Authors' response to Reviewer 2, Anonymous
For clarity, we have included the reviewer's comments in black; our response is in blue

General Comments:
The manuscript "Data-based mechanistic model of catchment phosphorus load improves predictions of storm transfers and annual loads in surface waters" addresses relevant scientific questions and should be published after some major revisions. The highlight of the value of high temporal resolution measurements to capture the system dynamics is very valid, as well as the benefit of the chosen approach in determining the "optimal" temporal resolution for future measurement campaigns. However, especially the "methods" section needs to be expanded by clearly stating and justifying the assumptions made.
See specific responses below
Specific comments:
Chapter 2.3:
What is exactly the meaning and the implications of not using a noise model in Eq 2? This should be explained in more detail. Any inference algorithm (in this case probably the RIV(C)BJ), needs to make assumptions about the errors to estimate parameters. Does not using a noise model mean that you assume the errors to be uncorrelated? Or is the error model inferred by the algorithm itself? The assumptions made in the inference process should be clearly stated and checked.
Noise models were not used for two reasons. (a) full models (input-output (I-O) plus noise) were actually initially evaluated and overall they did not produce better results; therefore, in order to keep a consistent approach for all catchments, noise models were not used in later model identification, (b) transfer function (TF) models with a noise component generally do not improve longer term predictions of processes which are I-O dominated, the noise being modelled as ARMA processes and thus not generating good longer term forecasts. The Reviewer is correct in that the residuals structure was not strong/consistent enough for a noise model to improve the model fit consistently. This will be mentioned in the manuscript.
Eq 4 leads to significant violation of the mass balance w.r.t. water if $Q(t-1)$ is larger than 1 (this depends strongly on the units of Q) and beta is larger than 0. This should be clearly stated, and then briefly mentioned why this is not a problem in this case (if it is not).
Sometimes rescaling is applied to Re, to ensure that total Re is equal to total R, and then the TF gain parameters can be interpreted as runoff coefficients. However, if rescaling is not done at that stage, it is balanced by the linear TF parameters (i.e. the rescaling takes place within the transfer function estimation process). In this application, rescaling at the Re stage would not be possible in the 'double layer' TP model, where Re and Q are simulated together, one step at a time.
Chapter 3:
It would be interesting to validate some models with noise to get the total predictive uncertainty, not just parameter uncertainty. This would be a more meaningful validation and would allow actual statements about the uncertainty related to predictions of TP loads.
Model residual variance is included in the parametric uncertainty. We propose to show 'double band' plots of model fit, where the data uncertainty will be illustrated as one band on the observations and model parametric uncertainty will be shown, as at present, on the model simulation, thus effectively comparing the distributions at each sample.
Most plots show observed and modeled quantities in the identification period. It would be interesting to have some more plots in the validation period (ev. including the total uncertainty bands including the noise). Also some zooms to specific time periods showing the strengths and the weaknesses of the models would be interesting.

We propose to show further plots, including some of validation period, and certainly some zoomed in periods. Thanks for this comment.

Page 11 L33: The comparison of the quality of fit to process-based models needs a more detailed analysis. How exactly do you compare the quality of fit? With NS coefficients only? Is this justified, are the assumptions underlying NS coefficients fulfilled?

Direct full comparisons are not currently possible, as the published results for process-based models used different catchments and data sets. Thus only generic model fit comparisons are possible at this time. However, we propose to tone down this comparison with other models, and accordingly we propose a change of title to "Prediction of storm transfers and annual loads with data-based mechanistic models using high-frequency data".

Technical corrections:

Page 8 L4: should be Table 2

This is true; thank you for noticing, this will be corrected

Page 10 L27/28: one of the "fast pathway" should probably mean "slow pathway"?

No, the text is correct. This paragraph compares the Blackwater TPload model with its corresponding runoff model, so the first 'fast pathway' refers to the proportion of TPload transferred on this pathway (p10, l27) and the second 'fast pathway' refers to the proportion of water transferred (p10, l28)

Page 11 L22: "in" should be "an"

Thank you for noticing, this will be corrected