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

Josephe P. Smith1, Andrew D. Gronewold2, Scott Steinschneider3
1Cooperative Institute for Great Lakes Research, University of Michigan, NOAA Great Lakes Environmental Research Laboratory, 2Department 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 lakes 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.

The Model

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

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

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

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

Data assimilation (Bayesian likelihood functions)

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

Likelihood function for changes in lake storage:

\[ y_{\Delta H,j,w} \sim N(\Delta H_{j,w}^\beta, \tau \Delta H_{j,w}) \]

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Data source and reference(s)</th>
<th>Years used</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCGHHD</td>
<td>(Gronewold et al., 2018)</td>
<td>2005 - 2015</td>
</tr>
<tr>
<td>CCGHHD</td>
<td>(Hunter et al., 2015)</td>
<td>2016 - 2017</td>
</tr>
<tr>
<td>GLM-HM</td>
<td>(Nondahl et al., 2012)</td>
<td>2005 - 2014</td>
</tr>
<tr>
<td>GLM-HM</td>
<td>(Hunter et al., 2015)</td>
<td>2016 - 2017</td>
</tr>
<tr>
<td>NOAA-GLERL</td>
<td>(Moore et al., 2014)</td>
<td>2005 - 2014</td>
</tr>
<tr>
<td>RCA</td>
<td>(Gronewold et al., 2015)</td>
<td>2015-2016</td>
</tr>
<tr>
<td>CCGHHD</td>
<td>(Gronewold et al., 2018)</td>
<td>2015-2016</td>
</tr>
</tbody>
</table>

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 modeled with a gamma prior probability distribution:

\[ \pi(P) = Ga(\psi(P), \varphi(P)) \]

With shape \( \psi \) and rate \( \varphi \) parameters calculated empirically (following Thorn, 1982) using GLM-HM values from 1990-2004:

\[ \psi(P) = \frac{1}{4\varphi(P)} \left( 1 + \sqrt{1 + 4\varphi(P)^2} \right) \]

\[ \varphi(P) = \ln(\mu(P), \sigma(P)) \]

\[ \psi(P) = \psi(1/\mu(P)) \]

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.

Our model yielded a new set of water balance estimates across the entire 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.

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 hydrologic 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 hydrologic causes of water level fluctuations, which can have outsized impacts on communities in those regions.