E1C1

Thank you for your revised manuscript, which has now been reviewed by two reviewers. The first reviewer still has concerns regarding the mathematical implementation of the model. A particular concern relates to the estimated slope coefficients relating transfers of the logarithm of TP from a source node to a recipient node being near 1, with the resulting implications for transfers on the raw (un-logged) scale increasing by about 2.8. The reviewer also has concerns regarding the calculations of the Bayesian R-sq values and why they are so high and would like to see a clearer explanation of the K-fold cross-validation calculation. The reviewer likes the idea of using a state-space model for this application but has concerns regarding the implementation of the approach. I agree with both reviewers that the application of the SSM approach to the understanding of total phosphorus pollution in a large lake system is novel and worth pursuing. Could you please therefore take careful account of the constructive criticism offered in the referee 1 report and consider revising your manuscript to address the raised concerns?

Dr. Glendell,

Thank you for the opportunity to respond to the peer reviewed comments. We have addressed each in the attached point-by-point response document and noted changes in the attached "tracked changes" manuscript copy with each reviewer comment (R1C1; reviewer 1 comment 1) noted.

Reviewer 1's particular concern stemming from the coefficients near 1 resulting in a doubling of values is not taking place in the model. We show unequivocally that the values do not double, moreover the code and data we provide at https://doi.org/10.5281/zenodo.5570508 can be run by anyone and will show no such doubling. The Bayesian R-sq values are so high because we use a daily time-step and the daily data available in the two river datasets create inflated R-sq values as the model passes through these observations with a very restricted predictive interval (illustrated in Figure 2a and 2d). This underscores the importance of using a leave-one-out K-fold cross validation. The nomenclature of the K-fold cross validation has been updated and the reference updated to provide the page number. Our explanation of the K-fold cross-validation calculation shows how we are removing a node with observations, predicting those observations, and ensuring that predictions are preforming consistently for each year and irrespective of which node was removed. Additionally, a discussion of Reviewer 2's concerns with dilution and settling were updated in the text.

As authors, we are pleased to submit this manuscript and feel the extensive feedback from the previous rounds of review have made a complete story we are eager to see finished. Please contact me with any subsequent questions. Thank you for your consideration.

Reviewer #1 (For easy of review, the comments from Reviewer #1 are pasted as images rather than coping the text off the submitted PDF)

Reviewer Comments – Black

Author Response – Green

Altered text – Blue

Page numbers referred to the "Tracked Changes" version of manuscript.

R1C1

The manuscript remains generally well organized and written. The authors have responded fairly thoroughly to the comments made in the first review, and I appreciate them checking out what happens with hourly step sizes to construct the adjacency matrix. Several substantive concerns remain, however.

That the posterior distributions for all four β's in the process model are so concentrated near 1 (across all years) seems remarkable. That coincidence aside, assuming that the latent states are the logarithms of the true TP value, the effects on the expected TP transfer from a "source" node (denoted k) to a corresponding "sink" or recipient node (denoted n) implies an increase in TP at the sink node on the raw scale. Letting $z_{n,t,y}$ be the raw scale TP value, thus $x_{n,t,y,} = \ln(z_{n,t,y})$, then the expected TP value in the sink node

$$E[z_{n,t,y}|z_{k,t,y}] = \exp\left(\beta + \ln(z_{k,t,y}) + \tau^2/2\right)$$

Substituting β=1 and τ^2=0.2^2 (a rough average from Table C1):

$$E[z_{n,t,y}|z_{k,t,y}] = \exp\left(1 + \ln(z_{k,t,y}) + 0.2^2/2\right) = 2.77 z_{k,t,y}$$

Values are logged prior to model fitting and remain in log space until the samples of the fitted posteriors are exponentiated.

E.g., a value of 100 ug/l total phosphorus

$y_{t=1}$ = log(100) = 4.60517

Sampling State Space Model: 4.60517 * **1** + data/process error $\approx$ 4.62

$\hat{y}_{t=2}$ = exp(4.62) = 101.5 ug/l total phosphorus

There is no doubling of concentrations between nodes. Please refer to R1C8.

