Reconstructing and Forecasting the Water Balance of Lake Victoria

Ari Basch
4/24/2023

1. Introduction

Lake Victoria is one of the largest freshwater lakes in the world, and serves as a critical resource to the region for energy production, the fishing industry, and agriculture. As one of the major sources of the Nile River, fluctuations in the water level of Lake Victoria have massive implications for the millions living in its basin and downstream. Here, we provide a comprehensive assessment of the Lake Victoria water balance using a combination of historical in situ data records, satellite observations, and a novel statistical model that has previously been applied to the Laurentian Great Lakes (in North America). More specifically, we employ this statistical model to develop historical estimates of Lake Victoria precipitation, evaporation, outflow, and tributary inflow that close the water balance over consecutive historical periods.

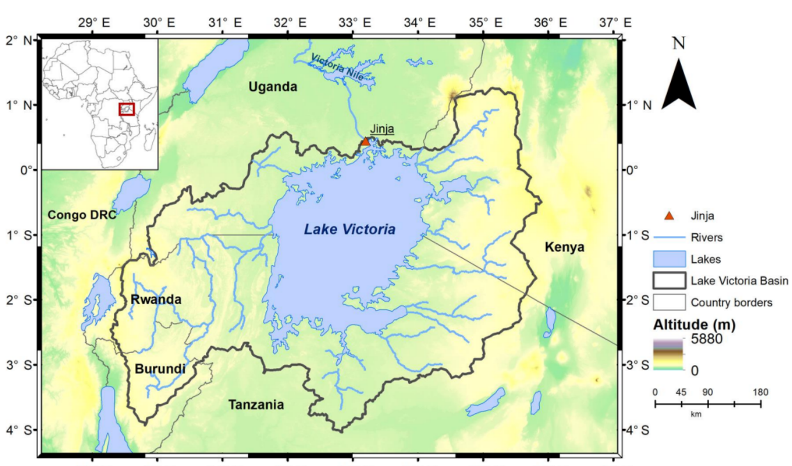

Figure 1. Map of Lake Victoria and its basin. Main tributaries are included along with the dam at Jinja. It's main tributary, the Kagera River, comes in from the west side of the lake (from Vanderkelen, 2018).

Previous studies on the lake have had limited success closing the water balance over long periods of time due largely to the lack of regular in situ measurements for some of the water balance components (Vanderkelen et al., 2018). A few studies have been able to recreate the water balance accurately using historical data, but in all such studies, one of the components was based off of observations of the other components using the water balance equation, resulting in the closure to be a priori. While some studies use observations for all the components, even the most comprehensive of these could not close the balance and had to include a bias correction.

This study uses an adaptation of the Large Lakes Statistical Water Balance Model (L2SWBM) which was developed for use on the Laurentian Great Lakes (Gronewold et al., 2020). We assume that the change in the volume of water stored in the lake is equivalent to the sum of volume fluxes into the lake subtracted by the fluxes out of the lake. The proposed model uses a rolling time window (of length w, in months) over which observed changes in lake storage (ΔV) across a w month period are equated to the cumulative sum of water balance components over the same period.

$$\Delta V_{j,w} = V_{j+w} - V_j = \sum_{i=j}^{j+w-1} (P_i - E_i + R_i - Q_i + \epsilon_i)$$

Equation 1. Assumption of the water balance encoded in the model

The L2SWBM utilizes a Bayesian framework to assimilate independent hydrological data products across Lake Victoria and subsequently infer the “true” monthly value for each water balance component. Although these monthly values are ultimately unknown, the L2SWBM is constrained by a traditional water balance equation, and uses multiple data sources to develop prior and likelihood functions for each component. Thus, the posterior estimates of the L2SWBM reconcile observed (or simulated) values from independent data products to close the water balance of the hydrologic system (Gronewold et al., 2020).

2. Configuring the Model

The model works best when datasets from a variety of sources are provided for each water balance component. This allows for flexibility in the construction of the prior distributions, and as such a majority of the time spent on this project was in finding and assessing the fidelity of datasets as opposed to working on the model. In situ data in the Lake Victoria basin is scarce, and the gauges that do exist are manually read and carried back to principal stations to be logged (Sangale et al., 2005). Therefore, we have to rely heavily on satellite data for most of the calibration, and even with these datasets, there are still decades where the model is missing all data for a water balance component. With that being said, the L2SWBM can simulate missing components effectively given a strong prior, and we can adjust uncertainty inputs to accurately represent the missing data. The following sections detail the data processing and subsequent model configuration that was done for each component.

