Quantifying data uncertainty and bias in a Bayesian model for large lake systems

H43D-2413

Joeseph P. Smith[1], Andrew D. Gronewold[2], Scott Steinschneider[3]

[1]Cooperative Institute for Great Lakes Research, U. of Michigan [2]NOAA Great Lakes Environmental Research Laboratory, [3]Department of Biological and Environmental Engineering, Cornell U.

Abstract

There is a growing need for water balance models that correctly attribute water demand to water use categories, anthropogenic controls, and the effects of climate change. Addressing this need requires explicit quantification of bias and uncertainty in model inputs. Here, we introduce recent advancements in a Bayesian water balance model for large lake systems that not only quantifies bias and uncertainty in multiple data sources across time, but demonstrates that doing so improves reconciliation of the regional water balance over multiple time horizons; an objective not often achieved in historical long-term water balance models. We present a case study in which a new version of the model is applied across the entire Laurentian Great Lakes system, and view this work as a stepping stone towards application to large lake systems around the world. We find that, for the Great Lakes, there are legacy data sources with severe seasonal biases that have historically propagated into regional lake water level management decisions and operational protocols. We anticipate that explicitly acknowledging and correcting these biases will lead to more accurate water balance component estimates, and a more robust basis for water management decisions conditioned on those estimates.

Study Area

Figure 1 – Site map of the Great lakes basin. The light brown region indicates the boundary of the Great Lakes basin.