R1C2

Another concern is with regard to the posteriors for the process and obs'n standard deviations. In Figure C1, the ranges of the joint priors for the pairs $(\beta_{rai}, \beta_{mau})$, $(\beta_{self}, \beta_{lake})$, and $(\ln(\sigma_y), \ln(\tau_x))$ (my added subscripts) are denoted by the red convex hulls. The black polygons denote posterior fitted values (for all 11 years).

- Are fitted values posterior means?
- I don't understand the sentence in the figure caption about the fitted values for each year not overlapping. Based on Table C1 there is considerable similarity in the β's: they are all very close to 1.
- With Figure C1.c it is disconcerting to show posterior means outside the support of the prior; that should not be so.
- Based on the `R` code the priors for the precision for the process and obs'n models ($1/\tau^2$ and $1/\sigma^2$) are Gamma(0.001, 0.001). The ranges for τ and σ are the positive real numbers (which in practice, see below, can be 0 and ∞), and the subsequent range for $\ln(\tau)$ and $\ln(\sigma)$ is much larger than Figure C1.c indicates (roughly -6 to 11).

- Fitted values are not represented as means, these black polygons are the extent of all sampled posteriors for all years.
- This sentence has been revised

Lines 446 to 447

The fitted values for every year (a. and b., all values represented as black polygon) do not appear to be overly influenced by the uninformed priors.

- The posterior values of Figure C1(c) are not disconcerting as the priors are purposefully broad and uninformed.
- The values in the figures are where generated via JAGS sampling of the defined uninformed priors, they do not represent the breath of theoretical values.

R1C3

Potential identifiability issues for τ and σ have not been addressed. What needs to be examined is the correlation between τ and σ. Scatterplots of sampled pairs from the posterior distribution, at a minimum, need to be examined (and shown). Figure 1 shows how the lower and upper bounds of the PIs relate, suggesting a negative association.

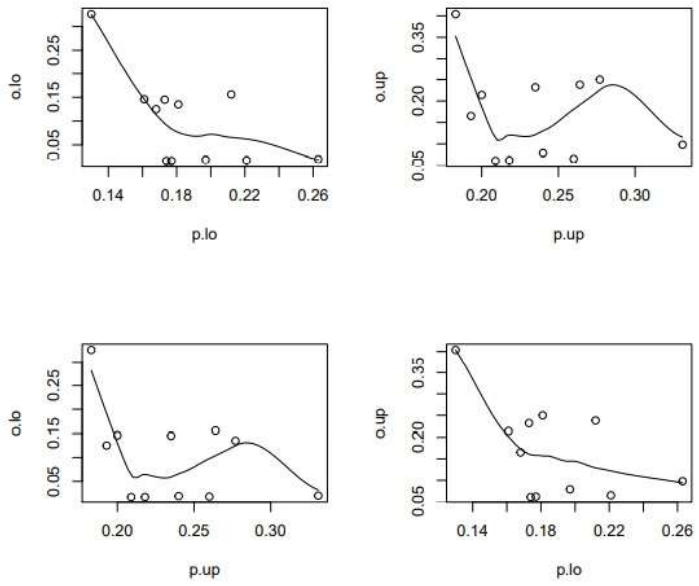

Figure 1: Scatterplots of lower and upper bounds of 95% PIs for τ and σ. p.lo and p.up are lower and upper endpoints for the process standard deviation; similarly, o.lo and o.up for observation.

The lack of σ and τ correlation is adequately displayed in Figure C1(c), which are the scatter plots that the reviewer specifies.

R1C4

My understanding of the Bayesian R^2 calculations based on Gelman et al (2019, eq'n 3) is that the calculated value of var_{res} would be the estimated value of the observation variance terms, $(\sigma^2)^s$ for sample s from the posterior. What is written in the text seems to match the Gelman et al eq'n 2 definition for *non*-Bayesian R^2. How much difference that makes needs to be examined—the fact that the majority of the average Bayesian R^2 are greater than 0.98 or so seems remarkable.

The estimates of Bayesian R2 have been recalculated via Gelman et al (2019) Eq 3, and values updated in Table 1. This approach only made our Bayesian R2 values closer to 1. The reason this model metric looks so "good" is discussed in the additional text below.

Lines 172 to 174

Bayesian R2 defined as the fitted variance (varfit) divided by the sum of varfit and the residual variance (varres) was calculated for each model year. Model varfit was the variance of the modelled predictive mean, while varres is estimated by squared standard deviation of the errors (Gelman et al., 2019).

