1^\circ$ TWS fields, see (Wahr et al., 1998). Colored/correlated noises in the products are reduced using the Kernel Fourier Integration (KeFIn) filter proposed by Khaki et al. (2018c), which also accounts for signal attenuations and leakage effects caused by smoothing. The KeFIn filter works through a two-step post-processing algorithm. The first step mitigates the measurement noise and the aliasing of unmodelled high-frequency mass variations, while the second step contains an efficient kernel to decrease the leakage errors. Details of this filter can be found in Khaki et al. (2018c).

2.2. Precipitation, Evaporation, and Discharge

Precipitation datasets from the Tropical Rainfall Measuring Mission Project (TRMM-3B43 version 7) products (TRMM, 2011), Global Precipitation Climatology Centre (GPCC; Schneider et al., 2008), and CPC unified gauge dataset (Chen et al., 2002) covering the period from 1998 to 2016 at monthly $1^\circ \times 1^\circ$ spatial resolution are employed.

Evaporation datasets are acquired from MODIS Global Evapotranspiration Project (MOD16;
Mu et al., 2007), Global Land Evaporation Amsterdam Model (GLEAM; Miralles et al., 2011), and ERA-interim (Simmons et al., 2007) for the same temporal period. These evaporation products are then converted to monthly 1°×1° spatial resolution similar to those of precipitation.

Water discharge time series are obtained from (i) the Jinja and Entebbe stations in Uganda, and (ii), from different sources including the Global Runoff Data Centre (GRDC; http://www.bafg.de/), a report produced by Power Planning Associates (PPA, 2007), and River Watch project (http://floodobservatory.colorado.edu/). Similar to precipitation and evaporation datasets, all these discharge products are resampled to monthly average height variations. Figure 1 shows the locations of discharge stations.

2.3. Satellite Radar Altimetry

TOPEX/Poseidon (T/P), Jason-1, and Jason-2 data (~9.915 day temporal resolution) of the Sensor Geographic Data Records (SGDR), which contains 20-Hz waveform data as well as ENVISAT 18-Hz SGDR product (35-day temporal resolution) from RA-2/MWR are used. This includes 360 cycles of T/P covering 1992–2002, 260 cycles of Jason-1 from 2002 to 2008, 277 cycles of Jason-2 covering 2008 to 2016, and 112 cycles of ENVISAT. T/P and Jason-1 data are both derived from the Physical Oceanography Distributed Active Archive Center (PO.DAAC), Jason-2 data is provided by AVISO, and ENVISAT data is obtained from European Space Agency (ESA). Before using these datasets geophysical corrections that include solid earth tide, pole tide, and dry tropospheric (Birkett, 1995) are applied. The data sets are then converted to a monthly scale and used to build virtual time series over different points (see Figure 1) located on the satellite ground tracks over Lake Victoria. At each virtual point, several points belonging to the same satellite cycle are considered, and the median value of the retracked altimetry-based water levels computed to address the hooking effects (Frappart et al., 2006). This effect is derived from off-nadir measurements when a satellite locks over a water body before or after passing above it (Seyler et al., 2008; Boergens et al., 2016). A summary of the datasets used in the present study is presented in Table 1.
Table 1: A summary of the datasets used in this study.

<table>
<thead>
<tr>
<th>Product</th>
<th>Platform</th>
<th>Reference</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terrestrial water storage (TWS)</td>
<td>GRACE</td>
<td>Mayer-Gürr et al. (2014)</td>
<td>Satellite remote sensing</td>
</tr>
<tr>
<td>Precipitation</td>
<td>TRMM-3B43</td>
<td>TRMM (2011)</td>
<td>Satellite remote sensing</td>
</tr>
<tr>
<td>Precipitation</td>
<td>GPCC</td>
<td>Schneider et al. (2008)</td>
<td>Based on in-situ data</td>
</tr>
<tr>
<td>Precipitation</td>
<td>CPC</td>
<td>Chen et al. (2002)</td>
<td>Gauge-based</td>
</tr>
<tr>
<td>Evapotranspiration</td>
<td>MOD16</td>
<td>Mu et al. (2007)</td>
<td>Satellite remote sensing</td>
</tr>
<tr>
<td>Evapotranspiration</td>
<td>GLEAM</td>
<td>Miralles et al. (2011)</td>
<td>Model and satellite remote sensing</td>
</tr>
<tr>
<td>Evapotranspiration</td>
<td>ERA-interim</td>
<td>Simmons et al. (2007)</td>
<td>Reanalysis dataset</td>
</tr>
<tr>
<td>Water discharge</td>
<td>Jinja and Entebbe stations</td>
<td>Ministry of Energy &amp; Mineral Development Kampala (Uganda)</td>
<td>In-situ</td>
</tr>
<tr>
<td>Water discharge</td>
<td>GRDC</td>
<td><a href="http://www.bafg.de/GRDC/EN/Home/homepage_node.html">http://www.bafg.de/GRDC/EN/Home/homepage_node.html</a></td>
<td>In-situ</td>
</tr>
<tr>
<td>Water discharge</td>
<td>PPA</td>
<td>Power Planning Associates (PPA, 2007)</td>
<td>In-situ</td>
</tr>
<tr>
<td>Water discharge</td>
<td>River Watch</td>
<td><a href="http://floodobservatory.colorado.edu/">http://floodobservatory.colorado.edu/</a></td>
<td>In-situ</td>
</tr>
</tbody>
</table>

