Reconciling the water balance of large lake systems

Andrew D. Gronewold[a,b,*], Joeseph P. Smith[c], Laura Read[d], James L. Crooks[e,f]

[a] *School for Environment and Sustainability, University of Michigan, Ann Arbor, Michigan USA, 48109*
[b] *Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, Michigan USA, 48109*
[c] *Cooperative Institute for Great Lakes Research, University of Michigan, Ann Arbor, Michigan USA, 48109*
[d] *National Center for Atmospheric Research, Boulder, Colorado USA, 80307*
[e] *Division of Biostatistics and Bioinformatics, National Jewish Health, Denver, Colorado USA, 80206*
[f] *Department of Epidemiology, Colorado School of Public Health, Aurora, Colorado USA, 80045*

Abstract

Water balance models are commonly employed to improve understanding of drivers behind changes in the hydrologic cycle across multiple space and time scales. Generally, these models are physically-based, a feature that can lead to unreconciled biases and uncertainties when a model is not encoded to be faithful to changes in water storage over time. Statistical methods represent one approach to addressing this problem. We find, however, that there are very few historical hydrological modeling studies in which bias correction and uncertainty quantification methods are routinely applied to ensure fidelity to the water balance. Importantly, we know of none (aside from preliminary applications of the model we advance in this study) applied specifically to large lake systems. We fill this gap by developing and applying a Bayesian statistical analysis framework for inferring water balance components specifically in large lake systems. The model behind this framework, which we refer to as the L2SWBM (large lake statistical water balance model), includes a conventional water balance model encoded to iteratively close the water balance over multiple consecutive time periods. Throughout these iterations, the L2SWBM can assimilate multiple preliminary estimates of each water balance component (from either historical model simulations or interpolated *in situ* monitoring data, for example), and it can accommodate those estimates even if they span different time periods. The L2SWBM can also be executed if data for a particular water balance component are unavailable, a feature that underscores its potential utility in data scarce regions. Here, we demonstrate the utility of our new framework through a customized application to the Laurentian Great Lakes, the largest system of lakes on Earth. Through this application, we find that the L2SWBM is able to infer new water balance component estimates that, to our are knowledge, are the first ever to close the water balance over a multi-decadal historical period for this massive lake system. More specifically, we find that posterior predictive intervals for changes in lake storage are consistent with observed changes in lake storage across this period over simulation time intervals of both 6 and 12 months. In additional to introducing a framework for developing definitive long-term hydrologic records for large lake systems, our study provides important insights into the origins of biases in both legacy and state-of-the-art hydrological models, as well as regional and global hydrological data sets.

Keywords: hydrologic cycle, large lakes, statistical modeling, Bayesian inference, water balance

1. Introduction

Hydrological models that simulate and forecast the water balance across a variety of space and time scales are needed to facilitate water resources management planning and, ultimately, to ensure human and environmental health (Vörösmarty et al., 2000; Pekel et al., 2016). This need is particularly pronounced in regions where rapid population growth coincides with changes in the spatiotemporal distribution of fresh water, and where the sustainability of future water supplies is uncertain (Schewe et al., 2014). To address this need, hydrological models need to clearly differentiate components of the hydrologic cycle that are often confounded (including, for example, evapotranspiration and irrigation water demand) and to quantify changes in those components over time (Nijssen et al., 2001; Kebede et al., 2006; Raes et al., 2006; Li et al., 2007; Gronewold and Stow, 2014).

*Corresponding author. Tel.: +1-734-764-6286
Email address: drewgron@umich.edu (Andrew D. Gronewold)

Global, continental, and basin-scale water balance modeling research typically focuses on improving representation of terrestrial and atmospheric physical processes collectively governing precipitation, evapotranspiration, and streamflow (Kim and Stricker, 1996; Vörösmarty et al., 1998; Crow et al., 2008; Senay et al., 2011; Milly and Dunne, 2017). This body of research, while providing foundational hydrologic data for much of the planet's land surface, rarely explicitly resolves mass and energy fluxes over large freshwater surfaces (Makhlouf and Michel, 1994; Xu and Singh, 1998; Arnell, 1999; Guo et al., 2002). Put differently, we find that the primary physical processes driving the water balance of large lakes, including over-lake evaporation (i.e. turbulent heat fluxes), over-lake precipitation, and predominant channel lake inflows and outflows, are represented poorly (if at all) in large-scale terrestrial land surface models and corresponding data sets.

There are, however, several state-of-the-art models that represent these processes specifically for lakes and for their interactions with the atmosphere. One particular example is WRF-lake, a one-dimensional (1-D) physically-based lake model that, in previous studies (Gu et al., 2013; Xiao et al., 2016), has been coupled with the Weather Research and Forecasting (WRF) model. A similar one-dimensional scheme has been applied to resolve large lakes (Holman et al., 2012; Notaro et al., 2013) within the Abdus Salam International Center for Theoretical Physics Regional Climate Model (ICTP RegCM4).

There have also been significant evolutions in three-dimensional lake models, including the transition of the National Oceanic and Atmospheric Administration (NOAA) Great Lakes Operational Forecasting System (GLOFS) from the Princeton Ocean Model (POM) to the Finite-Volume Community Ocean Model (or FVCOM; for details see Kelley et al., 2018), as well as the operationalization of the Nucleus for European Modelling of the Ocean (NEMO) model (Dupont et al., 2012) within Environment and Climate Change Canada's Water Cycle Prediction System (WCPS, described in Durnford et al., 2018).

While many of these (and other) existing lake models have been found to represent key physical processes, they are rarely (if at all) evaluated within the context of the overall hydrologic cycle, and reconciliation of changes in lake storage over multiple time periods. This missing piece of context in most lake modeling studies limits the extent to which scientists and practitioners, along with the general public, understand modes of variability in the storage of nearly all of the Earth's fresh unfrozen surface water (Gibson et al., 2006;

Swenson and Wahr, 2009; Xiao et al., 2018). Filling this gap requires a focused e ort on lake water balance models that resolve physical processes governing lake storage at appropriate space and time scales (Piper et al., 1986; Nicholson et al., 2000; Gibson et al., 2006; Peng et al., 2019).

Here, we address this gap by introducing a framework that employs a novel formulation of a lake water balance model in which historical monthly water balance components are estimated through Bayesian inference (Gelman et al., 2004). Empirical data sets and historical model simulations, if available, are incorporated into the framework through likelihood functions and prior probability distribution functions. This approach leads not only to probabilistic historical water balance component estimates that preserve spatial and temporal correlation across a lake (or lake system), but also to quantification of bias and uncertainty in the models and data sets that had previously been developed for that lake.

It is informative to note that a preliminary prototype of our framework was applied in a study assessing rapid water level changes between 2013 and 2015 on Lakes Superior, Michigan, and Huron (Gronewold et al., 2015, 2016). An analysis of an evolution of that prototype, which focused primarily on model selection, was presented in Smith and Gronewold (2018). The framework we present here is differentiated from the prototype in Gronewold et al. (2016) by (among other features) two key improvements. The first is a computationally-efficient filtering method (which we periodically refer to as a "rolling inference window") that facilitates inference over multi-decadal periods. The second is a skill assessment that reflects both the relative homoscedasticity of model residuals, as well as the extent to which the model closes a lake's water balance over consecutive multi-month time periods.

We also acknowledge that there are previous studies utilizing similar statistical methods, such as those that debias continental remote sensing data (Pan and Wood, 2006; Coccia et al., 2015). These studies, however, are typically focused exclusively on land surface processes and do not adequately resolve large lakes. The framework we develop here, therefore, is further distinguished by its focus on large lakes, including its ability to model multiple connected lakes in series. Our representative application also represents the first time a water balance model has been applied systematically to the entire Laurentian Great Lakes system (the largest system of lakes in Earth) that "closes" the water balance over multiple time periods while reconciling discrepancies between alternate measurements and model simulations of the same water balance component. The results of our application serve as both a solution to a long-standing water resources management problem for the Great Lakes, and as a stepping stone towards solving similar problems in large lake systems around the world.

2. Methods

In the following sections, we first (section 2.1) provide an overview of the generic formulation of our new framework. We then (section 2.2) describe the customization of our framework to the water balance of the entire Laurentian Great Lakes system. Section 2.2 also includes a description of our approach to evaluating the new framework using the results from our application to the Laurentian Great Lakes.

2.1. The model

We developed our modeling framework by first modifying the conventional formulation of a lake water balance model to represent changes in storage over a time window of w months:

$$\Delta H_{j,w} = H_{j+w} - H_j = \sum_{i=j}^{j+w-1} (P_i - E_i + R_i + I_i - Q_i + D_i + \epsilon_i) \qquad (1)$$

where $\Delta H_{j,w}$ represents the "true" change in lake storage over a period of w months (starting with month j), H_j and H_{j+w} represent "true" monthly average lake water levels (in mm) at the beginning of months j and $j + w$ (respectively), $j \in [1, T - w + 1]$, and T is the total number of months over which the model is run. Our use of i and j as month number indices within the context of equation 1 accommodates this rolling multi-month window approach. The value of the month index j in equation 1 can not, by definition, exceed $T - w + 1$. We index monthly water balance components outside of the context of equation 1 using $t \in [1, T]$.