Lines 263 to 274

The models consistently generated plausible posterior samples for mean TP concentration as each 95% CI of annual posterior predictive p-values included 0.5 and annual Bayesian R2 95% CI values ranged from 0.84 to 0.99 (Table 1). Annual posterior predictive p-values indicate that our model framework is preforming well predicting water quality within large water bodies even with sparse observations within the data. While our high Bayesian R2 values appear to support the use of our model, it is likely that they represent an inappropriate model metric. The state-space framework forces the model to pass through the observed data and the daily observation in the river datasets are likely driving Bayesian R2 values higher with their constrained predictive intervals. Because of these elevated Bayesian R2 estimates, the k-fold CV are important checks on the applicability of this state-space approach. The k-fold CV results generated by removing all the observations of a randomly selected lake node with at least 10 observations showed that model predictions were equally accurate across years and by node. Predicting equally well across the nodes and within any year provides strong support for this framework as being a useful application of Bayesian methods in water quality modelling.

R1C5

The description of the K-Fold cross-validation score in eq'n 8 is unclear. The summation index is d but the last value of the index is also d, the term d does not appear in the values being summed, and the conditioning notation $\bar{y}_{n,t,y}$ is is ambiguous. I am assuming that the eq'n is patterned after equation 3 in Piironen and Vehtari (2017) and conditioning would be on the set excluding the $y_{n,t,y}$ values; something like they used $D_{-y_{n,t,y}}$ might be more understandable.

The cross-validation scores are then regressed on $\mu_{o,n}$, μ_n's (eq'n 9), or $\mu_{o,y}$ and μ_y (eq'n 10), where these regressors are then given priors. The terms on the left-hand sides of eq'ns 9 and 10 need subscripting. More critically, are the regressors pure random effects and what are called priors actually the probability distributions for these random effects? I tried to find where in Kruschke (2014; that should be 2015 I believe) such a procedure was used—chapter or page numbers need to indicated.

The subscripts in Eqn 8 have been updated to reflect that the log predictive density from posterior samples come from the *k*th node which were observations were removed and subsequently fit. This is important because was we discuss in R1C4 the other measures of model performance do not give a complete accounting of how the model does within the lake where data is scarce. "d" represents the omitted observations. As per the sentences preceding Eqn 8. We are choosing to keep our $y_{k,t,y}$ nomenclature to remain consistent with the previously described data and process model.

The K-Fold CV's have been given subscripts defining the cross validation done by node "*n*" or year "*y*". The variables fit in Eq 9 and 10 are the between-group and among-group variance as described in "Doing Bayesian Data Analysis" page 560, the reference has been updated to reflect the correct year and page number.

R1C6

p3, L93: the 2.2 heading "Model description" still seems misleading, as the real model is the SSM. Subsection 2.2 is about creating the adjacency matrix, which is *not* a model, per se.

We do not feel the current headings are misleading.

R1C7

Constraints on State process values: Need to state that $x_{n,t,y}$ is the logarithm of TP: this is also important as the authors' reply about TP not being negative applies to the raw scale values, not the logarithm. While I can see the argument for constraining the range of $x_{n,t,y}$ based on expert opinion, I don't think that the observations in a given year should be used to determine those constraints: the prior needs to be independent of the data, nor does it make sense to me that measuring instrument limits constrain the true value. Why not simply say that our prior opinion is that log of TP lies between a and b and use the same values for each year?

I am possibly misinterpreting the scale for the obs'ns but based on the `R` code: `concs = log(concs*10^3)`, summaries of raw and logged TP values for the Maumee River, Raisin River, and the lake are the following:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
Maumee R	77.33	173.38	218.07	261.20	334.25	804.80	21.00
Raisin R	28.00	65.95	85.15	107.12	109.35	458.20	65.00
Lake	5.50	20.60	33.20	51.05	63.81	275.07	91789.00
log Maumee R	4.35	5.16	5.38	5.44	5.81	6.69	21.00
log Raisin R	3.33	4.19	4.44	4.51	4.69	6.13	65.00
log Lake	1.70	3.03	3.50	3.62	4.16	5.62	91789.00

The authors' reply that 0.7% were less than 10 ug/l is presumably referring to the lake samples; I found one out of 191 of the lake values less than 10. Note that the above means differ from those reported in the Results (lines 214-215), but maybe I've made mistakes.