3. Method

3.1. Data integration

A two-step data combination approach proposed by Aires (2014) is applied, where first, a Simple Weighting (SW) approach is employed to merge different precipitation and evaporation data sets leading to new merged (precipitation and evaporation) products. These merged precipitation and evaporation products together with those of GRACE TWS and discharge are passed through a Postprocessing Filtering (PF) procedure to generate improved precipitation \( (p) \), evaporation \( (e) \), water storage changes \( (\Delta s) \), and discharge \( (q) \) that accounts for water budget closure. Compared to other techniques, the SW approach has been shown to
perform better e.g., Aires (2014). During the SW step, the filter assigns a weight to each water component (e.g., for each precipitation and evaporation product). The PF step that is based on the water balance equation then checks the water budget closure (Eqs. 1 and 2) using a Kalman-based scheme.

\[ p - \Delta s - e - q = 0, \]  
\[ X^T \cdot G = 0, \]

with \( X^T = (p, \Delta s, e, q) \) (\( T \) indicates matrix transpose), calculated from the first step (SW), and \( G^T = (1, -1, -1, -1) \). A Kalman-like solution (following Pan and Wood, 2006) is then applied by,

\[ X_a = K \cdot X, \]
\[ K = (I - BG(G^T BG)^{-1} G^T), \]

where \( I \) is identity matrix and \( B \) is the error covariance matrix of \( X \) from SW. The application of PF guarantees that the estimated flux nets in \( X_a \) are balanced (see details in Aires, 2014; Munier et al., 2015).

3.2. Extrema Retracking (ExtR)

Satellite radar altimetry, originally designed to monitor sea level changes, nowadays are also used to monitor inland water bodies (see, e.g., Birkett, 1995) and rivers (see e.g., Birkett et al., 2002; Berry et al., 2005; Tseng et al., 2013). Nevertheless, the waveform retracking, which refers to the re-analysis of the waveforms, a time-series of returned power in the satellite antenna (Davis et al., 1995; Gomez-Enri et al., 2009), is required to improve the accuracy of measured ranges (or sea surface height; SSH) over inland waters (Brown, 1977). Here, to retrack satellite radar altimetry data, a developed Extrema Retracking (ExtR) algorithm proposed by Khaki et al. (2014) is applied to generate refined virtual lake level heights that are used in the water storage analysis step. It should be pointed out that this water analysis step uses improved water storage changes obtained from the SW and PF steps (see Figure 3). Our motivation for selecting the ExtR algorithm is due to its processing speed and its promising results that were obtained over the Caspian Sea when compared to the Off Center of Gravity (OCOG, Wingham et al., 1986), the NASA \( \beta \)-Parameter Retracking (Martin et al., 1983), and
Threshold Retracking (Davis, 1997). The ExtR is applied to the altimetry-derived waveforms to retrack datasets, which is necessary for inland applications of satellite radar altimetry. The Khaki et al. (2014) algorithm includes three steps; (i) applying a moving average filter to reduce the random noise of the waveforms, (ii) identifying extremum points of the filtered waveforms, and (iii), exploring the leading edges among all detected extremum points. Range corrections are applied using the offset between the position of the leading edges and their on-board values.