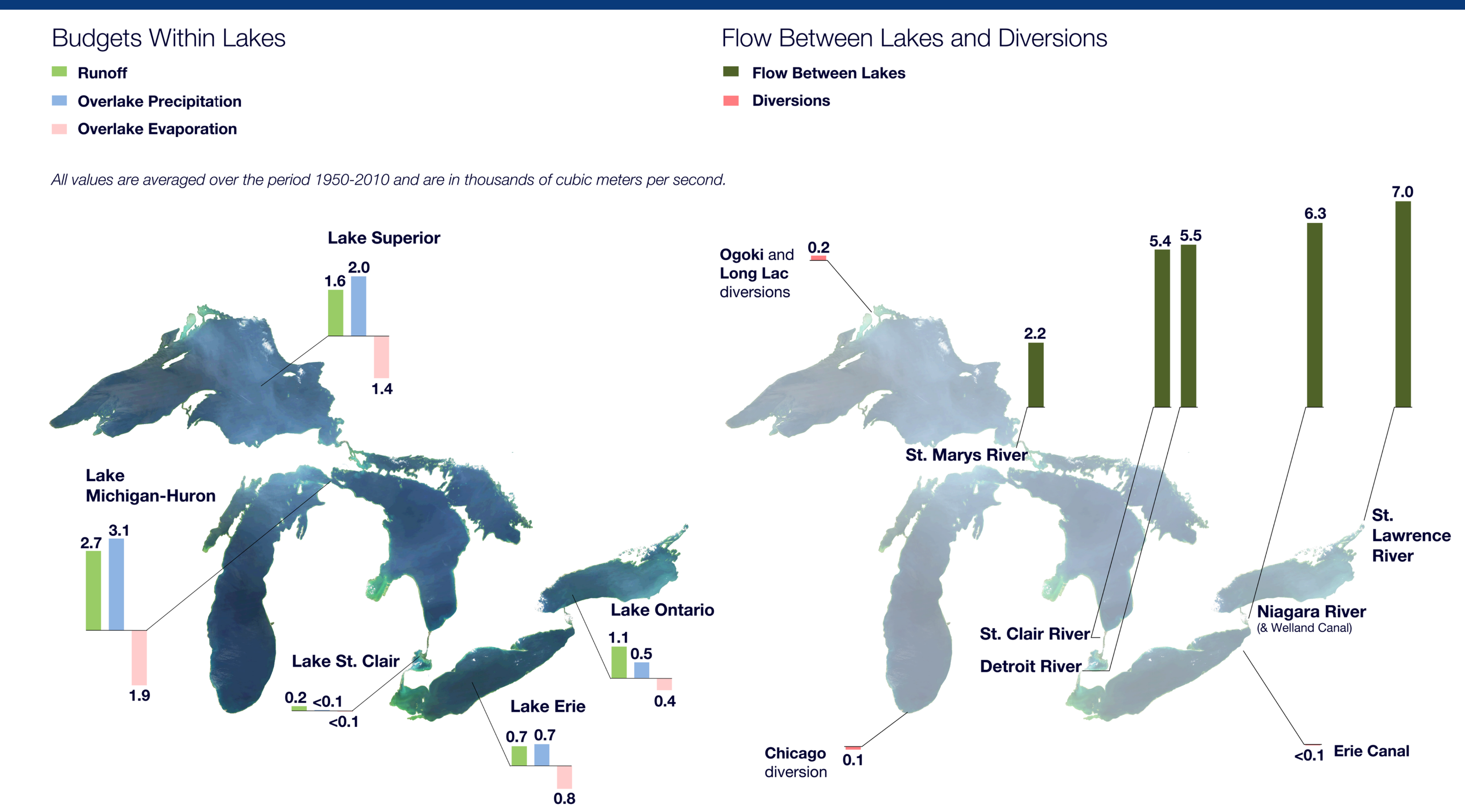

Figure 2 – Annual average water balance (in 1000s of cms) of the Great Lakes system, including magnitudes of lateral tributary inflow, over-lake precipitation, over-lake evaporation, interbasin diversions, and connecting channel flows.

We applied our model to the Laurentian Great Lakes (Figure 1), the largest system of lakes on Earth (Lake Superior alone is the largest lake on Earth by surface area). The Great Lakes are connected through channels (Figure 2) that propagate outflows from Lake Superior through Lake Michigan, Huron, Erie, and Ontario; continental scale outflows from Lake Ontario pass through the St. Lawrence River en route to the Gulf of St. Lawrence.

The Model

The proposed model uses a rolling time window (of length w, in months) over which observed changes in lake storage (ΔH) across a w month period are equated to the cumulative sum of water balance components over the same period.

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

Water Balance Components (all representing monthly totals, in mm over water surface)

P	=	Over-lake precipitation
E	=	Over-lake evaporation
R	=	Lateral tributary lake inflow (i.e. runoff)
I	=	Inflow from upstream lake (via connecting channel)
Q	=	Outflow to downstream lake (via connecting channel)
D	=	Flow through interbasin diversions
ϵ	=	Model error term

Data assimilation (Bayesian likelihood functions)

Historical estimates of water balance components (denoted by *y*, see Table 1), including changes in lake storage, are incorporated via likelihood functions in a Bayesian framework.

Likelihood function for changes in lake storage:

$$y_{\Delta H,j,w} \sim \mathsf{N}(\Delta H_{j,w}, \tau_{\Delta H,w})$$

Likelihood function(s) for water balance components (represented generically by θ):

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

Table 1 – Summary of data sources used to construct likelihood functions.

Variable	Data source and reference(s)	Years used
$y_{\Delta H}$	CCGLBHHD (Gronewold et al., 2018)	2005 - 2014
y^1_P	GLM-HMD (Hunter et al., 2015)	1950 - 2014
y^2_P	CaPA (Lespinas et al., 2015)	2005 - 2014
y^1_E	GLM-HMD; LLTM (Hunter et al., 2015)	1950 - 2014
y^2_E	GEM-MESH (Deacu et al., 2012)	2005 - 2014
y^1_R	GLM-HMD; ARM (Hunter et al., 2015)	1950 - 2014
y^2_R	NOAA-GLERL LBRM (Croley II, 1983)	1950 - 2014
y^1_Q	CCGLBHHD (Gronewold et al., 2018)	1950 - 2014
y^2_Q	IGS	2008 (Nov) - 2014
y_D	CCGLBHHD (Gronewold et al., 2018)	1950-2014