The "true" values for monthly water balance components (expressed in mm over a lake surface area) in equation 1 include over-lake precipitation P, over-lake evaporation E, lateral tributary runo R, inflow from an upstream channel I, discharge through a downstream channel Q, and the total of inter-basin diversions and consumptive uses D. The model also includes a process error term (ϵ) to account for potential sources of water balance variability not explained by components P, E, R, I, Q, and D alone including, for example, thermal expansion, glacial isostatic rebound, and groundwater fluxes (Quinn and Guerra, 1986; Mainville and Craymer, 2005). These terms could be added to equation 1 on a case-by-case basis depending on whether they are expected to be significant. Probabilistic estimates of each water balance component in equation 1 are inferred in a Bayesian framework (Press, 2003; Gelman et al., 2004) in which prior probability distributions and likelihood functions are parameterized using legacy models and data sets, as well as expert knowledge and opinion (specifically for prior probability distributions).

2.1.1. Likelihood functions

The likelihood function for the change in storage within a given lake over a period of w months is:

$$y_{\Delta H,j,w} \;=\; y_{H,j+w} - y_{H,j} \sim \mathsf{N}(\Delta H_{j,w}, \tau_{\Delta H,w}) \tag{2}$$

in which the observed change in storage $y_{\Delta H}$ starting in month j, and across a rolling window of length w, is the difference between water level measurements (y_H) at the beginning of months $j+w$ and j. We model this value with a normal distribution with mean $\Delta H_{j,w}$ and precision $\tau_{\Delta H,w}$. This approach allows for an explicit representation of uncertainty in water level data that can be differentiated from uncertainty in water balance component estimates. It is informative to note that rather than using variance (σ^2), we parameterize normal distributions using precision ($\tau = 1/\sigma^2$) following conventional practice for Bayesian inference (Casella and Berger, 2002; Gelman et al., 2004; Qian et al., 2009)

We then introduce three new parameters, I', Q' and D', to represent connecting channel inflows, outflows, and diversions (respectively) in units of m^3/s. We use these units because most water management practitioners are accustomed to recording and assessing these values in m^3/s, rather than mm over each lake surface. We encode the empirical relationship (i.e. the conversion of units) between parameters I, Q, and D and (respectively) I', Q' and D' using the surface area of each lake and the number of seconds in a particular month.

The likelihood functions for water balance components on the right-hand side of equation 1 (represented collectively by $\theta \in P, E, R, I', Q', D'$) is:

$$y^n_{t,\theta} \sim \mathsf{N}(\theta_t + \eta^n_{\theta,c_t}, \tau^n_{t,\theta}) \quad (3)$$

where $y^n_{t,\theta}$ is data source $n \in [1, N]$ for component θ at time step t, N is the total number of data sources for that component, η^n_{θ,c_t} is the bias of the n^{th} data source in calendar month c, and $\tau^n_{t,\theta}$ is the data source precision at time step t.

2.1.2. Prior probability distributions (standard formulation)

In Bayesian statistics, parameters are frequently modeled with normal probability distributions to support inference across a broad range of potential values. Alternate probability distribution families can be used, however, to reflect knowledge (or beliefs) that a parameter might have a more limited range of values.

We model E, I', Q', and D' with normal prior probability distributions:

$$\begin{aligned}
\pi(E_t) &= \mathsf{N}(\mu_{E,c_t}, \tau_{E,c_t}/2) && (4)\\
\pi(I'_t) &= \mathsf{N}(\mu_{I',c_t}, \tau_{I,c_t}) && (5)\\
\pi(Q'_t) &= \mathsf{N}(\mu_{Q',c_t}, \tau_{Q,c_t}) && (6)\\
\pi(D'_t) &= \mathsf{N}(\mu_{D',c_t}, \tau_{D',c_t}) && (7)
\end{aligned}$$

where prior means μ_{c_t} and precisions τ_{c_t} for each calendar month c are either calculated empirically using historical data records, or informed by expert opinion (for further reading on objective and subjective prior probability distributions, see Press, 2003). This approach allows for the possibility that lake evaporation can be both positive (i.e. a loss of water from a lake) and negative (i.e. when there is warm overlying air and condensation occurs). This approach is also suited for relative high values of connecting channel flows Q' and diversions D'. Future users of our framework could, should they choose to do so, select different prior probability distribution families (such as lognormal, for example).

We divide precision in half (i.e. double the variance) for prior probability distributions on over-lake evaporation E because, for many large lakes, evaporation has a very strong historical seasonal cycle with relatively low variability. That historical low variability could, when quantified in the parameters of a prior probability distribution, overly-constrain the range of inferred monthly evaporation estimates during a later period, particularly in lakes where climate change has led to a systematic increase in evaporation over time (Milly et al., 2008).

Total lateral tributary runo values aggregated over a lake basin and over monthly time steps are almost certainly positive, and we therefore model R with a log-normal prior probability distribution:

$$\pi(R_t) = \mathsf{LN}(\mu_{ln(R),c_t}, \tau_{ln(R),c_t}) \tag{8}$$

with prior means $\mu_{ln(R),c_t}$ and precisions $\tau_{ln(R),c_t}$. These values can be calculated for each calendar month c empirically using historical data records, or formulated to represent expert opinion.

For over-lake precipitation, we use a gamma prior probability distribution (Husak et al., 2007):

$$\pi(P_t) = \mathsf{Ga}(\psi^1_{c_t}, \psi^2_{c_t}) \tag{9}$$

with shape ψ^1 and rate ψ^2 (following Thom, 1958) defined as:

$$\begin{aligned} \psi^1_{c_t} &= \frac{1}{4\phi_{c_t}}\left(1 + \sqrt{1 + \frac{4\phi_{c_t}}{3}}\right) \\ \phi_{c_t} &= \ln(\mu_{P,c_t}) - \mu_{ln(P),c_t} \\ \psi^2_{c_t} &= \psi^1_{c_t}/\mu_{P,c_t} \end{aligned}$$

where μ_{P,c_t} is the mean historical precipitation for each month, and $\mu_{ln(P),c_t}$ is the mean of the logarithm of precipitation for each calendar month c.

We then model the bias of each contributing data set using normal prior probability distributions:

$$\pi(\eta^n_{\theta,c_t}) = \mathsf{N}(0, 0.01) \tag{10}$$

with mean 0 and precision 0.01. We note that this precision is equivalent to a standard deviation of 10, and is in units of mm over a lake surface for η_P, η_E, and η_R, while it is in units of m^3/s for $\eta_{Q'}$, $\eta_{I'}$, and $\eta_{D'}$. Users of our framework can customize these prior probability distributions by selecting mean and precision values that are unique to each bias term. Our representative application in the next section provides an example.

Finally, following Gelman (2006), we modelled $\tau_{\Delta H,w}$ and $\tau_{t,\theta}$ using a gamma $\mathsf{Ga}(0.1, 0.1)$ prior probability distribution with shape and scale parameter both equal to 0.1. Similarly, we constrained water balance model errors to one of 12 values corresponding to each of the 12 calendar months, with each error term having a common vague normal prior probability distribution with mean 0 and precision 0.01:

$$\begin{aligned} \epsilon_t &= \epsilon_{c_t} \\ \pi(\epsilon_{c_t}) &= \mathsf{N}(0, 0.01) \end{aligned} \tag{11}$$

We recognize that bias estimates in our model may be impacted by the classic problem of bias-variance tradeo (Geman et al., 1992). We view implementation of solutions to this problem, such as bias-variance decomposition (Valentini and Dietterich, 2004), as a potential future step in our research.

2.2. Representative application: the Laurentian Great Lakes

To demonstrate the utility of our model, we customized it to the entire Laurentian Great Lakes system (hereafter referred to simply as the "Great Lakes") to infer new monthly water balance components for the period 1980 through 2015. The Great Lakes system (figure 1) includes Lakes Superior, Michigan, Huron, St. Clair, Erie, and Ontario. Here, we follow conventional practice in Great Lakes hydrological modeling research at coarse time scales (e.g. monthly) by representing Lakes Michigan and Huron as a single lake (Lake Michigan-Huron) given the depth and breadth of the channel (i.e. the Straits of Mackinac) that connects them (Quinn and Edstrom, 2000; Pietroniro et al., 2007). Collectively, the Great Lakes represent the largest system of lakes on Earth; Lakes Superior and Michigan-Huron alone are the two largest lakes on Earth by surface area (Gronewold et al., 2013).

We encoded lake-to-lake connectivity within the Great Lakes system (i.e. through the St. Marys, St. Clair, Detroit, and Niagara Rivers) by defining the inflow to each lake through a major connecting channel (I') as the outflow from the adjacent upstream lake (Q'). For example, the inflow to Lake Michigan-Huron through the St. Marys River at each monthly time step t is encoded as $Q'_{SUP,t}$, the outflow from Lake Superior. There is no upstream connecting channel flowing into Lake Superior and therefore, in equation 1 for Lake Superior, $I_{SUP} = 0$. We obtained surface areas for each of the Great Lakes (table 1) from the National Oceanic and Atmospheric Administration (NOAA) Great Lakes Environmental Research Laboratory (GLERL) to calculate relationships between Q', D', Q and D . For details, see Hunter et al. (2015).