Two gauge stations around Lake Victoria (see Figure 1) located at Jinja (1992-1995) and Entebbe (1992-2009) from the Ministry of Energy & Mineral Development Kampala (Uganda), Old Aswan (1996-2009), Esna Barrage (1996-2009), and Naga Hammadi Barrage (1996-2007), and Assiut Barrage (1996-2009) from Ismail and Samuel (2011), and Nubaria (1997-2007) from Samuel (2014) are used to examine the performance of the ExtR filter. Retracked time series of two closest virtual stations to the in-situ stations are compared to the in-situ water level measurements. The average bias and standard deviation (STD) of average errors for both stations are presented in Figure 2. Significant decreases in both bias and STD are found after applying the filtering process (cf. Figure 2). The figure shows the capability of the ExtR filter for reducing errors and justifying its usage in this study.

3.2.1. Climate Variability Impacts

In order to investigate the impacts of climate variability/change on LV’s water storage changes, correlation analysis is used. Hereafter, reported correlation values between any two variables are calculated as the average of correlation between their time series at all grid points. Furthermore, principal component analysis (PCA; Lorenz, 1956; Preisendorfer, 1988) is applied on the improved precipitation, evaporation, and water storage time series (after filtering in Section Section 3.1) to better analyze the spatio-temporal changes of water storages and climatic indicators. This is done to examine the climate patterns within the LV area and to investigate their connections to water storage changes. Since precipitation and evaporation are the major effective parameters on water storage recharge, the process helps to study the role of climate variability on water storage variations. A schematic illustration of the applied processing steps in this study, i.e., data integration procedure, retracking, and climatic impacts exploration, is provided in Figure 3.
4. Results and Discussion

First, we present the two-step (SW+PF) filtering results and the impact of the process on individual data type, as well as on the balance between them. Afterwards, spatio-temporal variations of precipitation and TWS and their interactions as major components of water fluxes are investigated.

4.1. Coherent filtered products

To demonstrate the usefulness of the SW+PF for filtering remotely sensed datasets for Lake Victoria, correlations between the original and filtered time series of each water cycle component and other filtered components are calculated in order to assess how the filtering process increases the agreement between different water cycle components. Table 2 shows the average improvements in the correlations between every two products, i.e., how each estimated water flux is correlated with other flux observations. Note that this is done to assess the effect of the applied method to produce a more coherent data, which does not necessarily lead to a higher accuracy. Correlation values between the original and filtered water fluxes, e.g., original
Figure 3: A schematic illustration of the applied methodology. The algorithm operates in two steps. First, the SW approach is employed to merge the precipitation and evaporation data sets. The merged products together with those of GRACE TWS and discharge are then subjected to the PF filter in the second step to produce improved water budget parameters.

and filtered \( p \) and all data products of \( e, \Delta s, \) and \( q \) are calculated to allow for estimation of achieved improvement in the filtered data. It is evident that in all the cases, improvements are achieved between any two filtered datasets. For example, between the original GPCC products and filtered precipitation time series, the later is 13.48% more correlated to filtered water storage changes. It can also be seen that the obtained improvements are different for various products. In general, for precipitation, higher increase in correlation is achieved from GPCC while less improvements are found in CPC, which is gauge-based. A similar correlation improvement can also be seen for evaporation, where different products (e.g., MOD16, GLEAM, and ERA-interim) receive various weights in the process, which correspondingly lead to various
levels of improvements. Based on the results in Table 2, it can be concluded that the filtered
p, e, ∆s, and q products are largely in agreements.

Table 2: Average correlation improvements (%) between different variables. Note that p, e, ∆s, and q refer
to the filtered water cycle components of precipitation, evaporation, water storage changes, and discharge.
Improvements in the correlation (r) values are calculated as \( \left( \frac{r_{\text{filtered results}} - r_{\text{original data}}}{r_{\text{original data}}} \right) \times 100(\%) \).

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>e</th>
<th>∆s</th>
<th>q</th>
</tr>
</thead>
<tbody>
<tr>
<td>p (compared to TRMM-3B43)</td>
<td>0</td>
<td>8.51</td>
<td>12.93</td>
<td>9.81</td>
</tr>
<tr>
<td>p (compared to GPCC)</td>
<td>0</td>
<td>13.48</td>
<td>16.75</td>
<td>11.32</td>
</tr>
<tr>
<td>p (compared to CPC)</td>
<td>0</td>
<td>2.73</td>
<td>6.14</td>
<td>5.73</td>
</tr>
<tr>
<td>e (compared to MOD16)</td>
<td>9.50</td>
<td>0</td>
<td>11.65</td>
<td>6.69</td>
</tr>
<tr>
<td>e (compared to GLEAM)</td>
<td>8.36</td>
<td>0</td>
<td>8.01</td>
<td>7.17</td>
</tr>
<tr>
<td>e (compared to ERA-interim)</td>
<td>18.18</td>
<td>0</td>
<td>14.44</td>
<td>9.20</td>
</tr>
<tr>
<td>∆s (compared to GRACE TWS)</td>
<td>16.56</td>
<td>12.94</td>
<td>0</td>
<td>15.58</td>
</tr>
<tr>
<td>q (compared to initial discharge)</td>
<td>11.11</td>
<td>07.08</td>
<td>9.52</td>
<td>0</td>
</tr>
</tbody>
</table>