Prior probability distributions

Parameters for prior probability distributions for each water balance component are estimated empirically (Figure 3). For example, monthly total over-lake precipitation is modelled with a gamma prior probability distribution:

$$\pi(P_t) = \mathsf{Ga}(\psi^1_{c(t)}, \psi^2_{c(t)})$$

With shape ψ^1 and rate ψ^2 parameters calculated empirically (following Thom, 1958) using GLM-HMD values from 1950-2004:

$$\psi^1_{c(t)} = \frac{1}{4\phi_{c(t)}}\left(1 + \sqrt{1 + \frac{4\phi_{c(t)}}{3}}\right)$$

$$\phi_{c(t)} = \ln(\mu'_{P,c(t)}) - \mu'_{ln(P),c(t)}$$

$$\psi^2_{c(t)} = \psi^1/\mu'_{P,c(t)}$$

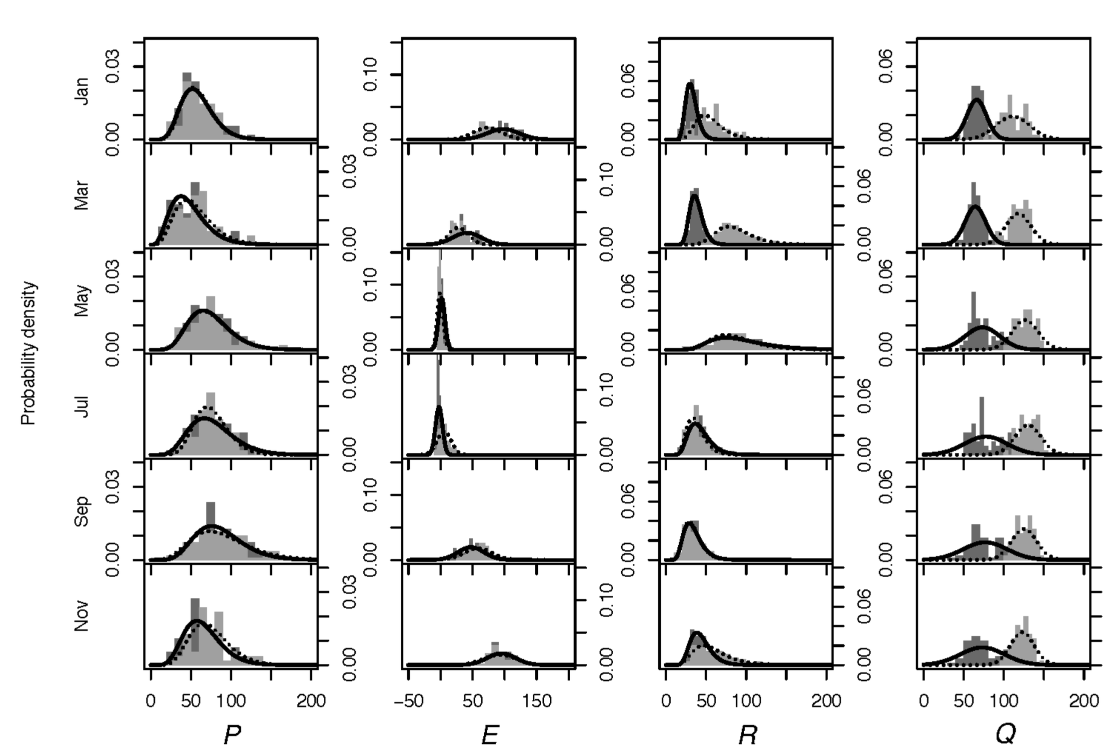

Figure 3 – Prior probability distributions (only odd-numbered months shown for simplicity). Probability distribution parameters are estimated empirically from historical estimates between 1900 and 1950. Dark-colored histograms (and corresponding solid lines) represent Lake Superior, and light-colored histograms (with corresponding dashed lines) represent Lake Michigan-Huron.