2.1 Precipitation

As expressed more clearly in the results below, precipitation is responsible for around 70% of the water going into Lake Victoria (where runoff accounts for the other 30%). It has a strong seasonal cycle with two wet seasons in a year; the first happening in the spring and the second in the fall. Precipitation is gauged at several points around the lake, but Victoria is large enough that lake effect rainfall leads to significantly increased precipitation over the lake compared to the shoreline gauges (Yin, Nicholson, 1998). While there exists island gauges that can occasionally capture this discrepancy, such as the one in the northwestern Ssese Islands, records from these gauges are very infrequent (Nicholson et al., 2021). Thus we make use of three datasets derived from satellite observations, supplemented by data collected during Zoe Khatami's past work on an L2SWBM model for Victoria. These datasets are visualized below over a ten-year period.

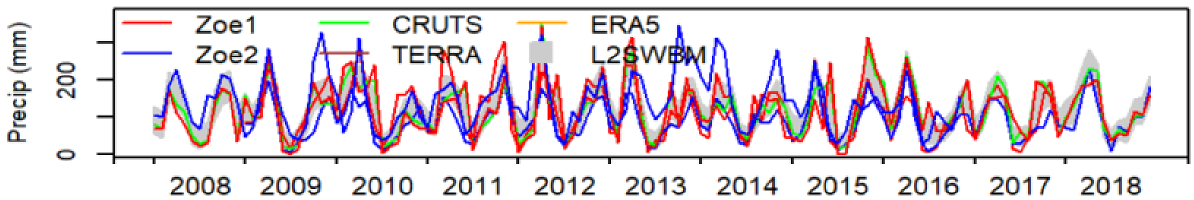

Figure 2. Precipitation datasets used in the L2SWBM from 2008 to 2018

When creating the prior distributions for each month, we want the resulting distribution to generally avoid values that would be completely unreasonable for the respective component to take on during that month. Thus we want to be careful when fitting a distribution to our data, as sometimes a normal distribution doesn't accurately represent it. Below we've created frequency charts for each month based off of one of our precipitation datasets, the Climatic Research Unit gridded Time Series (CRUTS). Typically, precipitation distributions lend themselves well to gamma-like distributions (Martinez-Villalobos, Neelin, 2019), and this seems to hold true for the CRUTS one below. A gamma distribution is fit to the data for each month to help assess its accuracy.

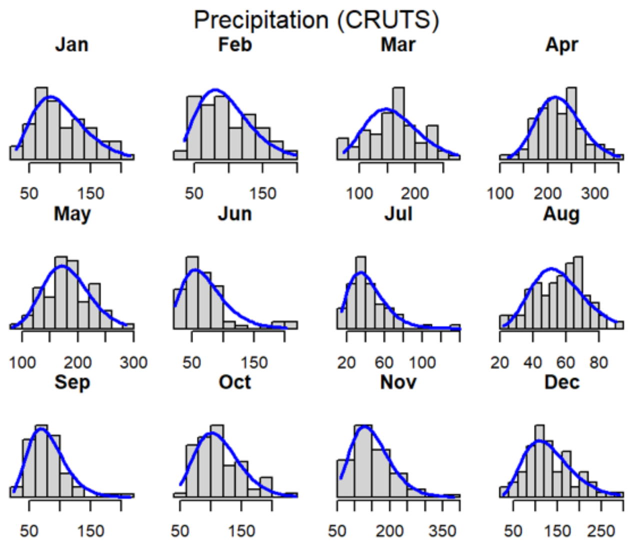

Figure 3. Monthly frequency charts for the CRUTS precipitation dataset

2.2 Evaporation

Similar to precipitation, evaporation is a major component of the water balance responsible for around 70% of the water leaving Victoria. The seasonal cycle is *much* less pronounced, to the point where some studies consider evaporation to be constant (Vanderkelen et al., 2018). Historical studies disagree on how uncertain evaporation estimates should be for the region, as Yin and Nicholson (1998) had it as the most uncertain component while Vanderkelen et al. (2018) had it as the least. This discrepancy follows from the multitude of ways to calculate the component used across different studies, such as the Penman Equation or latent heat flux formulas.