In addition to correlation improvements above, based on water balance equation, the SW+PF
filtering algorithms also corrects for imbalance between the water cycle components. The post-
processed water fluxes from the application of the SW+PF filter is displayed in Figure 4. Four
filtered water budget components of precipitation, evaporation, TWS changes and discharge
show different performance in the water balance equation (Figure 4a). Precipitation and water
storage changes are seen to have the largest contributions, i.e., 37% and 35%, respectively.
Evaporation shows 19% contribution while runoff depicts the least contribution of 9%. The
large value of the evaporation contribution, corroborated also by the findings of (Mohamed
et al., 2005), indicates that a big part of precipitation over the lake area cannot recharge the
outflow river (e.g., White Nile).

Figure 4b illustrates the imbalance error before and after using the SW+PF filtering, show-
casing the capability of the filters to reduce the imbalance between water cycle components in
order to provide a more coherent data sets. This is also evident from the correlations between
altimetry level variations and the fluxes before and after filtering. The average correlation
improvement of 12% is obtained between lake height variations and all the four components
after applying the SW+PF filtering algorithm. In spite of this improvement after applying the
filter, the imbalance between component can still be seen. This can be attributed to various
factors such as observation errors, the contribution of groundwater in- and out-flows and its
interaction with surface storage, and also the impact of extreme climatic impacts, which can be
underestimated in reanalysis and remote sensing observations contrary to the in-situ discharge
measurements. In what follows, these improved (merged and filtered) products are used to analyze trends in Lake Victoria’s water in the face of climate variation/change impacts.

4.2. Climatic variation/change impacts

To investigate the impacts of climatic variation/change on the Lake Victoria water storage, first a comparison of the average variations of precipitation, evaporation, and water storage changes within the lake is made. Figure 5 shows the average time series of three data types from various sources including filtered and unfiltered products over the entire LVB. Note that 6 months running mean is applied to filter out high-frequency variations leading to better representations of variations and trends. As can be clearly seen, the filtered results are in a large agreement (0.91 average correlation between each two time series) compared to the two other samples from original (unfiltered) datasets, thus indicating the capability of the applied filtering method for achieving coherent data. Water storage changes largely follow the precipitation and evaporation patterns in Figure 5 top panel. This shows that climate is the most effective factor in the lake’s water storage changes. As expected, there is also a large agreement between precipitation and water discharge, especially after the filtering process. This agreement is better discussed by comparing Figures 5 top and bottom panels, in which
the original datasets are plotted. In addition to the time series’ patterns, it can also be seen that the filtering approach affects the time series’ magnitudes. For example, the magnitude of precipitation after the filtering is different from the original data in Figure 5 middle and bottom panels. From Figure 5, several significant positive and negative variations can be seen. Increases in rainfall in 2005, 2007, 2013, and 2016 cause similar rise in water storages. On the contrary, declines are observed in all time series during 2006 and 2014. Furthermore, it can be seen in Figure 5 (bottom and middle panels) that a larger discrepancy exists between precipitation and two other flux observations. Due to the large evaporation rate over Lake Victoria, this larger interaction with water storage is expected. Such a connection, which can be absent on other water bodies depend on their characteristics, and can affect the water flux covariance matrix and violate the assumption of independence observation made on water budget closure. Here, however, the impact of this partial dependency between water storage and evaporation is neglected mainly due to the fact that no information is available in this regard.