Results

Our Bayesian model yielded a new set of water balance estimates (Figure 4) that reconcile not only differences between historical water balance component estimates, but also changes in lake storage across the entire Great Lakes system over multiple time periods.

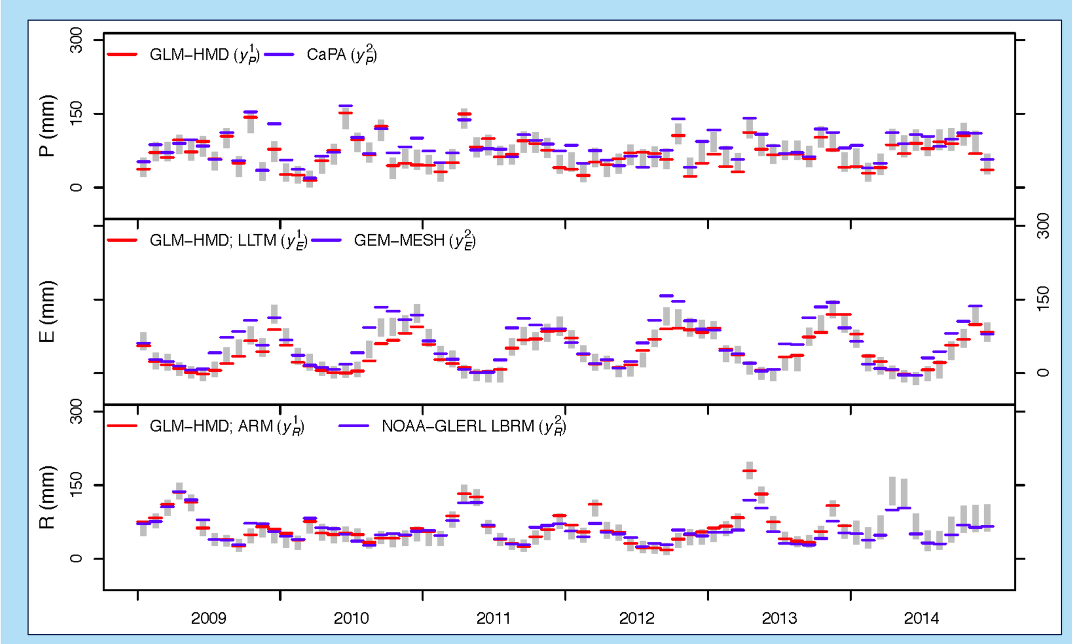

Figure 4 – Representative results for Lake Michigan-Huron with new water balance estimates (grey 95% credible intervals) compared to deterministic historical (legacy) estimates (blue and red dashes). The new model results underscore the importance of quantifying uncertainty when there are limited reference data sets, or when data sets (including observed changes in lake storage) are divergent.