2.2.1. Model modifications for the water balance of Lake St. Clair

We model Lake St. Clair differently from the other Great Lakes because its surface area is relatively small (table 1), and because its hydrologic cycle is dominated by inflow from the St. Clair River and outflow to the Detroit River. More specifically, rather than differentiating precipitation, evaporation and runo for Lake St. Clair, we represented them collectively as a single term commonly referred to as net basin supply ($NBS' = P' - E' + R'$). We model the NBS' for Lake St. Clair using modified versions of equations 1 and 2 (in units of m^3/s) as follows:

$$\Delta H'_{j,w} = \sum_{i=j}^{j+w-1} \left(NBS'_i + Q'_{MHU_i} - Q'_i + D'_i + \epsilon_{STC_i} \right) \quad (12)$$

$$y_{\Delta H',j,w} \sim \mathsf{N}(\Delta H'_{j,w}, \tau_{\Delta H',w}) \quad (13)$$

where Q'_{MHU} is the outflow from Lake Michigan-Huron (i.e. the inflow to Lake St. Clair through the St. Clair River).

We model Lake St. Clair NBS' values with a normal prior probability distribution:

$$\pi(NBS'_t) = \mathsf{N}(\mu_{NBS',c_t}, \tau_{NBS',c_t}) \quad (14)$$

and a normal likelihood function:

$$y^n_{t,NBS'} \sim \mathsf{N}(NBS'_t + \eta^n_{NBS',c_t}, \tau^n_{t,NBS'}) \quad (15)$$

where η^n_{NBS',c_t} is the bias of NBS' estimate $n \in [1, N]$ in calendar month c, N is the total number of NBS' data sources, and $\tau^n_{t,NBS'}$ is the precision of each data source.

We then adapt the prior probability distributions from equations 10 and 11 for NBS' data bias and Lake St. Clair model error such that the prior precision for each ($\tau = 0.0625 = 1/\sigma^2$) is equivalent to a standard deviation (σ) of 4 m^3/s, or roughly 10mm of water on the surface of Lake St. Clair (table 1) over the course of a month:

$$\pi(\eta^n_{NBS',ct}) = \mathsf{N}(0, 0.0625) \quad (16)$$

$$\pi(\epsilon_{STC}) = \mathsf{N}(0, 0.0625) \quad (17)$$

2.2.2. Data for application to the Great Lakes

We obtained beginning-of-month lakewide-average water level data (y_H) for each of the Great lakes, as well as historical records of interbasin diversions ($y_{D'}$), channel flows ($y_{Q'}$), and estimates of Lake St. Clair's net basin supply from the Coordinating Committee on Great Lakes Basic Hydraulic and Hydrologic Data (CCGLBHHD). For further reading on the CCGLBHHD, see Gronewold et al. (2018).

The water level data we obtained from the CCGLBHHD is derived from water level measurements at gauges located around the coastline of each of the Great Lakes that are maintained by both the NOAA National Ocean Service Center for Operational Oceanographic Products and Services (NOAA-NOS CO-OPS) and the Canadian Department of Fisheries and Oceans' Canadian Hydrographic Service (DFO-CHS). We recognize that alternate sources of water level information are available, including those based on satellites and other remote sensing products (Alsdorf et al., 2001; Crétaux et al., 2011; Schwatke et al., 2015). For this application, we utilize the relatively robust network of Great Lakes water level gauging stations synthesized in the CCGLBHHD records, and leave assimilation of remotely-sensed water levels to future research on either the Great Lakes, or other large lake systems.

Similarly, we obtained data on diversions into, out of, or within each lake basin from the CCGLBHHD including the Ogoki River and Long-Lac diversions into Lake Superior, the Chicago River diversion out of Lake Michigan-Huron, and the Welland Canal that runs parallel the Niagara River (figure 1).

We then obtained two sets of connecting channel flow data. The first ($y^1_{Q'}$) includes estimates for each of the Great Lakes connecting channels derived by the CCGLBHHD using a variety of standard methods. These methods include the aggregation of discrete flow measurements through dams and marine navigation locks, and the application of stage-fall discharge equations. The second set of connecting channel flow estimates ($y^2_{Q'}$) is based on Acoustic Doppler Velocity Meters (ADVMs) located at International Gauging Stations (IGS) maintained through a partnership between the United States Geological Survey (USGS) and Water Survey Canada (WSC). These IGS measurements are available only for the St. Marys, St. Clair, and Detroit Rivers.

We then obtained a set of data for over-lake precipitation, over-lake evaporation, and runo from the NOAA-GLERL Great Lakes Monthly Hydrometeorological Database (or GLM-HMD, described in Hunter et al., 2015). The GLM-HMD utilizes a suite of models and interpolation schemes to generate 1-dimensional estimates of water balance components over the land and lake surfaces of each of the Great Lakes. More specifically, over-lake precipitation estimates in the GLM-HMD are based on Thiessen weighting (Croley II and Hartmann, 1985) of land-based meteorological station data (for further discussion, see Holman et al., 2012). Over-lake evaporation simulations in the GLM-HMD are derived from the legacy Large Lakes Thermodynamics Model (LLTM) which utilizes wind speed, dew point, cloud cover, and lake surface temperature to simulate heat exchange and ice cover across the lakes (Croley II, 1989, 1992). Finally, runo estimates in the GLM-HMD are derived from an area-ratio based interpolation of USGS and WSC streamflow gages across the basin (for further reading, see Fry et al., 2013).

We obtained additional data for over-lake precipitation, over-lake evaporation, and runo from two Canadian federal government products; the Canadian Precipitation Analysis (or CaPA) and GEM-MESH. GEM-MESH is a configuration of the Modélisation Environmentale-Surface et Hydrologie (MESH) forced by the Canadian Global Environmental Multiscale (GEM) numerical weather prediction model (Deacu et al., 2012; Lespinas et al., 2015).

We utilized each of these data sets during our model inference routine (described below in section 2.2.4) for the period 1980 to 2015. We also used a different subset of these data for calculating prior probability distribution hyper-parameters, as described in the following section. A complete summary of the data used for our representative application to the Great Lakes, including an indication of how each data set was used in either prior probability distributions or likelihood functions, is included in Appendix A.

2.2.3. *Prior probability distributions: application to the Great Lakes*

For our application to the Great Lakes, we employed prior probability distributions for model parameters P, E, R, Q', and D' prescribed by equations 4 through 9. For P, E, and R, we calculated hyper-parameters empirically using historical data from 1950 through 1979 from the GLM-HMD. For Q' and D', as well as NBS' values for Lake St. Clair, we calculated hyper-parameters using data from 1950 through 1979 from the CCGLBHHD.

Similarly, we employed equation 10 as a prior probability distribution for bias in data sources for over-lake precipitation, over-lake evaporation, and runo (i.e. η_P, η_E, and η_R), and equation 16 as a prior probability distribution for bias in data sources for Lake St. Clair NBS'. However, we modified the prior probability distributions for bias in channel flow and diversion data by calculating the prior probability distribution precision as $\tau = 1/\sigma^2$ and $\sigma = \lambda * \mu$, where μ is the historical empirical monthly mean (of Q' and D'), and λ is a coefficient of variation unique to a particular source of data for Q' and D' (table 2) reflecting information we obtained from regional experts (for further information on expert opinion solicitation, see Borsuk et al., 2001; Voinov and Bousquet, 2010).

2.2.4. *Model inference and analysis*

We implemented three configurations of our model, each with either a 1-month, 6-month, or 12-month rolling inference window. We encoded these configurations in JAGS (Just Another Gibbs Sampler; Plummer, 2003), and executed the JAGS model inference routine through the 'rjags' package in the `R` statistical software environment (R core team, 2017). JAGS is an open-source, cross-platform engine of the BUGS (Bayesian inference Using Gibbs Sampling) language (Lunn et al., 2000) which has been applied in numerous Bayesian inference studies across a range of disciplines (Lunn et al., 2009; Kéry, 2010; Ntzoufras, 2011; Parkes and Demeritt, 2016). JAGS model code is included for reference in on-line supplementary material.

We ran each model for $K = 1{,}000{,}000$ Markov chain Monte Carlo (MCMC) iterations across three parallel MCMC chains. We omitted the first 500,000 iterations as a 'burn-in' period, and then thinned the remaining 500,000 iterations at even intervals such that each chain had 1,000 values. The resulting 3,000 MCMC samples serve as the basis for our estimates of the posterior probability distributions for each model parameter.

We evaluated each configuration by first assessing homoscedasticity of model errors (i.e. ϵ), and then by assessing the extent to which inferred water balance components closed the water balance over different time horizons. This evaluation allowed us to better understand relationships between the length of an inference rolling window, and the range of time horizons over which the corresponding model provides results that close the water balance. Some water management agencies, for example, need monthly water balance component estimates that are consistent with observed changes in lake storage on only a month-to-month basis. Others, such as seasonal forecasting authorities, may be concerned with changes in the water balance over longer time horizons.

