Authors' response to Reviewer 3, Anonymous
For clarity, we have included the reviewer's comments in black; our response is in blue

General comments

This paper explores the use of Data-Based Mechanistic modelling to gain understanding of the dynamics of phosphorus transport using a high resolution TP concentration and load time series data set. To my knowledge, such a study has never been conducted using an hourly time series data set. As such, the paper will be useful to anyone working in phosphorus transport – a key component to understanding eutrophication issues. It is particularly worth noting that the authors are making the data openly available for which they are to be applauded. Presumably this will include the necessary meta-data to enable future users to understand the uncertainty in the measurements. Overall, the paper is very well written, and suitable for publication after the issues noted below have been dealt with.

We are very grateful for this recognition of our contribution

Specific comments

Given that approximately 3 years of data have been used in this study, could the authors comment on the magnitude of the error/uncertainty in the result if only daily data were used, or even worse, data collected at varying intervals (as is frequently the case)?

The errors resulting from sampling well below the catchment dynamics have been well documented elsewhere, e.g. Lloyd et al., 2016; Johnes, 2007; Jones et al., 2012, Moatar et al., 2013. However, we propose to include these references.

Lloyd, C. E. M., Freer, J. E., Johnes, P. J., Coxon, G., and Collins, A. L.: Discharge and nutrient uncertainty: implications for nutrient flux estimation in small streams, Hydrol. Process., 30, 135-152, 10.1002/hyp.10574, 2016.

Johnes, P. J.: Uncertainties in annual riverine phosphorus load estimation: Impact of load estimation methodology, sampling frequency, baseflow index and catchment population density, J. Hydrol., 332, 241-258, 10.1016/j.jhydrol.2006.07.006, 2007.

Jones, A. S., Horsburgh, J. S., Mesner, N. O., Ryel, R. J., and Stevens, D. K.: Influence of Sampling Frequency on Estimation of Annual Total Phosphorus and Total Suspended Solids Loads, 48, 1258-1275, 10.1111/j.1752-1688.2012.00684.x, 2012.

Moatar, F., Meybeck, M., Raymond, S., Birgand, F., and Curie, F.: River flux uncertainties predicted by hydrological variability and riverine material behaviour, 27, 3535-3546, 10.1002/hyp.9464, 2013.

The authors cite Young (2010) who argued that the sampling period should be at most 1/6th of the quickest time constant. Given this and finding that the time constant for the fast pathway varied between 2 and 15 hours, using hourly data at the low end of this scale would mean the sampling frequency would be too slow – questioning the statement in the abstract that hourly data are "necessary". For time constants near 2 hours, hourly data would be inadequate by the condition given by Young (2010) – does this mean that the authors are revising the recommendation given by Young (2010)? My thought is this is not the case, so the phrasing in the abstract seems a little strange.

The reviewer is correct in that we are not revising Young's recommendation. Wherever possible we used continuous time models, which are less critical in this respect. However, we propose to edit the statement in the abstract to tone down the requirement for hourly data and state "high temporal resolution data are necessary to capture the dynamic responses in small catchments".

Page 2, line 25: The authors state that USLE is a process-based model. In a broad sense it could be called this, but I feel it would be more accurate to describe it as an empirically-based model.

Accepted. We will change this to semi-empirical

Page 4, line 23: Can visual inspection really determine whether there has been a significant loss of information? Was the inspection merely looking at the hydrograph (i.e. plotting the data)? Could some objective measure be developed for this?
There is no objective measure for this. Close visual inspection/comparison is very valuable in identifying whether the dynamics are still captured by the lower resolution data

Page 6, line 26: A minor point, but there are some issues with the description of NSE. The numerator is the variance of the model residuals if and only if the mean residual (i.e. bias) is identically zero.
Accepted. The transfer function (TF) estimates using Instrumental Variable methods produce unbiased parameter estimates; the bias of residuals is only asymptotically zero. NS measure with all its shortcomings is still a commonly used model metric. However, we propose to replace the word 'variance' with 'variance estimate' in the descriptions of NSE.

It is a reasonable approximation to the variance if the bias is small, however it is a biased estimate as it will always be a little larger than the actual variance. On page 8 (line 22), the authors state that the model bias was less than ± 10% for all three catchments. What would the impact of this amount of model bias be on the comparison between R2t and NSE?
The difference between R_t^2 and NSE is that NSE can be equivalent to either R_t^2 or R^2 depending on the application.
An even more minor point is that the normalisation factor for the variance is usually 1/(N-1) rather than (1/N).
This is indeed a minor issue given the number of samples (thousands). However, we propose to replace the word 'variance' with 'variance estimate' in the descriptions of NSE and Rt2 to address this query.

Page 7, line 25: Text seems to jump a little at this point. Maybe add a sentence bringing "storage" into the picture before the closing sentence for this paragraph?
Accepted. We propose to add (before the last sentence of paragraph) "This indicates that antecedent conditions and the storage state of the catchment are important in determining the response."
Page 8, line 11: Presumably, the rating curve would need to be extrapolated to get to this flow level, which would also contribute to the uncertainty in the estimated discharge?
Exactly. A review of measurement data uncertainty is presented by McMillan et al., 2012 (and references therein) and rigorous treatment of the uncertainties in high frequency nutrient data and its subsequent impact on loads is given by Lloyd et al., 2016. We propose to show 'double banded' plots with one band on the observations to show measurement uncertainty on the discharge and phosphorus load, and one band on the model simulation to show model parametric uncertainty.
McMillan, H., Krueger, T., and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, Hydrol. Process., 26, 4078-4111, 10.1002/hyp.9384, 2012.
Lloyd, C. E. M., Freer, J. E., Johnes, P. J., Coxon, G., and Collins, A. L.: Discharge and nutrient uncertainty: implications for nutrient flux estimation in small streams, Hydrol. Process., 30, 135-152, 10.1002/hyp.10574, 2016.

Technical corrections
Page 7, line 14: space missing after Avon
Thanks for noticing these. Space will be added.
Page 9, line 23+24: The flow fractions for the slow flow component are not really

needed as this are just 100

This is true, but we prefer to include them so that reader does not have to search back several lines to find the % for the fast components.

Page 9, line 32: Space missing after “drain flow”

Space will be added.

Page 12, line 7: space missing after “have”

Space will be added before ‘have’.

Page 12, line 31: space missing before “Yang”

Space will be added.