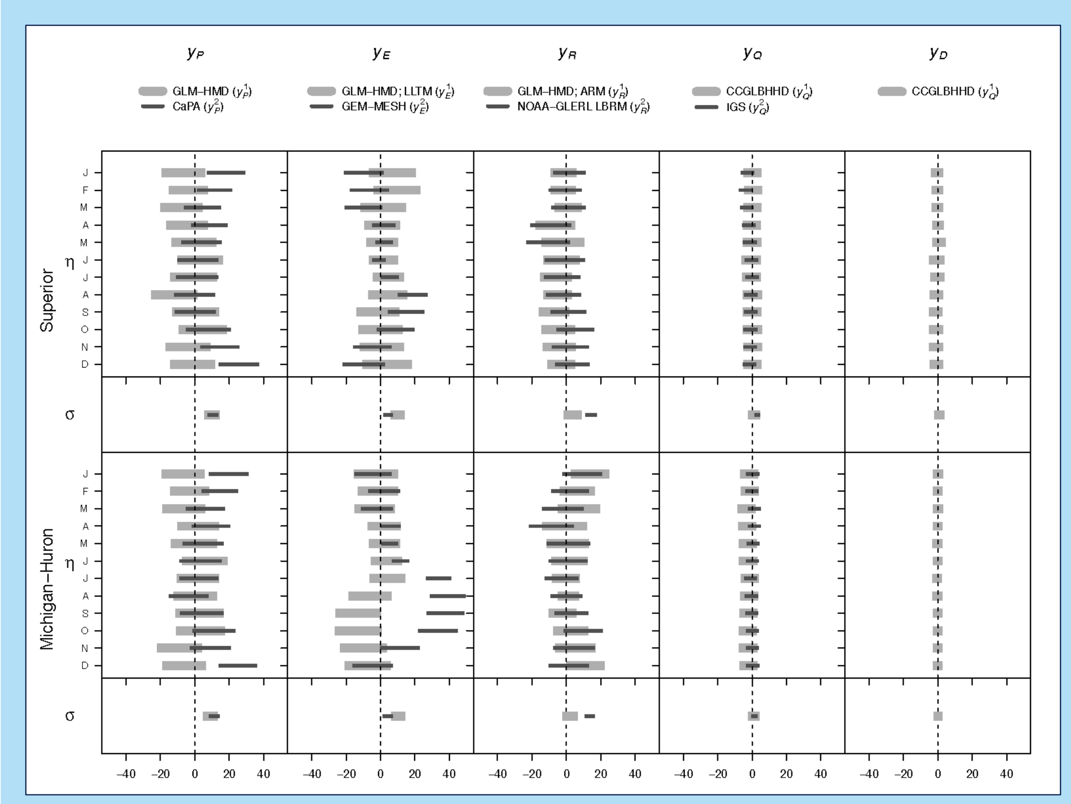

Figure 6 – Inferred bias and error estimates in historical water balance data.

We assess fidelity of our new estimates relative to long-term changes in lake storage by using the new estimates to simulate changes in lake storage over 1, 12 and 60 month periods (Figure 5)