We then used the inferred water balance component estimates (and other model parameters) from each model configuration to simulate the posterior predictive distribution (Gelman et al., 2002; Kruschke, 2013) of observed changes in lake storage (i.e. left side of equation 2). It is informative to note that after inferring monthly water balance components, we can use those components to simulate changes in lake storage over *any* time horizon; we are not, in other words, constrained to simulating over only 1, 6, and 12-month windows (i.e. the time windows we used to infer the water balance components). To address potential water resources management planning needs over a range of time scales, we elected to calculate the posterior predictive distribution for observed changes in storage across consecutive time windows of 1 month, 12 months, and 60 months, all between 1980 and 2015.

3. Results and discussion

3.1. Model diagnostics

3.1.1. Process model error distribution (i.e. homoscedasticity)

Our assessment of monthly model process errors indicates that errors in the model configuration with a 1-month inference window (left column, figure 2) reflect significant seasonality, particularly for Lakes Superior and Michigan-Huron. This finding indicates that there is an important mode of variability in the Great Lakes seasonal cycle that is not represented in water balance component estimates derived from a model with a 1-month inference window.

The errors in the model configuration with a 6-month inference window also reflect seasonality for Lakes Superior and Michigan-Huron, though not with nearly the same severity as the model with a 1-month inference window. It is interesting to note that model errors are relatively uniform for Lake Erie and Ontario for both the 1-month and 6-month configurations. This finding most likely reflects the fact that connecting channel flows represent a higher proportion of each lake's water balance moving downstream through the system from Lake Superior to Lake Ontario.

Errors in the model configuration with a 12-month inference window (right column, figure 2) do not follow any noticeable seasonal pattern. There does, however, appear to be some evidence of a positive bias (where the mean model error is slightly less than zero) in models for Lakes Superior and Michigan-Huron, though this evidence is very weak (i.e. the uncertainty bounds suggest there is no evidence for error values other than zero).

3.1.2. Long-term water balance closure

We find that our model configuration with a 1-month inference window only closed the water balance when simulating changes in lake storage over a 1-month period (top row, figure 3). When used to simulate changes in storage over consecutive 12- and 60-month simulation periods, water balance components from the model with a 1-month inference window accumulate severe biases and lead to very wide prediction intervals (middle row and bottom row, figure 3). In contrast, we find that our model configuration with a 12-month inference window (figure 4) consistently closes the water balance across consecutive 1, 12, and 60 month time horizons.

Furthermore, the inferred water balance components, and their uncertainties, may help identify months in which there is a need for additional information; perhaps in the form of expanded or improved monitoring infrastructure. Knowledge of how to either expand or consolidate monitoring infrastructure is critical to long-term understanding of hydrologic response, and the L2SWBM provides a potential pathway towards that understanding.

3.2. Inferred water balance component values, data bias, and data error

A visual assessment of a representative time series of inferred values of P, E, and R for Lake Erie from our model configuration with a 12-month rolling window (figure 5) indicates that while inferred estimates are generally consistent with historical data, there are also important differences both among the historical data sets and between those historical data sets and our new estimates. For example, we find that runo estimates from the GEM-MESH system (bottom panel, figure 5) tend to be systematically lower than those of the GLM-HMD in late winter and early spring. We also note that during periods when only one data source is available for a particular water balance component (i.e. Lake Erie evaporation in 2015, and Lake Erie runo in 2014 and 2015), inferred estimates have a higher degree of uncertainty. Summary statistics for each water balance component and Lake (table 3) underscore the relative contribution of each lake's water balance components, as well as their magnitudes relative to connecting channel flows.

It is informative to note that multiple additional models and data products could have been either assimilated into our application of the L2SWBM, or used as an independent basis for comparison with our new inferred water balance components (i.e. figure 5). The primary goal of this study, however, was to provide a robust demonstration of how the L2SWBM can close the water balance over multiple consecutive multi-month time steps. This demonstration provides a basis for future comparisons to (and perhaps assimilation of) those products. A recent study using preliminary results from the L2SWBM provides a representative example of this potential (Gronewold et al., 2019).

It is also worth noting that the larger uncertainty in bias for Lake Erie and Lake Ontario channel outflows reflect overall uncertainty in the water balance for both lakes. Erie and Ontario have roughly a quarter of the surface area of Superior and Michigan-Huron (table 1). Thus, less water is required to raise and lower the water level for both lakes, and uncertainties in other components of their water balances can be magnified. In the case of Erie and Ontario, channel flows are the dominant factor in the balance (table 3), and therefore any uncertainty in those estimates is magnified in the model. In contrast, the individual models for Lake Superior and Lake Michigan-Huron can absorb greater amounts of uncertainty in water balance components with respect to the water level and their surface areas.

An examination of inferred data bias and error estimates (figure 6) further underscores the ability of our framework to reconcile disparate historical data sets, and to close the water balance of a large lake system. For example, the bias and error results indicate that CaPA over-lake precipitation estimates tend to be positively biased relative to the overall water balance, particularly in winter months. These results are interesting in light of previous findings (Holman et al., 2012) suggesting that precipitation estimates across the Great Lakes based on terrestrial monitoring stations (such as those in the GLM-HMD) misrepresent winter atmospheric stability dynamics over large lakes and are therefore expected to show a strong seasonal bias as well.

We also find that both sources of legacy evaporation data (GLM-HMD LLTM and GEM-MESH) have seasonal biases (figure 6) relative to the water balance, with particularly severe biases in GEM-MESH evaporation estimates for Lakes Michigan-Huron and Ontario. This is not entirely surprising, given the challenge of accurately measuring (Blanken et al., 2011; Spence et al., 2011) and simulating (Fujisaki-Manome et al., 2017; Charusombat et al., 2018) turbulent heat fluxes across the vast surfaces of the Great Lakes. These challenges are particularly pronounced in the fall months; a period when evaporation rates increase rapidly, and when there can be significant year-to-year variability (Lenters, 2001; Spence et al., 2013).

4. Conclusions

We developed, tested, and applied a new Bayesian statistical analysis framework that reconciles the water balance of large lake systems. We hereafter propose formally referring to this product as the Large Lake Statistical Water Balance Model (or L2SWBM). Significant contributions to hydrological modeling represented by the new L2SWBM include explicit closure of the water balance over multiple time horizons through the use of a fixed-length rolling window, and a formulation for monthly model error distinct from water level measurement error and water balance component estimate uncertainty. We have also demonstrated how the L2SWBM can incorporate expert opinion through informative prior probability distributions on the bias in historical measurements of certain water balance components.

It is informative to note that our framework was recently adopted by Great Lakes regional management authorities, including the United States Army Corps of Engineers, and Environment and Climate Change Canada, as a step towards generating a new set of internationally-coordinated water balance component estimates for the entire Great Lakes system. It is our understanding that our framework is the first to provide a definitive approach to reconciling differences between water balance estimates for this system, and for closing the water balance over multiple time periods.

Moving forward, we anticipate applying the L2SWBM to other large lakes and large lake systems around the world. We recognize that, for many global lake systems, water balance data sets are based on sparse monitoring networks. In some cases, monitoring networks are nonexistent, and coarse model simulations are used to provide estimates of a lakes water balance components. The L2SWBM provides an ideal platform for utilizing *any* available information about a lake's water balance to reconcile changes in storage, and to explicitly allocate uncertainty and bias within historical data and water balance component estimates. As another potential future step in the evolution of the L2SWBM, we envision replacing water balance component terms (e.g. P, E, and R) with physically-based models. Potential examples include replacing the term for evaporation E in equation 1 with a lake surface energy balance model based on eddy diffusion (Hostetler and Bartlein, 1990), a formulation of Penman or Priestley-Taylor equations (Penman, 1948; Priestley and Taylor, 1972), or the Surface Energy Balance System (Su, 2002). This approach would allow direct approximation of parameters for those models that are faithful not only to governing physical processes and environmental observations, but to the water balance of a lake (or system of lakes) as well.

Data Availability Statement

The data that support the findings of this study are openly available from federal agencies including the National Oceanic Atmospheric Administration, Environment and Climate Change Canada, and the Coordinating Committee on Great Lakes Basic Hydraulic and Hydrologic Data.

Acknowledgements

The authors thank Song Qian, Yves Atchade, Kerby Shedden, Edward Ionides, Vincent Fortin, Bryan Tolson, Freya Rowland, and Craig Stow for helpful discussions on Bayesian inference and alternative formulations of our water balance model. Jacob Bruxer, Frank Seglenieks, Tim Hunter, Tim Calappi and Lauren Fry o ered expert opinions and facilitated access to publicly-available water balance data. Nicole Rice provided graphical and editorial support. Funding was provided by the International Joint Commission (IJC) International Watersheds Initiative (IWI) to NOAA and the Cooperative Institute for Great Lakes Research (CIGLR) through a NOAA Cooperative Agreement with the University of Michigan (NA12OAR4320071). The use of product names, commercial and otherwise, in this paper does not imply endorsement by NOAA, NOAA-GLERL, CIGLR, or any other contributing agency or organization. This is NOAA-GLERL contribution number XXXX and CIGLR contribution number ZZZZ.

Alsdorf, D., Birkett, C., Dunne, T., Melack, J., Hess, L., 2001. Water level changes in a large Amazon lake measured with spaceborn radar interferometry and altimetry. Geophysical Research Letters 28 (14), 2671–2674.