We generated evaporation priors using a variety of satellite products that calculate evaporation using the Penman equation. Monthly average values from Yin and Nicholson (1998), a study which uses latent heat flux formulas, were also used. The disagreement between evaporation datasets seen below, especially when it comes to seasonal cycles, is significantly more than those for other components such as precipitation. This is in part due to the sensitivity of the Penman equation to small changes in its variables, coupled with the imprecision of satellite measurements as opposed to in situ ones. Since evaporation also varies less than precipitation seasonally, differences can be more pronounced in a smaller range. These datasets are shown below, graphed during our simulation period.

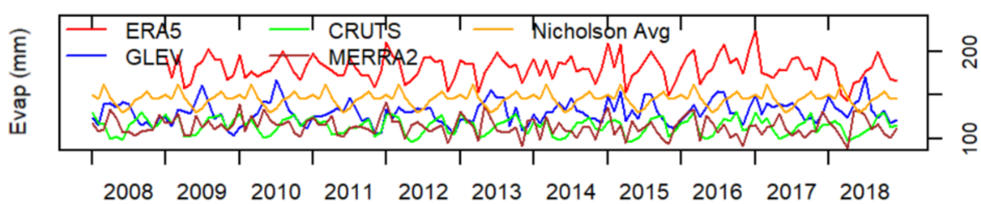

Figure 4. Evaporation datasets used in the L2SWBM from 2008 to 2018

Frequency charts were also created for evaporation to advise fitting prior distributions to the data. Creating an initial guess for the prior distribution is more of an approximation than an exact science since as the model runs and more datasets are included, the shape of the prior distributions will adjust to accommodate them. In this case, normal distributions fit the best, as visualized for the dataset below.

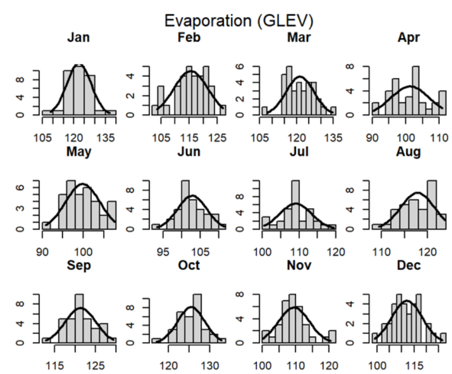

Figure 5. Monthly frequency charts for the CRUTS precipitation dataset

2.3 Outflow

Lake Victoria's outlet is on the north side of the lake at the Jinja dam and becomes one of the main sources of the Nile. Since the 1950s, outflow at the dam was designed to mimic naturalistic flow based on an "Agreed Curve" (Sene, 2000). This was followed until around 2000, when a second parallel dam was built to address increasing power demands. At this point, outflow deviated from the Agreed Curve and lake levels began dropping (Vanderkelen et al., 2018). Finally, in 2006 outflow measurements stopped being received (likely due to political dissonance in the region) but most studies agree that outflow continued to exceed the Agreed Curve after this period (Sutcliffe, Petersen, 2009).

We have very little outflow data to generate priors and likelihood functions. A dataset mirroring the Agreed Curve was provided although it's unclear if it's accurate to use the dataset for the simulation period, which takes place after outflow measurements deviated from the curve. Therefore, multiple runs of the model were conducted, some which take into account the likelihood functions and some only the priors and other water balance components. Two such runs can be compared below, where the gray band indicates the model's 95% confidence interval for outflow. For the final model run in the results below, a version of the model was chosen where outflow likelihood functions were *not* taken into account.

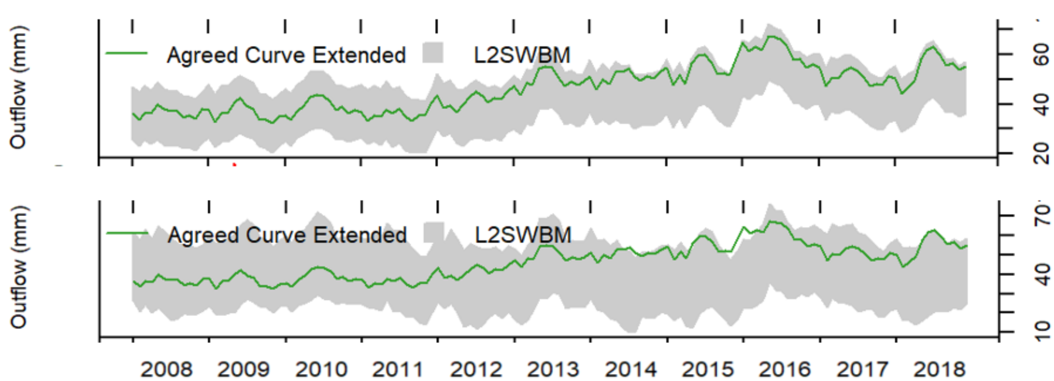

Figure 6. Model results with the outflow likelihood functions incorporated (on top) and without them (on bottom)