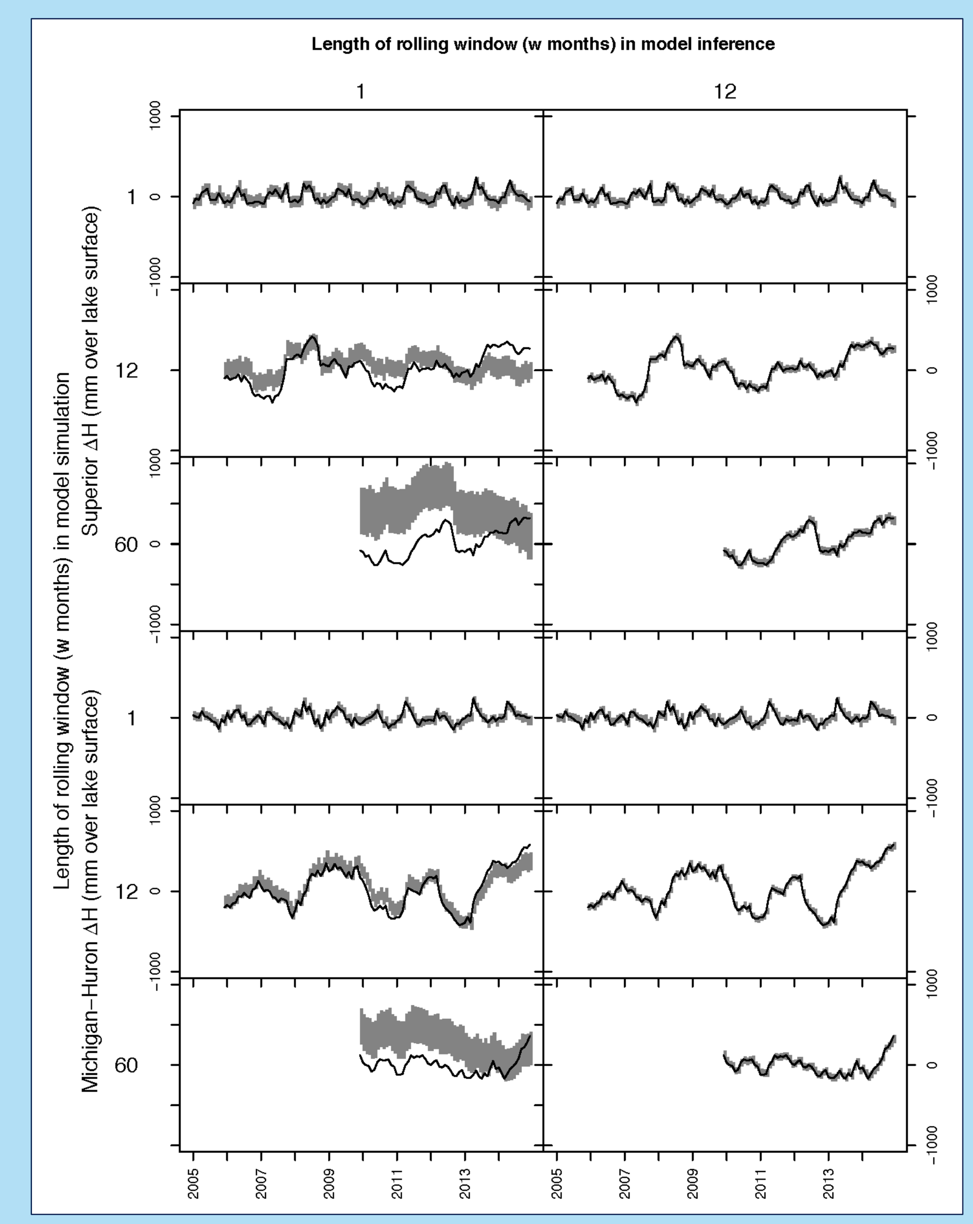

Figure 5 – Observed (black line) and simulated (grey 95% credible intervals) changes in lake storage over periods of 1, 12, and 60 months. Model results are presented for rolling inference windows of 1 month (left column) and 12 months (right column).

One of the most intriguing results from the L2SWBM is the estimation of bias and error for each of the historical data sets (Figure 6). We find, for example, that there are strong seasonal biases in estimates of over-lake evaporation for Lake Michigan-Huron. Importantly, the L2SWBM not only identifies these biases, but generates new water balance component estimates that accommodate them.

Conclusions and Future Work

We have presented a hierarchical Bayesian model that can integrate available observations and their uncertainties to close the water balance for the Laurentian Great Lakes. These estimates can be used to better ascertain the hydroclimatic drivers of water level variability in the Great Lakes system. In the future, we intend to use this Bayesian model to test how improved estimates and uncertainty quantification of specific components of the water balance (e.g., regional tributary inflows, over-lake evaporation) propagate into the overall quantification of all other water balance components.

In this way, the Bayesian model can reveal the value of additional information from new measurement and estimation techniques. In addition, new structural features will be tested to further improve the representation of uncertainty in the model, e.g., an accounting of spatial and temporal autocorrelation in different water level component estimates. Finally, future efforts will extend this model to other large lake systems (e.g., African Great Lakes) to help determine the hydroclimatic causes of water level fluctuations, which can have outsized impacts on communities in those regions.

Acknowledgements & References

Kaye Lafond and Nicole Rice provided graphical and editorial support. This project was supported with funding from the International Joint Commission (IJC) through the International Watersheds Initiative (IWI). Funding was also provided by NOAA.

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

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, runoff, and air temperature. *Journal of Great Lakes Research*, 41(1), 65–77.

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