Arnell, N. W., 1999. A simple water balance model for the simulation of streamflow over a large geographic domain. Journal of Hydrology 217 (3), 314–335.

Blanken, P. D., Spence, C., Hedstrom, N., Lenters, J. D., 2011. Evaporation from Lake Superior: 1. Physical controls and processes. Journal of Great Lakes Research 37 (4), 707–716.

Borsuk, M. E., Clemen, R., Maguire, L., Reckhow, K. H., 2001. Stakeholder values and scientific modeling in the Neuse River watershed. Group Decision and Negotiation 10 (4), 355–373.

Casella, G., Berger, R. L., 2002. Statistical Inference. Duxbury, Pacific Grove, California.

Charusombat, U., Fujisaki-Manome, A., Gronewold, A. D., Lofgren, B. M., Anderson, E. J., Blanken, P. D., Spence, C., Lenters, J. D., Xiao, C., Fitzpatrick, L., Cutrell, G., 2018. Evaluating and improving modeled turbulent heat fluxes across the North American Great Lakes. Hydrology and Earth System Sciences 22 (10), 5559–5578.

Coccia, G., Siemann, A. L., Pan, M., Wood, E. F., 2015. Creating consistent datasets by combining remotely-sensed data and land surface model estimates through Bayesian uncertainty post-processing: The case of land surface temperature from HIRS. Remote Sensing of the Environment 170, 290–305.

Crétaux, J.-F., Jelinski, W., Calmant, S., Kouraev, A., Vuglinski, V. S., Bergé-Nguyen, M., Gennero, M. C., Nino, F., Del Rio, R. A., Cazenave, A., 2011. SOLS: A lake database to monitor in the near real-time water level and storage variations from remote sensing data. Advances in Space Research 47 (9), 1497–1507.

Croley II, T. E., 1989. Verifiable evaporation modeling on the Laurentian Great Lakes. Water Resources Research 25 (5), 781–792.

Croley II, T. E., 1992. Long-term heat storage in the Great Lakes. Water Resources Research 28 (1), 69–81.

Croley II, T. E., Hartmann, H. C., 1985. Resolving Thiessen polygons. Journal of Hydrology 76 (3–4), 363–379.

Crow, W. T., Kustas, W. P., Prueger, J. H., 2008. Monitoring root-zone soil moisture through the assimilation of a thermal remote sensing-based soil moisture proxy into a water balance model. Remote Sensing of the Environment 112 (4), 1268–1281.

Deacu, D., Fortin, V., Klyszejko, E., Spence, C., Blanken, P. D., 2012. Predicting the net basin supply to the Great Lakes with a hydrometeorological model. Journal of Hydrometeorology 13 (6), 1739–1759.

Dupont, F., Chittibabu, P., Fortin, V., Rao, Y. R., Lu, Y., 2012. Assessment of a NEMO-based hydrodynamic modelling system for the Great Lakes. Water Quality Research Journal of Canada 47 (3-4), 198–214.

Durnford, D., Fortin, V., Smith, G. C., Archambault, B., Deacu, D., Dupont, F., Dyck, S., Martinez, Y., Klyszejko, E., MacKay, M., Liu, L., 2018. Toward an operational water cycle prediction system for the Great Lakes and St. Lawrence River. Bulletin of the American Meteorological Society 99 (3), 521–546.

Fry, L. M., Hunter, T. S., Phanikumar, M. S., Fortin, V., Gronewold, A. D., 2013. Identifying streamgage networks for maximizing the e ectiveness of regional water balance modeling. Water Resources Research 49 (5), 2689–2700.

Fujisaki-Manome, A., Fitzpatrick, L., Gronewold, A. D., Anderson, E. J., Lofgren, B. M., Spence, C., Chen, J., Shao, C., Wright, D. M., Xiao, C., 2017. Turbulent heat fluxes during an extreme lake e ect snow event. Journal of Hydrometeorology 18 (2), 3145–3163.

Gelman, A. J., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Analysis 1 (3), 515–534.

Gelman, A. J., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004. Bayesian Data Analysis. Chapman & Hall/CRC, Boca Raton, Florida.

Gelman, A. J., Goegebeur, Y., Tuerlinckx, F., Van Mechelen, I., 2002. Diagnostic checks for discrete data regression models using posterior predictive simulations. Journal of the Royal Statistical Society Series C - Applied Statistics 49 (2), 247–268.

Geman, S., Bienenstock, E., Doursat, R., 1992. Neural networks and the bias/variance dilemma. Neural Computation 4 (1), 1–58.

Gibson, J. J., Prowse, T. D., Peters, D. L., 2006. Hydroclimatic controls on water balance and water level variability in Great Slave Lake. Hydrological Processes 20 (19), 4155–4172.

Gronewold, A. D., Anderson, E. J., Smith, J. P., 2019. Evaluating operational hydrodynamic models for real-time simulation of evaporation from large lakes. Geophysical Research Letters 46 (6), 3263–3269.

Gronewold, A. D., Bruxer, J., Durnford, D., Smith, J. P., Clites, A. H., Seglenieks, F., Qian, S. S., Hunter, T. S., Fortin, V., 2016. Hydrological drivers of record-setting water level rise on Earth's largest lake system. Water Resources Research 52 (5), 4026–4042.

Gronewold, A. D., Clites, A. H., Bruxer, J., Kompoltowicz, K., Smith, J. P., Hunter, T. S., Wong, C., 2015. Water levels surge on Great Lakes. Eos, Transactions American Geophysical Union 96 (6), 14–17.

Gronewold, A. D., Fortin, V., Caldwell, R., Noel, J., 2018. Resolving hydrometeorological data discontinuities along an international border. Bulletin of the American Meteorological Society 99 (5), 899–910.

Gronewold, A. D., Fortin, V., Lofgren, B. M., Clites, A. H., Stow, C. A., Quinn, F. H., 2013. Coasts, water levels, and climate change: A Great Lakes perspective. Climatic Change 120 (4), 697–711.

Gronewold, A. D., Stow, C. A., 2014. Water loss from the Great Lakes. Science 343 (6175), 1084–1085.

Gu, H., Jin, J., Wu, Y., Ek, M. B., Subin, Z., 2013. Calibration and validation of lake surface temperature simulations with the coupled WRF-lake model. Climatic Change 129 (3), 471–483.

Guo, S., Wang, J., Xiong, L., Ying, A., Li, D., 2002. A macro-scale and semi-distributed monthly water balance model to predict climate chagne impacts in China. Journal of Hydrology 268 (1), 1–15.

Holman, K. D., Gronewold, A. D., Notaro, M., Zarrin, A., 2012. Improving historical precipitation estimates over the Lake Superior basin. Geophysical Research Letters 39 (3), L03405.

Hostetler, S. W., Bartlein, P. J., 1990. Simulation of lake evaporation with application to modeling lake level variations of Harney-Malheur Lake, Oregon. Water Resources Research 26 (10), 2603–2612.

Hunter, T. S., Clites, A. H., Campbell, K. B., Gronewold, A. D., 2015. Development and application of a monthly hydrometeorological database for the North American Great Lakes - Part I: precipitation, evaporation, runo , and air temperature. Journal of Great Lakes Research 41 (1), 65–77.

Husak, G. J., Michaelsen, J., Funk, C., 2007. Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. International Journal of Climatology 27 (7), 935–944.

Kebede, S., Travi, Y., Alemayehu, T., Marc, V., 2006. Water balance of Lake Tana and its sensitivity to fluctuations in rainfall, Blue Nile basin, Ethiopia. Journal of Hydrology 316, 233–247.

Kelley, J. G. W., Anderson, E. J., Xu, J., 2018. Upgrade of NOS Lake Erie Operational Forecast System (LEOFS) to FVCOM: model development and hindcast skill assessment. NOAA Technical Memorandum, NOS CS 40. Tech. rep.

Kéry, M., 2010. Introduction to WinBUGS for ecologists: A Bayesian approach to regression, ANOVA, mixed models and reltaed analyses. Academic Press.

Kim, C. P., Stricker, J., 1996. Influence of spatially variable soil hydraulic properties and rainfall intensity on the water budget. Water Resources Research 32 (6), 1699–1712.

Kruschke, J. K., 2013. Posterior predictive checks can and should be Bayesian: comment on Gelman and Shalizi, ‘Philosophy and the practice of Bayesian statistics’. British Journal of Mathematical and Statistical Psychology 66 (1), 45–56.

Lenters, J. D., 2001. Long-term trends in the seasonal cycle of Great Lakes water levels. Journal of Great Lakes Research 27 (3), 342–353.

Lespinas, F., Fortin, V., Roy, G., Rasmussen, P., Stadnyk, T., 2015. Performance evaluation of the Canadian Precipitation Analysis (CaPA). Journal of Hydrometeorology 16 (5), 2045–2064.

Li, X.-Y., Xu, H.-Y., Sun, Y.-L., Zhang, D.-S., Yang, Z.-P., 2007. Lake-level change and water balance analysis at Lake Qinghai, west China during recent decades. Water Resources Management 21 (9), 1505–1516.