The values within the Lake and within the Maumee River are too highly variable for a constant constraint between two values to be useful. Truncating the values of the latent state does not set a prior based on the data, rather it constrains the latent state within each year to values that represent reality based on the source dataset. The effectiveness of our approach is evident in Figure 4 where the culmination of our modeling scheme generates the a linear expression of Maumee River TP spring load to western Lake Erie TP concentration.

The values the reviewer generated the above table were only from 2018 which we provided in the online repository as an example. The values in the manuscript are representative of all our available data.

R1C8

p4, L122-135: I find the explanation of the SSM awkward. Need to state somewhere that the latent states are log transformed (only by checking the R code did that become clear). Consider: "The SSM consists of two models, one for the data (y) called the observation model and one for latent states (x) called the process model." The observations were modelled as follows.

$$y_{n,t,y} \sim \text{Normal}\left(x_{n,t,y}, \sigma^2\right) \tag{1}$$

where $y_{n,t,y}$ is the natural logarithm of the measured TP concentration at node n on day t of year y, $x_{n,t,y}$ is latent (unobserved) true log TP concentration, and σ^2 is the observation standard deviation. The process model is first order Markov, only depending on the value of the node at time $t-1$ which transported to TP to node n at time t. That source node is denoted k and for nodes in the river, k=n, and for nodes in the lake, k is determined from the time t adjacency matrix.

$$x_{n,t,y} \sim \text{Truncated Normal}\left(f(x_{k,t-1,y}), \tau^2\right) \quad I(a \leq x_{n,t,y} \leq b), \tag{2}$$

where

$$f(x_{k,t-1,y}) = \begin{cases} \beta_{mau} * x_{k,t-1,y} & \text{if } n\text{=Maumee River node} \\ \beta_{rai} * x_{k,t-1,y} & \text{if } n\text{=River Raisin node} \\ \beta_{self} * x_{k,t-1,y} & \text{if } n\text{=same lake node} \\ \beta_{lake} * x_{k,t-1,y} & \text{if } n\text{=a different lake node} \end{cases} \tag{3}$$

The process standard deviation is τ and the values for $x_{n,t,y}$ are restricted to $[a, b]$

Also please write the eq'ns for the priors for the β's to ease comprehension for the reader. It's not clear to me what the sentence at lines 136-137 is saying. Only by examining the R code could I tell that:

$$\begin{aligned} \beta_{self}, \beta_{lake} &\overset{iid}{\sim} \text{Normal}\left(0, 10,000\right) \\ \beta^* &\sim \text{Normal}\left(0, 10,000\right) \\ \tau_{mau}, \tau_{rai} &\overset{iid}{\sim} \text{Gamma}\left(0.001, 0.001\right) \\ \beta_{mau} &\sim \text{Normal}\left(\beta^*, \tau_{mau}^{-2}\right) \\ \beta_{rai} &\sim \text{Normal}\left(\beta^*, \tau_{rai}^{-2}\right) \end{aligned}$$

Also write the priors for $1/\tau^2$ and $1/\sigma^2$.

The text immediately after the model definitions in our revised text noted the logged concentrations. To again emphasize this, we have added text to describe how the data were logged prior to fitting the model and that output MCMC samples were not exponentiated until plotting. Thank you for the suggested text however we intend to keep the current revised version, The priors for σ and τ were uniformed gamma distributions which are now in the text.

Lines 135 to 136

The data were logged prior to fitting the model and MCMC samples remained in log space until exponentiated for plotting.

Lines 147 to 150

The latent state ($x_{n,t,y}$; Eq 6) is sampled from a normal distribution of a predicted latent state ($xp_{n,t,y}$, 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), $xp_{n,t,y}$ was defined depending on the node n as being a river or lake node (Eq 7), and σ and τ were fit with uniformed *gamma*(0.001,0.001) priors.

R1C9

Results. Major comments about the results were given in Section 1. One other point is the regression of mean TP concentration on effective Spring TP load: doesn't distance of the lake nodes from the Maumee River have an effect on the relationship? Distance is not included in the regression.