2.4 Runoff

The lake's largest tributary is the Kagera River which accounts for around 30% of the total inflow (Sutcliffe, Petersen, 2009). Since the 1970s, several tributaries have been gauged around the basin but data for all of them is scarce and rarely available (Yin, Nicholson, 1998). Likelihood functions for runoff are based on lognormal distributions, which best fit the datasets.

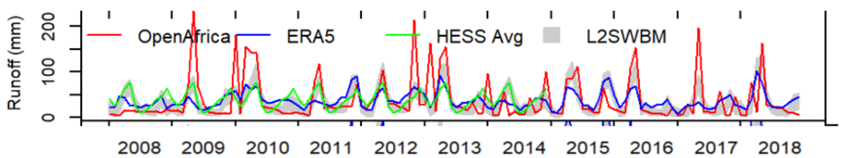

Figure 7. Runoff datasets used in the L2SWBM from 2008 to 2018

3. Results and Discussion

Provided in this section are the results from an example run of the model. We used a prior range of 1990-2007 with a posterior range of 2008-2018, and 4000 Markov Chain Monte Carlo iterations across 5 parallel chains. Likelihood functions for outflow were not taken into account, as discussed above, and range was standardized between all components to highlight impact. All values in the components and lake storage are measured in millimeters of lake level rise/fall. The surface area of the lake was assumed to be constant due to its large size, as changes due to rising/falling water levels would be insignificant.

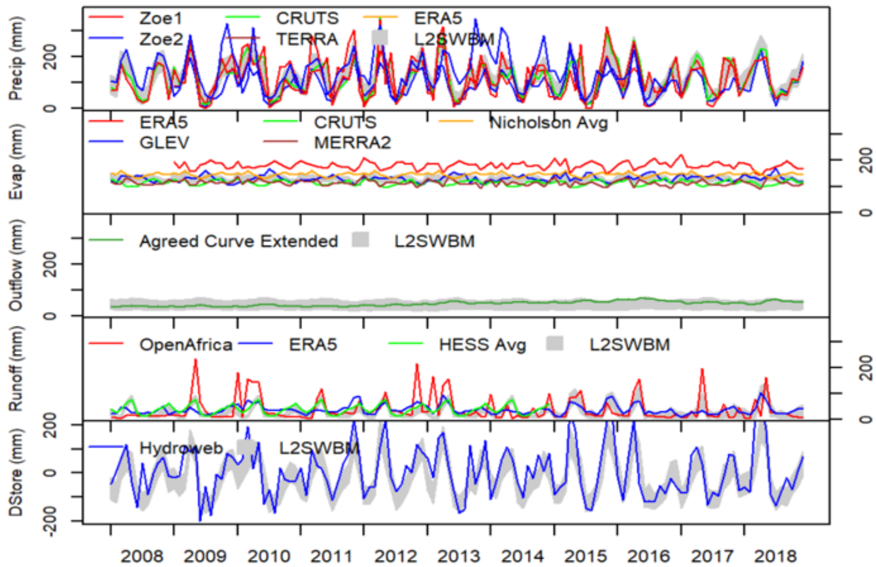

Figure 8. Model results with water balance component observations from Lake Victoria 2008-2018

Provided above are all the datasets for each component along with the change in storage at the bottom. The gray bar represents a 95% confidence interval from the model. With the components plotted next to each other, their respective influence on the water balance is better visualized. Uncertainty in the model results was kept to around ±20mm for the lake level and components which is relatively constrained. Once we obtain results from a run, we typically want to assess the fidelity of them and simulate the water balance based on the estimates; this is done below.