Lunn, D. J., Spiegelhalter, D. J., Thomas, A., Best, N. G., 2009. The BUGS project: Evolution, critique and future directions. Statistics in Medicine 28 (25), 3049 – 3067.

Lunn, D. J., Thomas, A., Best, N. G., Spiegelhalter, D. J., 2000. WinBUGS-A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing 10 (4), 325–337.

Mainville, A., Craymer, M. R., 2005. Present-day tilting of the Great Lakes region based on water level gauges. Geological Society of America Bulletin 117 (7), 1070–1080.

Makhlouf, Z., Michel, C., 1994. A two-parameter monthly water balance model for French watersheds. Journal of Hydrology 162 (3), 299–318.

Milly, P. C., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., Stou er, R. J., 2008. Stationarity is dead: Whither water management? Science 319 (5863), 573–574.

Milly, P. C., Dunne, K. A., 2017. A hydrologic drying bias in water-resource impact analyses of anthropogenic climate change. JAWRA Journal of the American Water Resources Association.

Nicholson, S. E., Yin, X., Ba, M. B., 2000. On the feasibility of using a lake water balance model to infer rainfall: an example from Lake Victoria. Hydrological Sciences Journal 45 (1), 75–95.

Nijssen, B., O'Donnell, G. M., Lettenmaier, D. P., Lohmann, D., Wood, E. F., 2001. Predicting the discharge of global rivers. Journal of Climate 14 (15), 3307–3323.

Notaro, M., Holman, K. D., Zarrin, A., Fluck, E., Vavrus, S. J., Bennington, V., 2013. Influence of the Laurentian Great Lakes on regional climate. Journal of Climate 26 (3), 789–804.

Ntzoufras, I., 2011. Bayesian modeling using WinBUGS. John Wiley & Sons.

Pan, M., Wood, E. F., 2006. Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter. Journal of Hydrometeorology 7 (3), 534–547.

Parkes, B., Demeritt, D., 2016. Defining the hundred year flood: A Bayesian approach for using historic data to reduce uncertainty in flood forecasting estimates. Journal of Hydrology 540, 1189–1208.

Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A. S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540 (7633), 418.

Peng, Z., Hu, W., Liu, G., Wei, W., 2019. Estimating daily inflows of large lakes using a water balance based runo coefficient scaling approach. Hydrological Processes.

Penman, H. L., 1948. Natural evaporation from open water, bare soil and grass. Proceedings of the Royal Society of London Series {A}– Mathematical and Physical Sciences 193 (1032), 120–145.

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., Verseghy, D., Soulis, E. D., Caldwell, R., Evora, N., 2007. Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. Hydrology and Earth System Sciences 11 (4), 1279–1294.

Piper, B. S., Plinston, D. T., Sutcliffe, J. V., 1986. The water balance of Lake Victoria. Hydrological Sciences Journal 31 (1), 25–37.

Plummer, M., 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd international workshop on distributed statistical computing. Technische Universit at Wien, p. 125.

Press, S. J., 2003. Subjective and Objective Bayesian Statistics: Principles, Models, and Applications. Wiley-Interscience, Hoboken, NJ.

Priestley, C. H. B., Taylor, R. J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Monthly Weather Review 100, 81–92.

Qian, S. S., Craig, J. K., Baustian, M. M., Rabalais, N. N., dec 2009. A Bayesian hierarchical modeling approach for analyzing observational data from marine ecological studies. Marine pollution bulletin 58 (12), 1916–1921.

Quinn, F. H., Edstrom, J., 2000. Great Lakes diversions and other removals. Canadian Water Resources Journal 25 (2), 125–151.

Quinn, F. H., Guerra, B., 1986. Current perspectives on the Lake Erie water balance. Journal of Great Lakes Research 12 (2), 109–116.

R core team, 2017. R: A language and environment for statistical computing. URL `http://www.r-project.org`

Raes, D., Geerts, S., Kipkorir, E., Wellens, J., Sahli, A., 2006. Simulation of yield decline as a result of water stress with a robust soil water balance model. Agricultural Water Management 81 (3), 335–357.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., 2014. Multimodel assessment of water scarcity under climate change. Proceedings of the National Academy of Sciences of the United States of America 111 (9), 3245–3250.

Schwatke, C., Dettmering, D., Bosch, W., Seitz, F., 2015. DAHITI - an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. Hydrology and Earth System Sciences 19 (10), 4345–4364.

Senay, G. B., Leake, S., Nagler, P. L., Artan, G., Dickinson, J., Cordova, J. T., Glenn, E. P., 2011. Estimating basin scale evapotranspiration (ET) by water balance and remote sensing methods. Hydrological Processes 25 (26), 4037–4049.

Smith, J. P., Gronewold, A. D., 2018. Development and analysis of a Bayesian water balance model for large lake systems. Tech. rep.

Spence, C., Blanken, P. D., Hedstrom, N., Fortin, V., Wilson, H., 2011. Evaporation from Lake Superior: 2: Spatial distribution and variability. Journal of Great Lakes Research 37 (4), 717–724.

Spence, C., Blanken, P. D., Lenters, J. D., Hedstrom, N., 2013. The importance of spring and autumn atmospheric conditions for the evaporation regime of Lake Superior. Journal of Hydrometeorology 14 (5), 1647–1658.

Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat flux. Hydrology and Earth System Sciences 6, 85–99.

Swenson, S., Wahr, J., 2009. Monitoring the water balance of Lake Victoria, East Africa, from space. Journal of Hydrology 370 (1-4), 163–176.

Thom, H. C., 1958. A note on the gamma distribution. Monthly Weather Review 86 (4), 117–122.

Valentini, G., Dietterich, T. G., 2004. Bias-variance analysis to support vector machines for the development of SVM-based ensemble methods. Journal of Machine Learning Research 5, 725–775.

Voinov, A. A., Bousquet, F., 2010. Modelling with stakeholders. Environmental Modelling & Software 25 (11), 1268–1281.

Vörösmarty, C. J., Federer, C. A., Schloss, A. L., 1998. Potential evaporation functions compared on US watersheds: Possible implications for global-scale water balance and terrestrial ecosystem modeling. Journal of Hydrology 207 (3), 147–169.

Vörösmarty, C. J., Green, P., Salisbury, J., Lammers, R. B., 2000. Global water resources: vulnerability from climate change and population growth. Science 289 (5477), 284–288.

Xiao, C., Lofgren, B. M., Wang, J., Chu, P. Y., 2016. Improving the lake scheme within a coupled WRF-lake model in the Laurentian Great Lakes. Journal of Advances in Modeling Earth Systems 8 (4), 1969–1985.

Xiao, K., Griffis, T. J., Baker, J. M., Bolstad, P. V., Erickson, M. D., Lee, X., Wood, J. D., Hu, C., Nieber, J. L., 2018. Evaporation from a temperate closed-basin lake and its impact on present, past, and future water level. Journal of Hydrology 561, 59–75.

Xu, C.-Y., Singh, V. P., 1998. A review on monthly water balance models for water resources investigations. Water Resources Management 12 (1), 20–50.

Figure 1: The Laurentian Great Lakes basin (shaded region) including location of major cities, interbasin diversions, and connecting channels (Source: NOAA-GLERL and USACE-Detroit).

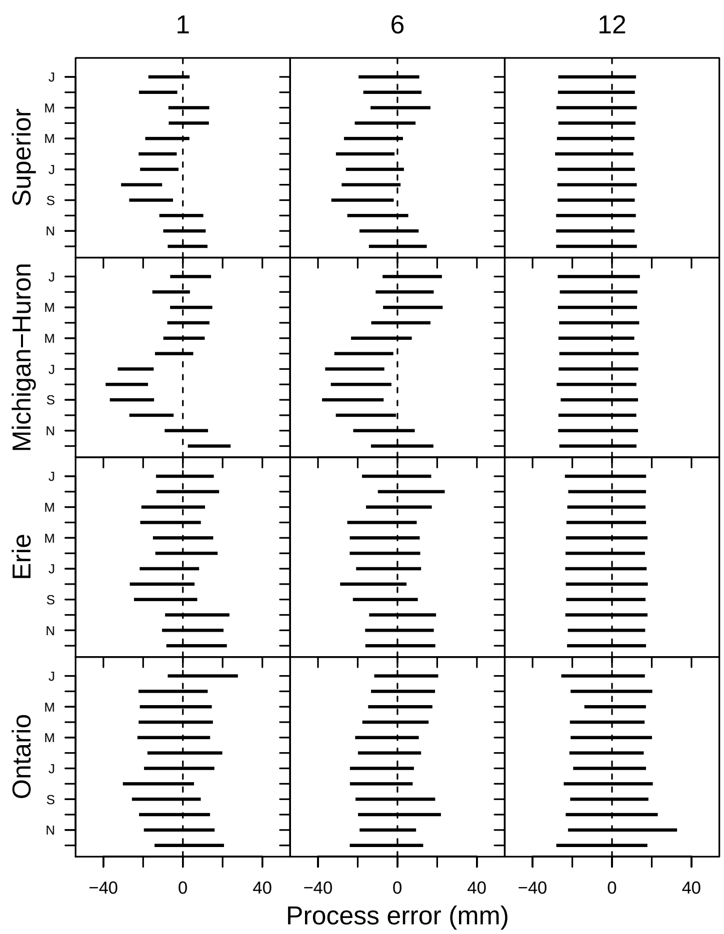

Figure 2: 95% credible intervals for model process errors ϵ_{c_t} from model configurations with a 1-month inference window (left), 6-month inference window (center), and a 12-month inference window (right).