To better understand water storage changes and the associated climatic impacts over the lake, PCA method is employed to the merged filtered precipitation, evaporation, and water storage changes. Figure 6 shows the spatial variations of these datasets within LVB corresponding to the first three empirical orthogonal functions (EOFs) of the PCA analysis. The spatial variability of water storage changes matches those of precipitation and evaporation in most of the areas proving that they are the dominant indicator of the impacts of climate indicators on LVB water storage. It can be seen that the main rainfall pattern exists in the central (EOF1) parts of the lake corresponding to a similar pattern in evaporation and water storage. The main water storage patterns, as expected, are observed in the central parts of the Lake Victoria (EOF1) as a result of recharge from rainfall in this area. EOF2 shows considerable positive signals in the western parts. This could be attributed to the contribution of Kagera river to the lake’s water changes. Kagera river, which originates from Burundi, is the largest inflow into Lake Victoria. To a lesser degree, larger rainfall, evaporation and water storage can be observed in the eastern parts (EOF3) of the lake resulting from the effect of the south east monsoon trade winds (e.g., Awange et al., 2008, 2013a) and possible recharge from the Grummeti, Simivu, and Mara rivers. Major rainfall spatial variabilities in the central (EOF1) and western (EOF2) parts corroborate the findings in (Awange et al., 2008, 2013a), which shows that these parts are responsible for most of the rainfalls occurring over Lake Victoria and its water recharge thereby resulting in a larger evaporation and water storage changes.
Figure 5: Average (6 months running mean) precipitation (red), evaporation (green), water discharge (black), and water storage variations (blue) from different sources including filtered products (top), original TRMM, MOD16, and GRACE (middle), and original GPCC, GLEAM, and GRACE (bottom). Note that normalized values (based on the time series STD) are represented for a better visual comparison. Note that precipitation and evaporation data in the middle and bottom panels are selected randomly to show how different they can act.

Figure 7 shows the corresponding first three principal components (PCs) time series. The dominant seasonal (PC1) and annual (PC2) rainfall patterns can be seen, which are in agreements with those of evaporation and water storage changes. Some significant anomalies can also be observed in precipitation, e.g., the large negative anomalies observed in 2006 (PC2) and 2014 (PC1 and PC3), and significant positive variations observed, e.g., in 2005 and 2007 (PC1, PC2, and PC3), 2010 (PC1), 2013 (PC1 and PC2), and 2016 (PC1). These variations can also be seen in evaporation and water storage changes time series, especially the rises in 2005, 2007, and 2013. The 2007 ENSO rainfall effect (e.g., Omondi et al., 2013; Awange et al., 2014) is evident in PC1 for all three datasets. A negative trend is found between 2003 and 2005 for water storage time series similar to those of evaporation (PC1). This could be attributed to excessive water usages reported in the works of Awange et al. (2008) and Swenson and Wahr (2009) while such a negative trend is absent in rainfall time series. Large rainfalls, generally
after 2013 result in positive water storage and evaporation trends. A decrease in rainfall is captured after 2011 due to the drought that affected the region (see, e.g., Awange et al., 2013b), which causes a decline in water storage variations. These similar patterns suggest a close tie between water storage variabilities and climatic impacts.

These similar patterns in climatic indicators and water storage changes, in terms of spatial (cf. Figure 6) and temporal (cf. Figure 7) variabilities, suggest a close tie between water storage variabilities and climatic impacts. This can also be seen in Figures 8 and 9, which show annual spatial variations and trends, and temporal variations of precipitation and water storage over the area, respectively.

It can be seen in Figure 8 that the major variations exist in the south-eastern parts, for both precipitation and water storage. A similar pattern is found for trends (right panels in Figure 8). This means that precipitation plays the main role in the lake’s water storage variations as already reported in other studies (e.g., Nicholson et al., 2003). One can see the same effects from the time series in Figure 9, where the agreement between rainfall changes (top panel) and the lake’s water storage changes (bottom panel) emphasizes the impact of rainfall on Lake Victoria.
Figure 7: The first six principal components (PCs) time series of precipitation, evaporation, and water storage changes from PCA. While the seasonal pattern (PC1) and annual (PC2) are dominant, several considerable positive and negative variations (e.g., in 2005 and 2007 (PC1 and PC2), and 2014 (PC3)) can also be seen in PCs.

Above-average rainfalls during 2007 El’niño significantly affected water storage variations corresponding to the large anomalies. Nevertheless, as previously mentioned, the large contribution of evaporation does not allow water storage changes to perfectly match precipitation time series variations, e.g., in 2012 and after 2014. The impact of climate variability influences the balance between precipitation and evaporation and subsequently impacts the lake’s depth and arguably its areal extent and correspondingly water storage changes (e.g., Owor et al., 2011). Another effective factor on water storage changes is groundwater within the area, which has been under larger influences by the growing population in recent years. Nevertheless, there is not much strong evidence of interaction between groundwater and surface water mainly due to the lack of ground-based groundwater measurements.