Distance is wrapped up in the deflection of the predictive intervals through time because it isn't just the distance but also whether the water mass from the Maumee physically moves toward a node. There are several days in which even the closer nodes 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 load and observed in-lake concentrations, which we overcome here.

R1C10

Discussion. Not clear what in Auger-Methe, et al (2021) is being referred to regarding identifiability and visual determination of priors dominating. Is what is meant that the posterior and prior will not differ much for parameters that are unidentiable or weakly identifiable?

Yes, as we discussed in our prior revisions, identifiability as defined by the priors dominating the parameter fit was not observed.

R1C11

p1, L22: "estimated that, in the absence of the Maumee River load, lake concentrations..."

Change made

R1C12

p2, L46: “affects”

Change made

R1C13

p2, L52: here refer to soluble reactive phosphorous (SRP), but later (starting with p3, L80) refer to TP, without defining what TP means. I’m no expert on water chemistry, but my understanding is that TP, total phosphorous, includes SRP. TP needs to be defined, and if SRP is not referred to again, perhaps do not add the abbreviation.

We have made these changes.

Lines 53 to 54

Spring Maumee River soluble reactive phosphorus export correlates with western Lake Erie HABs extent; this pattern has been observed since the soluble reactive phosphorus loads started to increase in the 1990s (Ho and Michalak, 2017; Michalak et al., 2013; Stow et al., 2015).

Lines 82 to 84

Here, we quantified how well our model fits the data and generated predictions of total phosphorus (TP) concentrations across western Lake Erie. TP includes the dissolved and particulate forms of phosphorus.

R1C14

p2, L60: perhaps “Bayesian inference” instead of “Bayesian frameworks”. This paragraph is more about SSMs than about Bayesian inference, and it might be better to make a statement about SSMs first (the topic sentence), e.g., use the 2nd sentence without the adjective Bayesian: “State-space models (SSMs) have been used...”. Then discuss the application areas and then add a sentence or two about Bayesian SSMs.

We have made these changes.

Lines 63 to 70

Bayesian state-space models have been used in ecology to incorporate temporal and spatial autocorrelation and quantify observation error separate from the error attributable to the modelled ecological process (Auger-Méthé et al., 2021; Durbin and Koopman, 2012; Shumway and Stoffer, 2019). State-space models are widely used in ecology to model animal populations (Buckland et al., 2004), movement (Royer et al., 2005), and fisheries stocks (Meyer and Millar, 1999). Bayesian inference can quantify uncertainty in the effect of nutrient load on nutrient distribution within a dynamic system such as Lake Erie. Non-stationary time-series models have been used in the Great Lakes to model water levels (Lamon and Stow, 2010; Sellinger et al., 2008) and to predict polychlorinated biphenyls concentration in trout (Stow et al., 2004).

R1C15

p3, L68: "While spatial models"

Please refer to R2C4.

R1C16

p3, L71: "incorporate concentration data"

Change made

R1C17

p3, L80: define TP

Please refer to R1C13.

R1C18

p4, L100-110. Perhaps: "Hourly northward and eastward transport ... was expressed in radians:

$$dLat_t = \frac{dN_t}{R}$$
$$dLon_t = \frac{dE_t}{R\cos\left(\pi\frac{Lat_t}{180}\right)}$$

Then add "The latitude and longitude at time $t+1$, given the latitude and longitude at time t and the above derivatives, was calculated as follows:

$$Lat_{t+1} = Lat_t + dLat_t * \left(\frac{180}{\pi}\right)$$
$$Lon_{t+1} = Lon_t + dLon_t * \left(\frac{180}{\pi}\right)$$

I don't think Eqs 1-4 ever get referred to and they do not need to be numbered. I found the use of 0 and 1 confusing as this procedure is carried out at every time step.

We have changed Eq 3 and 4 to use the "t" notation, "t" is already defined at the beginning of that paragraph, so we did not need to add the sentence suggested. Eq 1 and 2 are referred to in the preceding paragraph, so their numbered designations remain.

R1C19

p5, L145. The material beginning "The model was run" belongs in a section labelled Fitting the SSM, not in the SSM model description section. It might be more appropriate to name Section 2.2.2 SSM fitting and diagnostics.