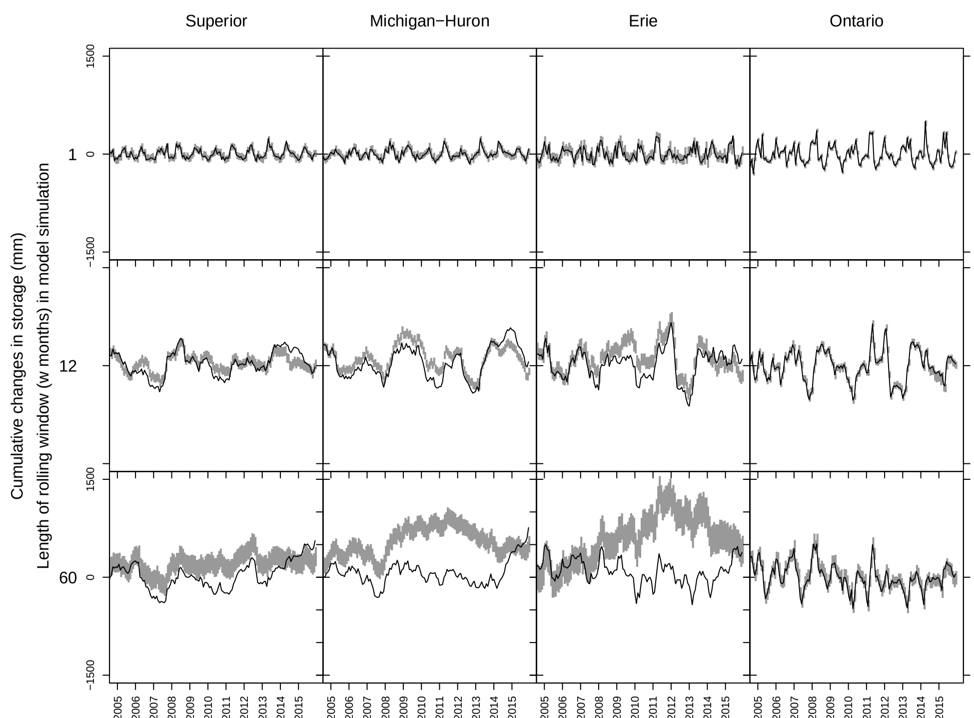

Figure 3: Water balance closure assessment for the version of our model configured with a 1-month inference window. Simulated changes in storage from the model are presented as grey bands (95% posterior predictive intervals) over cumulative one month (top row), 12 month (middle row), and 60 month periods. Observed cumulative changes in storage over corresponding time horizons are represented by a black line.

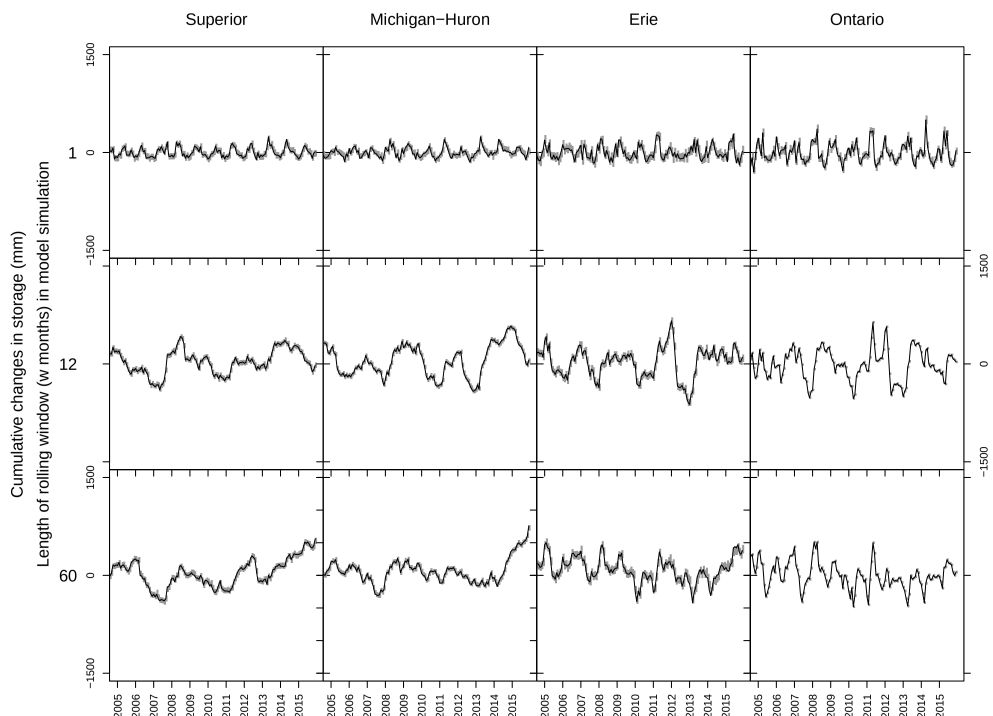

Figure 4: Water balance closure assessment for the version of our model configured with a 12-month inference window. Simulated changes in storage from the model are presented as grey bands (95% posterior predictive intervals) over cumulative one month (top row), 12 month (middle row), and 60 month periods. Observed cumulative changes in storage over corresponding time horizons are represented by a black line. Vertical axis scale in each panel is the same as in figure 3 to facilitate comparison.

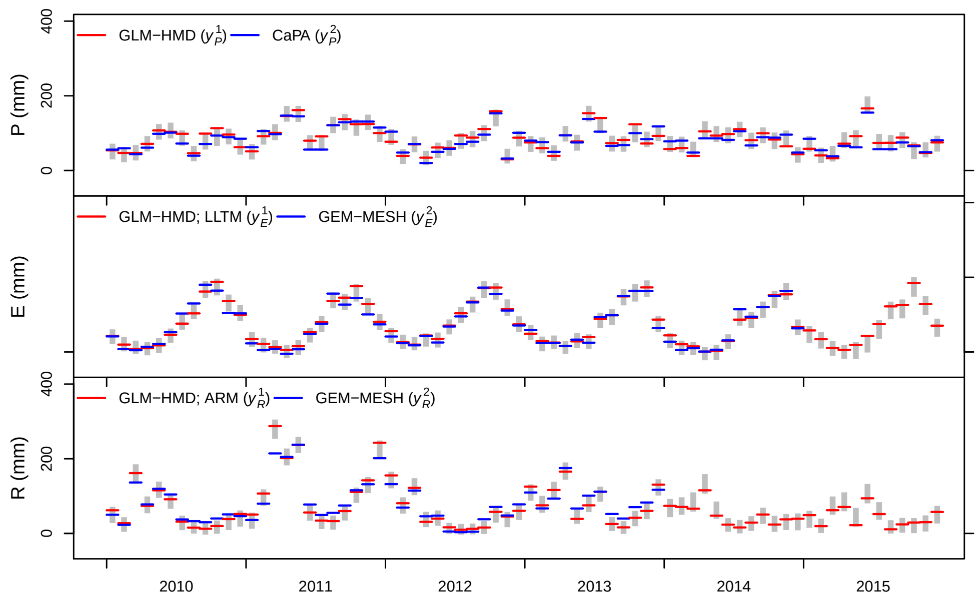

Figure 5: Comparison between newly-derived water balance component estimates from our model for Lake Erie configured with a 12-month inference window (vertical grey bars representing 95% credible intervals) and corresponding observations from legacy regional data records and models (blue and red horizontal dashes). Results are presented only for years 2010 through 2015 for clarity. Additional results are included in online Supplementary Material.