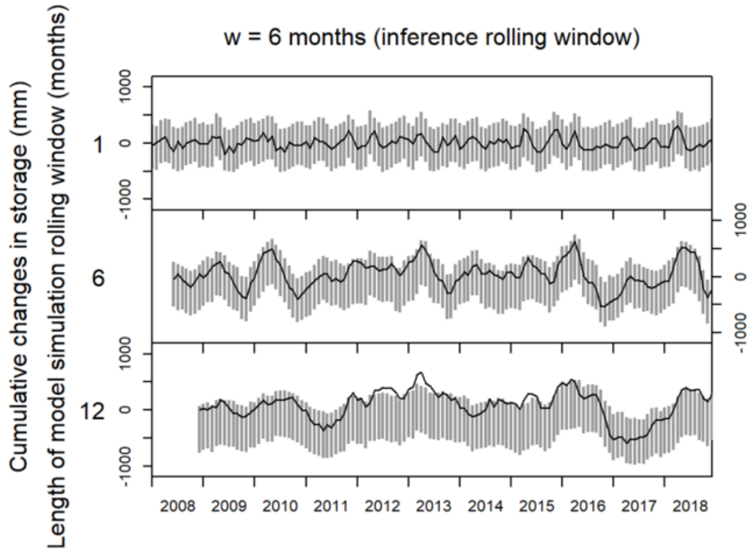

Figure 8. Observed (black line) and simulated (grey 95% credible intervals) changes in lake storage over periods of 1, 6, and 12 months. Model results are presented for a rolling inference window of 6 months.

We can assess the fidelity of our model results by taking the results and simulating the water balance using the given estimates. We can then compare the results of the simulation with the true change in storage to get a sense of how close the estimates were. This was done above across a 1, 6, and 12 month simulation window. As we attempt to simulate the water balance for longer and longer periods of time, the model's results begin to deviate from the truth. Over the course of this project, one of the main goals driving many of our decisions was to both shrink the uncertainty in the model and improve the performance of long-term simulations based on the model's estimates.

4. Conclusion

We have demonstrated the adapted L2SWBM model for Lake Victoria developed during this project can decently close the water balance of the lake with limited uncertainty. A variety of datasets were collected for each of the components that significantly expanded upon past implementations of the model. Moving forward, we'd like to continue to find more data and improve the long-term simulations done on results. If possible, collecting data onsite and following up with local authorities on data availability would be very beneficial. We should also continue to tweak how uncertainty is applied to the different components and work toward even better simulations. If we can achieve a model that closes the water balance decades into the future, we can provide critical hydrological insight to the region and help local governments better adapt to the dynamic lake.

Works Cited

Gronewold, A. D., SmithLa, J. P., Read, L. K., & Crooks, J. L. (2020, January 10). *Reconciling the water balance of Large Lake Systems*. Advances in Water Resources. Retrieved April 27, 2023, from https://www.sciencedirect.com/science/article/abs/pii/S0309170819305263

Martinez-Villalobos, C., & Neelin, D. (2019, November 1). *Why do precipitation intensities tend to follow gamma distributions?* American Meteorological Society. Retrieved April 27, 2023, from https://journals.ametsoc.org/view/journals/atsc/76/11/jas-d-18-0343.1.xml

Nicholson, S., Klotter, D., & Hartman, A. (2021, June 1). *Lake-effect rains over Lake Victoria and their association with Mesoscale Convective Systems*. American Meteorological Society. Retrieved April 27, 2023, from https://journals.ametsoc.org/view/journals/hydr/22/6/JHM-D-20-0244.1.xml?rskey=r9CtMR&result=3

Sangale, F., Njuguna, H., Okungu, J., Abuodha, J. O. Z., & Hecky, R. E. (2005). *Meteorology and hydrology of the Lake Victoria Basin: Kenyan sector.* Aquadocs. Retrieved April 27, 2023, from https://aquadocs.org/bitstream/handle/1834/7148/ktf0091.pdf?sequence=1

Sene, K. J. (2000, February). *Theoretical estimates for the influence of Lake Victoria on flows in the upper White Nile*. Hydrological Sciences. Retrieved April 27, 2023.

Sutcliffe, J. V., & Petersen, G. (2009, December 15). *Lake Victoria: derivation of a corrected natural water level series*. Hydrological Sciences. Retrieved April 27, 2023, from https://www.tandfonline.com/doi/abs/10.1623/hysj.52.6.1316

Vanderkelen, I., van Lipzig, N., & Thiery, W. (2018, October 25). *Modelling the water balance of Lake Victoria (East Africa) – part 1: Observational analysis*. Hydrology and Earth System Sciences. Retrieved April 27, 2023, from https://hess.copernicus.org/articles/22/5509/2018/#bib1.bibx51

Yin, X., & Nicholson, S. (1998). *The water balance of Lake Victoria*. Hydrological Sciences. Retrieved April 27, 2023, from https://www.tandfonline.com/doi/epdf/10.1080/02626669809492173?needAccess=true&role=button