Our opinion that the sentence should remain in Section 2.2.1. because it precedes the description of iterations, thin, and chain numbers, all consistent with what reader would expect following the model parameterization description at the beginning of this paragraph. We have added "Diagnostics" to the Section 2.2.2. heading.

Reviewer #2

R2C1 This manuscript presents an interesting application of state-space models for predicting the distribution of TP across a large body of water. The presentation is clear, and the resulting model provides a useful contribution to the literature. I have only minor comments. Spatial resolution: The selection of the 2 km x 2 km grid seems to derive from the spatial scale of currents. Is this a correct assumption? Have other spatial resolutions been considered for discretizing the distribution of TP concentrations? For example, a coarser resolution might be necessary to feasibly model a larger body of water, whereas a finer resolution might enable one to perceive smaller scale features. Or, is the modeling resolution strictly determined by the spatial scale of the current database?

For our application, yes 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 a 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 larger data sparse systems or hourly data in smaller 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.

R2C2 Dilution from Maumee to Erie: I would have guessed that there would be some dilution of Maumee River flow as it enters Lake Erie, with an associated decrease in TP concentration. However, the coefficient Beta_mau is 1, indicating no change in concentration. Is this because the Maumee River flow accounts for the majority of the near shore water volume?

The Maumee coefficient (like the Raisin and two lake coefficients) was close to 1 because on a 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. 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 be negative, this could be very useful to others, so we maintain the coefficient use here.

R2C3 Line 31: “Excessive nutrient export primarily from agricultural watersheds…” The phrasing of this sentence seems to imply that eutrophication can mostly be attributed to agricultural nutrient loads, which may be true for certain receiving waters, but is not true for all water bodies.

We have altered that sentence to frame it as an example of the point and non-point sources introduced in the previous sentence.

Lines 30 to 31

"Observed concentrations are driven by both point and non-point sources. E.g., Wastewater effluent and excessive nutrient export primarily from agricultural watersheds may lead to eutrophication, harmful algae blooms (HABs), and threatens drinking water contamination (Brooks et al., 2016; Mellios et al., 2020; Schneider and Bláha, 2020)."

R2C4 Line 68: "Spatial" does not need to be capitalized.

"Spatial" in this sentence does not need to be capitalized.

Lines 74 to 75

"While spatial models have been used in the Great Lakes for predicting HABs biomass, HABs extent, and nutrient transport (Fang et al., 2019; Schwab et al., 2009), we proposed a Bayesian framework for similar spatial data."

R2C5 Line 77: Might want to break this section into at least two sentences.

We have broken this sentence into three to aid the reader interpreting what model we built and how we used the model to test a hypothesis.

Lines 82 to 87

"Together our contribution will fit values in the absence of observations and allow experimentation in archival data previously not possible. Here, we quantified how well our model fits the data and generated predictions of TP concentrations across western Lake Erie. Additionally, we experimentally manipulated observed concentrations to estimate the spatial and temporal impact from the Maumee River plume. Using this delineated Maumee River impact in time and space we tested the hypothesis that when water movement is incorporated, there is a linear relationship between river load and western Lake Erie water TP concentrations."

R2C6/C7 Line 254: TP is conservative in the context of this model only if we neglect settling, right? Line 256: I'm assuming that the dilution effect that is mentioned here is due to the increasing depth? What would be magnitude of dilution one would expect, given the known bathymetry? So, given that settling and dilution are known processes that would reduce TP concentration, why does the model still yield beta values that are 1?

Our model of TP is conservative to settling and dilution specifically in our extant within Western Lake Erie. While both settling and dilution are happening within the lake our model lacks the requisite detail to represent it and these effects are therefore subsumed by the data and process error leaving the coefficients close to 1. A more mechanistic process model which defined the effect of dilution and settling is an appropriate next step to our model and we hope that this manuscript prompts such models. We will adjust the text to make this more apparent

Lines 276 - 286

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. β_{mau}, β_{rai}, β_{self}, and β_{lake} near 1 would then be expected in the absence of dilution and settling. βs larger than 1 would indicate in-lake sources of TP. Every β_{mau}, β_{ras}, β_{lake}, and β_{self} fit in our models had 95% predictive intervals encompassing a value of 1. However, dilution, settling, and internal loading of TP are happening within our modeled 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 the 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.