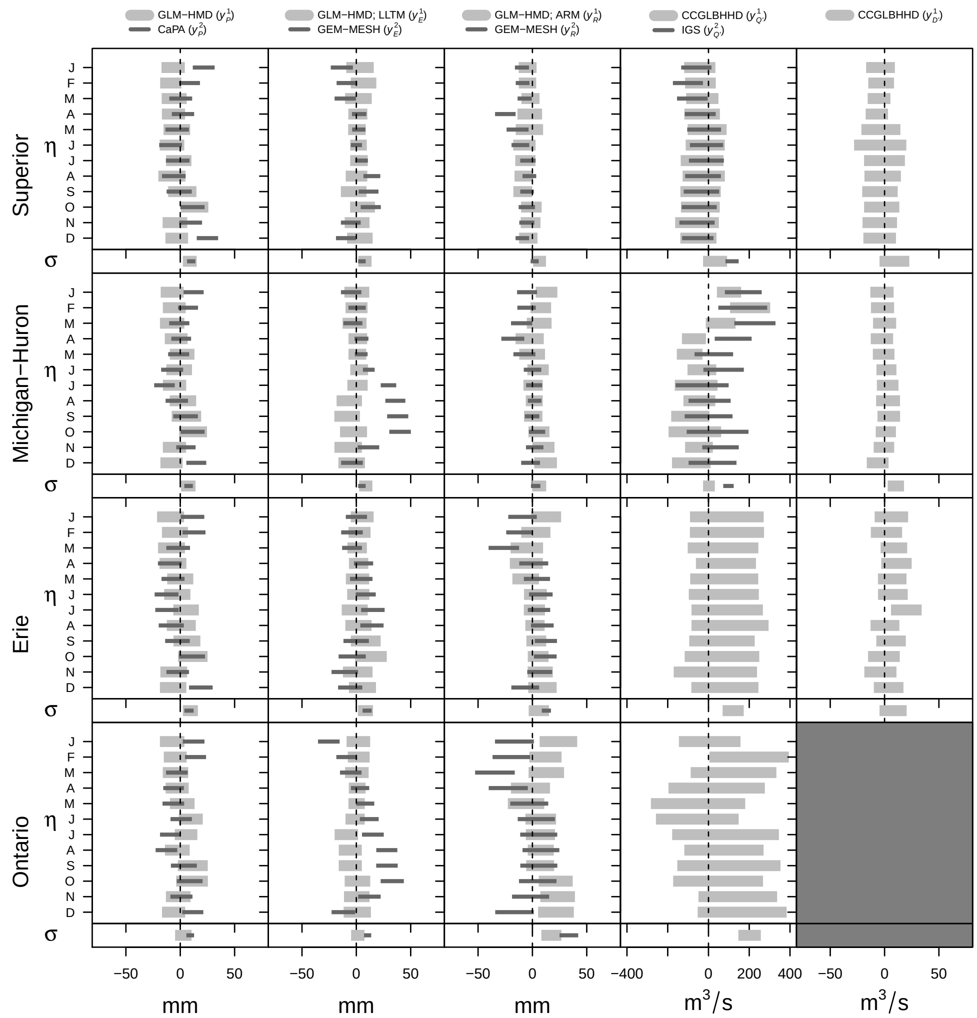

Figure 6: Inferred bias (η) and error (σ) in historical water balance component data based a configuration of our model with a 12-month inference window. Columns, from left to right, correspond to monthly data for over-lake precipitation, over-lake evaporation, lateral tributary runo , outflow, and interbasin diversions. Horizontal bars represent 95% credible intervals. For details on data sources, see table A1.

Lake	Surface Area (km^2)
Superior	81,925
Michigan-Huron	116,850
St. Clair	1,109
Erie	25,404
Ontario	19,121

Table 1: Surface areas for each of the Laurentian Great Lakes (for details, see Hunter et al., 2015).

Channel Flow or Diversion	λ	Mean (cms)	SD (cms)
Superior Outflow (St. Marys River)	0.02	2,111	494
Superior Diversion (in, via Ogoki & Long-Lac)	0.04	145	83
Michigan-Huron Outflow (St. Clair River)	0.03	5,189	636
Michigan-Huron Diversion (out, via Chicago)	0.04	133	66
St. Clair Outflow (Detroit River)	0.03	5,323	637
Erie Outflow (Niagara)	0.02	5,784	666
Erie Diversion (Welland Canal)	0.04	159	80
Ontario Outflow (St. Lawrence River)	0.02	6,949	931

Table 2: Coefficients of variation (λ), and historical mean and standard deviation of monthly average values used in empirical estimation of prior standard deviation on the bias in historical data sources for channel flows and diversions on the Great Lakes. Channel flow estimates are from data collected from 1900 through 2010. Chicago River diversion estimates are based on data available from 1900 through 2008, Ogoko & Long-Lac diversion estimates are based on data available from 1939 to 2012, and the Welland Canal diversion estimates are based on data from 1900 through 2012.

	Average annual values, 1980-2015 (in mm over lake surface area)			
	P	E	R	Q
Superior	766.97	540.52	621.14	808.43
Michigan-Huron	823.29	535.72	689.99	1441.06
Erie	891.67	854.14	842.70	7454.20
Ontario	854.57	669.57	1664.14	12102.40

Table 3: Average annual totals for major water balance components on each of the Great Lakes from 1980 to 2015 based on our new L2SWBM results.

APPENDIX A - Data

This Appendix includes a summary (Table A1) of data sources for populating water balance component "observations" (y) and for calculating prior probability distribution hyperparameters.

Variable or parameter	Data source and reference(s)	Years used
$y_{\Delta H}$	CCGLBHHD (Gronewold et al., 2018)	1980 - 2015
y_P^1	GLM-HMD (Hunter et al., 2015)	1980 - 2015
y_P^2	CaPA (Lespinas et al., 2015)	2006 - 2015
μ_P, $\mu_{ln(P)}$	GLM-HMD (Hunter et al., 2015)	1950 - 1979
y_E^1	GLM-HMD; LLTM (Hunter et al., 2015)	1980 - 2015
y_E^2	GEM-MESH (Deacu et al., 2012)	2004 - 2014
μ_E	GLM-HMD; LLTM (Hunter et al., 2015)	1950 - 1979
y_R^1	GLM-HMD; ARM (Hunter et al., 2015)	1980 - 2015
y_R^2	GEM-MESH (Superior and Michigan-Huron)	2004 (June) - 2009
y_R^2	GEM-MESH (Erie and Ontario)	2004 (June) - 2013
$\mu_{ln(R)}$	GLM-HMD; ARM (Hunter et al., 2015)	1950 - 1979
$y_{NBS'}^1$	GLM-HMD (Hunter et al., 2015)	1980 - 2015
$y_{NBS'}^2$	GEM-MESH (Deacu et al., 2012)	2004 (June) - 2012
$y_{NBS'}^3$	CCGLBHHD Residual (Gronewold et al., 2018)	1980 - 2015
$\mu_{NBS'}$	CCGLBHHD Residual (Gronewold et al., 2018)	1950 - 1979
$y_{Q'}^1$	CCGLBHHD (Gronewold et al., 2018)	1980 - 2015
$y_{Q'}^2$	IGS (for St. Marys, St. Clair, and Detroit Rivers only)	2008 (Nov) - 2014
$\mu_{Q'}$	CCGLBHHD (Gronewold et al., 2018)	1950 - 1979
$y_{D'}$	CCGLBHHD (Gronewold et al., 2018)	1980 - 2015
$\mu_{D'}$	CCGLBHHD (Gronewold et al., 2018)	1950 - 1979

Table A1: Summary of data sets used in our study. Unless indicated otherwise, date ranges include the entire calendar year. Variable definitions are included in table B1.

APPENDIX B - Notation

A summary of notation used in our study is included in table B1.

Symbol	Description
Indices and related variables	
$c(t)$	Calendar month $c \in [1, 12]$ of time step t
i	Index for months within a water balance window of length w
j	Index $j \in [1, T - w + 1]$ for the first month of a rolling window
l	Index for an individual lake; $l \in [SUP, MHU]$
n	Index of data sources for a particular water balance component; $n \in [1, N]$
N	Total number of data sources for a water balance component (in this study, typically 2)
t	Index for month number in the sequence $[1, T]$
T	Total number of months in study. Here, $T = 120$ (January 2005 through December 2014)
w	Length of rolling window (in months) for water balance inference
"True" (unobserved) monthly average water balance components (all in mm over lake surface)	
$\Delta H_{l,j,w}$	Change in water level for lake l from beginning of month j to beginning of month $j + w$
$D_{l,t}$	Monthly diversion from lake l in month t
$E_{l,t}$	Evaporation from lake l in month t
$I_{l,t}$	Connecting channel inflow for lake l in month t
$P_{l,t}$	Precipitation over lake l in month t
$Q_{l,t}$	Connecting channel outflow for lake l in month t
$R_{l,t}$	Basin runo into lake l in month t
θ	A parameter from Θ
Θ	The parameter set P, E, R, Q, D
Data and model-based estimates (subscript l removed from each for clarity)	
$y_{H,t}$	Water level measurement at beginning of month t
$y_{\Delta H,j,w}$	Observed water level difference from beginning of month j to beginning of month $j + w$
$y^n_{\theta,t}$	n^{th} estimate of θ_t
Prior probability distribution parameters (subscripts l and c(t) removed from each for clarity)	
μ_E, μ_Q, μ_D	Historical empirical mean of E, Q, and D from the GLM-HMD
$\mu_{ln(R)}$	Historical empirical log-mean of R from the GLM-HMD
τ_E, τ_Q, τ_D	Historical empirical precision of E, Q, and D from the GLM-HMD
$\tau_{ln(R)}$	Historical empirical precision of natural logarithm of R from GLM-HMD
ψ^1, ψ^2	Shape and rate parameters for $\pi(P)$
Hyperparameters (subscripts l removed for clarity)	
$\epsilon_t = \epsilon_{c(t)}$	Water balance model process error for calendar month $c(t)$
$\eta^n_{\theta,c(t)}$	Seasonal bias of $y^n_{\theta,t}$ by calendar month $c(t)$
$\tau_{\Delta H,w}$	Precision of $y_{\Delta H,j,w}$ for all j
τ^n_θ	Precision of $y^n_{\theta,t}$ (for all t)

Table B1: Summary of notation used in our study.