Thank you for your reply to reviewer comments and manuscript revisions. I believe that the manuscript has been substantially improved and clarified during the review process and I hope that you found the reviewer comments helpful. I have a few remaining suggestions.

Thank you for the opportunity to revise our manuscript. Below are the changes based on your comments, our comments are in green, text added to the manuscript are blue, and your original comments are in black. Please note that line numbers indicated on this document reflect those of the "track changes" version of the revised manuscript. The authors also made small grammatical corrections and added the Great Lakes Restoration Initiative to the acknowledgement section, noted in the tracked changes document as "Author edits".

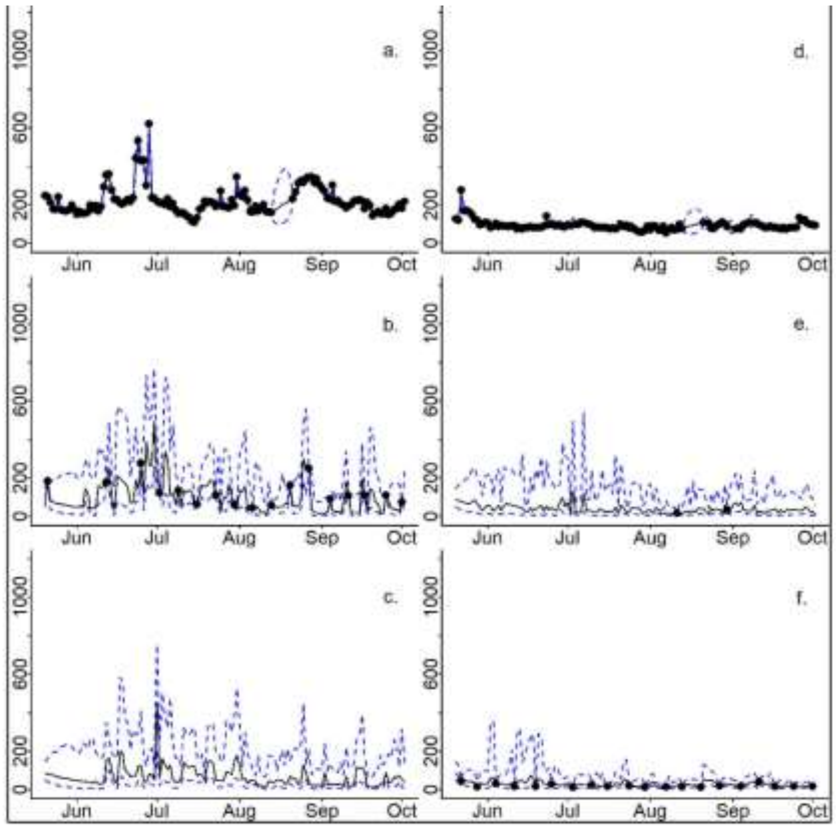
E1C1 R1C7 In your reply to reviewer comment, you state that truncating the values of the latent state does not set a prior based on the data, rather it constrains the later state to plausible values. Does this point relate to your explanation of the high Bayesian R2 values (in response to R1C4), whereby 'the model passes through the observed data and the daily observation in the river datasets are likely driving Bayesian R2 values higher with their constrained predictive intervals'? You also say that the Bayesian R2 may not be a fair metric to evaluate the model performance. Hence, is there a link between the Reviewer's point that such constraints (or priors) should not be informed by observations to which the data is fitted and your explanation of the high Bayesian R2 values? Or perhaps I have misunderstood. However, if there is a link, could this be reflected in the discussion as a potential limitation and area for future improvement?

Our reasoning that the upper truncation on the latent values is not associated with the anomalously high Bayesian R2 values. The upper truncation is based on the condition that no concentration observed in the lake will ever be greater than the largest concentration in the Maumee River.

Lines 148 to 149

The maximum observed Maumee River concentration was used as the upper truncating value because no observation in the lake will exceed this value.

Figure 2(a-f) shows why our Bayesian R2 values are likely not an accurate representation of model fit. River samples (a) and (d) contain the bulk of our data, thus the model passing through those daily values does a great job fitting posteriors with constrained predictive intervals. While (b,c,e,and f) show how difficult it is to fit values out in the lake where data is sparse. The posterior predictive p-values and leave-one-out metrics of model fit specifically address these places where data is sparse making up for the Bayesian R2 values which are not telling a complete story. We have added a reference to Figure 2a & 2d to the text we inserted during the previous revision at Line 277.



Cropped Figure 2(a-f).

E1C2 R1C8 I found the format suggested by the reviewer to explain the model clear and helpful. Can you please take another look at the wording proposed by the reviewer and consider adopting it, or parts of it, in the manuscript? I believe it makes the model structure clear and explicit, especially the explanation of the two models – data=observation and latent=process.

We have amended and added text to the manuscript based on the previous responses.