To better study Lake Victoria’s stored water changes, analyzing its surface water variations is essential. To this end, altimetry-derived surface water changes are plotted in Figure 10 and compared with TWS changes from GRACE. The water level height variations in Figure 10 depicts a large agreement to water storage variations. A large negative trend is found between
Figure 8: Average annual variations and trends of precipitation and the filtered water storage time series at each grid point.

Figure 9: Spatially averaged time series of precipitation and the filtered water storage changes within the Lake Victoria.
1998 and 2006 before a remarkable positive anomaly due to the effects of 2007 ENSO. The lake’s water level variation also closely follows rainfall pattern. Larger rainfalls before 1998 result in a water level increase in the same period, which ends with a remarkable positive anomaly in 1998 due to an excessive ENSO rainfall in 1997. Decreases in water level are also observed for the period of 2002 to 2004 and after 2007 similar to water storage variations. This large agreement between GRACE-derived TWS variations and altimetry-derived surface water changes suggests that the impact of groundwater and its interaction with surface storage is minimal. This justifies the common assumption that has been made by a number of previous studies (see, e.g., Kite, 1982; Sene and Plintson, 1994; Nicholson and Yin, 2001), in which they ignore the groundwater contribution to the Lake Victoria.

![Figure 10: Comparison between average time series of altimetry-derived lake level variation (after applying ExtR retracking) and the filtered water storage variations.](image)

5. Conclusion

The present study (i) investigated the capability of the two-step data-driven approach of Simple Weighting (SW) and Postprocessing Filtering (PF) to improve remotely sensed datasets through merging and filtering to obtain four water cycle key components; precipitation, evaporation, discharge, and water storage variations, and (ii), explored the impacts of climate variabilities on water storage changes using the improved datasets in (i). The application of this approach, for the first time for LVB, results in a more efficient analysis of water fluxes that preserve water balance. The filtered water fluxes were largely in agreement compared to the original unfiltered datasets. This shows that SW+PF merging and filtering approach can effec-
tively reduce imbalances between different observations over a limited scale inland water bodies. It was also found that there is a remarkably smaller imbalance between the post-processed time series with various rates of contribution for each water component, e.g., 37% and 35% for precipitation and water storage changes, respectively, as the largest contributions. The achieved coherent datasets allowed for a better analysis of the lake’s water changes. Based on these, major rainfall spatial variabilities were observed in the central and western parts of the lake corresponding to the similar pattern in water storage changes. In addition, various strong anomalies were found in the filtered time series, e.g., in 2006 and 2014 (being negative), and 2005, 2007, and 2010 (being positive). The study showed that the climatic variation/change through the precipitation and evaporation (as indicators) are the main sources of the water storage changes within the lake. Moreover, an average correlation of 0.93 was found between water storage changes and the lake’s water level variations, which suggests that the main part of water storage changes within the lake refers to the variation of surface storages. These findings suggest the possible application of the applied algorithm to any inland lake that permits the use of satellite remote sensing, especially GRACE for studying water storage changes.

References


Birkett, C., Murtugudde, R., Allan, T. (1999); Indian Ocean climate event brings floods to East Africa’s lakes and the Sudd Marsh. GRL, 26(8), pp.1031-1034.


Consulate General of Ethiopia, (2012); Grand Ethiopian Renaissance Dam, Los Angeles, CA.


Frappart, F., Calmant, S., Cauhope, M., Seyler, F., Cazenave, A. (2006); Preliminary results of envisat ra-2-derived water levels validation over the amazon basin, Remote Sensing of Environment, 100(2), 252-264.


Khaki, M., Schumacher, M., J., Forootan, Kuhn, M., Awange, E., van Dijk, A.I.J.M. (2017c); Accounting for Spatial Correlation Errors in the Assimilation of GRACE into Hydrological


Kite, G.W. (1982); Analysis of Lake Victoria levels. Hydrological Sciences Journal, 27(2,6), 99-110.


Schneider, U., Fuchs, T., Meyer-Christoffer, A., Rudolf, B. (2008); In G. P. C. Centre (Ed.), Internet publication.


Tropical Rainfall Measuring Mission (TRMM) (2011); TRMM (TMPA/3B43) Rainfall Estimate L3 1 month 0.25 degree x 0.25 degree V7, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed [Data Access Date] https://disc.gsfc.nasa.gov/datacollection/TRMM_3B43_7.html.