Lines 124 - 167

The distance between each daily offset surface current location (Lat_1, Lon_1) and each 2 km-by-2 km concentration node was measured and the node *n* with the shortest distance defined the adjacency matrix to associate each node *n* on day *t* with the node *k* on day *t-1*. The state-space model consists of two models, an observation model of data (y) and a latent state (x) process model.

$$y_{n,t,y} \sim N\left(x_{n,t,y}\,, \sigma^2\right) \tag{5}$$

Log-transformed TP concentration observations (y) at the n^{th} node on the t^{th} day of the y^{th} year was estimated with a normal data model sampled from the unobserved latent state variable (x) at the n^{th} node on the t^{th} day of the y^{th} year with standard deviation σ (Eq 5). The process model is a first order Markov, only depending on the value of the node at time *t-1* which transported TP to node n at time *t*, that source node is denoted *k*. For nodes in the river, *k*=*n* and for nodes in the lake, *k*, is determined from the time *t* adjacency matrix.

$$x_{n,t,y} \sim Truncated\ N\left(f\left(x_{n,t-1,y}\right), \tau^2\right) \quad I\left(a \le x_{n,t,y} \le b\right) \tag{6}$$

where

$$f\left(x_{n,t-1,y}\right) = \begin{cases} x_{k,t-1,y} * \beta_{mau} & if\ n = Maumee\ River\ Node \\ x_{k,t-1,y} * \beta_{rai} & if\ n = River\ Raisin\ Node \\ x_{k,t-1,y} * \beta_{self} & if\ n = same\ lake\ Node \\ x_{k,t-1,y} * \beta_{lake} & if\ n = different\ lake\ Node \end{cases} \tag{7}$$

The latent state ($x_{n,t,y}$; Eq 6) is sampled from a normal distribution of a predicted latent state ($f(x_{n,t-1,y})$, Eq 7) and standard deviation τ. $x_{n,t,y}$ was truncated by the detection limit of TP laboratory analysis (5 μg l^{-1}, *a*, Eq 6) and the maximum value observed in each year (y) within the Maumee River (*b*, Eq 6). The maximum observed Maumee River concentration was used as the upper truncating value because no observation in the lake will exceed this value. Priors were uninformative and defined as;

$$\sigma, \tau \overset{\text{iid}}{\sim} Gamma(0.001, 0.001)$$

$$\beta_{self}, \beta_{lake} \overset{\text{iid}}{\sim} Normal(0,10,000)$$

$$\beta` \sim Normal(0,10,000)$$

$$\tau`_{mau}, \tau`_{rai} \overset{\text{iid}}{\sim} Gamma(0.001, 0.001)$$

$$\beta_{mau} \sim Normal(\beta`, \tau`^{2}_{mau})$$

$$\beta_{rai} \sim Normal(\beta`, \tau`^{2}_{rai})$$

River model coefficients (β_{mau} and β_{rai}) were fit hierarchically ($N(\beta`,\tau`^{2})$) because the ecological and anthropogenic processes enacted on these watersheds are similar, if at different scales. The two lake models were fit with two independent β coefficients depending on if the nearest adjacent node *k* is the same as the estimated node *n* (β_{self}) or if a different node *k* is the nearest (β_{lake}). Separate independent in-lake models were used to capture different potential drivers of TP concentration through time depending on whether each node was subject to little surface water movement (β_{self}) or active surface water movement (β_{lake}).

E1C3 R1C9 I think this explanation is of interest and would be helpful to include in the manuscript for a better understanding of the physical processes operating on the lake.

We have amended and added text to the manuscript based on the previous responses.

Lines 303 - 304

The lack of σ and τ correlation was also visually assessed (Figure C1(c)).

Lines 356 - 361

Distance and direction across our model extent is wrapped up in the change observed in predictive intervals through time. In our framework distance and whether the water mass from the Maumee River physically moves toward a node combine. There are several days in which even the nodes closest to the Maumee River are bypassed because the currents take Maumee River water in a different direction. The dual complications of distance and movement have complicated previous attempts at defining a single relationship between Maumee River load and observed in-lake concentrations, which we overcome here.

E1C4 R2C1and R2C2 I believe that your response to these reviewer questions may be of interest to the readers. Can you please consider including these points in the manuscript?

We have amended and added text to the manuscript based on the previous responses.

Lines 283 – 300

TP is a conservative water quality constituent. TP observations are insensitive to biogeochemical transformations of phosphorus form because these data represent both the organic and inorganic forms of phosphorus occurring in the water column. βmau, βras, βlake, and βself fit in our models had 95% predictive intervals encompassing a value of 1. Coefficients were close to 1 because on our daily time-step, the TP concentrations do not widely vary (e.g., the concentration today is similar to the concentration yesterday). The uncertainty in the process and data models allows the model predictions to trend toward the observations where available and be constrained where previous time-steps passed through observations. Were these coefficients to exceed 1 this would be evidence of other inputs of P or less that 1 would indicate some internal loss such as settling or dilution. TP is conservative to processes within the water column because it accounts for the dissolved and particulate P, if our model was applied only to dissolved P which is subject to strong assimilation pressure by phytoplankton the model coefficients would likely be negative. While dilution, settling, and internal loading of TP are happening within our modelled extent in western Lake Erie. Our model lacks the specificity to capture dilution, settling, and internal loading and therefore their effect is being accumulated in our error terms. However, this state-space framework could be defined with a mechanistic process model that did capture these effects. Additionally, while our framework could be implemented with the coefficients (βMau, βRas, βLake, and βSelf) fit hierarchically by year potentially defining the overall effect of dilution, settling, and internal loading, current restrictions on computer memory prevented that use here. However, for smaller spatial and temporal models it could be effective.

Lines 322 – 333

Our model framework allows information from discrete grab samples to be shared across any waterbody where the movement pattern of water is available. Additionally, this model can generate estimates at unobserved nodes or at unobserved time-steps of observed nodes without requiring defined biogeochemical processes of a mechanistic model. For our application, the 2km x 2km grid was chosen to match the surface current dataset. While we did not experiment with other discrete grid distances, any applicable configuration will work. Defining a reasonable grid distance could be based on the spatial distribution of the available data and the user's willingness to extrapolate or average surface current direction and magnitude. Similarly, our temporal time-step was daily, but this could also be applied to monthly data in data sparse systems or hourly data in data rich applications. This spatial and temporal flexibility or using state space frameworks gives users the capacity to tune the computational runtime and resolution of models to fit the hypothesis tested.