Referees comments are in italics, reply from author is in plain text

Outram et al. describe research on concentration responses of stream nitrogen and phosphorus species to a drought-ending, extreme rainfall event that spanned three agricultural catchments of the UK Demonstration Test Catchments program. The authors have investigated high-frequency data from sensors using flow/solute duration curves and hysteresis analysis. They have discussed flushing patterns that control the variation of nutrient responses and why patterns vary among nitrate, ammonium, TP, and TRP. A strength of the manuscript is the explicit linkage of monitoring data, environmental policy, and management directives/strategies – the authors clearly and nicely set forth and discuss this framework. Beyond that, the manuscript is grammatically well-written. The topic area is appropriate for HESS. While I feel that the overall message is well crafted and the topic has merit, I am highly critical of several aspects of the manuscript and feel that major revisions are necessary before it may be further considered for publication. Substantial revisions are needed to address: 1) how sensor data were validated, operated, and QA/QC'ed; 2) content on hydrological flowpaths that is not directly supported by any strong evidence or cited, relevant literature; and 3) manuscript organization.

The authors thank the reviewer for this feedback, we are pleased you feel the topic has merit and is relevant for publication in HESS. We will reply to your individual concerns below.

1) Validation and QA/QC information for sensor data: No information is presented to verify how sensors for streamflow or solute concentrations were calibrated, operated, validated, or assessed. In these types of studies, stream stage or flow is typically measured (e.g. readings from staff gages or measurements with flow meters) to validate calculated values of stream discharge from stage height monitoring; and grab water samples must be analyzed using benchtop instruments and standard methods to calibrate and validate sensor data. Furthermore, these validation measurements are typically done with regularity to span flow conditions. While the authors mention stream flow measurements with Doppler flow meters at two (of three) of the sites, there is no mention of how this information was used or how frequent measurements were made. And, I am especially concerned about the lack of descriptions of validation and maintenance methods for nutrient sensors. Without validation for each catchment and descriptions of QA/QC procedures, the data are unpublishable.

We have now included a section on QA/QC procedures which we agree are essential for validating the high resolution data (section 2.3). We have included a comparison table (Table 2) between data collected using bankside analysers and grab samples analysed using standard laboratory procedures which show the measurements are in good agreement.

2) Discussions of hydrological flowpaths have not been supported by data: My contention here is that the authors have over interpreted hysteresis loops and that any resulting attribution of stream concentration variation to inputs from particular flowpaths (e.g. overland, shallow, near-surface, deep) is both completely unsubstantiated and unnecessary to the focus of the manuscript. It has been shown that identical hysteresis patterns may arise from various mixing processes (Chanat, J. G., K. C. Rice, and G. M. Hornberger (2002), Consistency of patterns in concentration-discharge plots, Water Resour. Res., 38(8), 1147, doi:10.1029/2001WR000971). Consequently, mixing from distinct

hydrological flowpaths cannot be deciphered from hysteresis patterns unless there is additional supporting information such as concentrations measured along various flowpaths and evidence that shows when water may have been flowing along those flowpaths. The authors have presented no such supporting information. In short, the analysis and interpretation regarding flowpaths are not credible. Since most of this interpretation appears in the results (see my criticism of this organization in the next comment), exclusion of this topic would have little effect on the strengths of the discussion section as it had been written.

Hysteresis concentration-discharge plots have been used many times in the literature to infer flow pathways, including several highly regarded papers in this journal. We have now included a section in the introduction (page 3 line 23 – page 4 line 9) reviewing some of those papers whilst also making reference to the Chanat et al. 2002 paper suggested by the reviewer to make it clear that hysteresis interpretation is not powerful enough to distinguish absolute pathways for pollutants during storm events. We have also toned down some of the language to make it clear that we are inferring pathways as well as including information on concentrations of end members where applicable, such as groundwater and tile drain concentrations, to back up our interpretation of hysteresis plots (see page 9 lines 8-9; page 13 lines 29 – page 14 line1; page 17 lines 5-6). The hysteresis loops produced as a result of the storm event studied in the three catchments were so different that the authors believe it is possible to infer the controlling pathways which can then be investigated further with tracers and mixing models. All three catchments are engaging in this type of analysis but this was deemed beyond the scope of this paper the main purpose was to highlight the benefits of bankside monitoring which provides a wealth of data for understanding catchment processes.

3) Manuscript organization: Many sentences or paragraphs in the manuscript need to be moved to proper locations. All methods need to be consolidated in the methods section. For example, the calculation of the hysteresis index is a method, not a result. Interpretations (if supported by data) that currently appear in the results need to be moved to the discussion section. See line-by-line comments below for many examples.

Thank you for this comment, the methods, results and discussion sections have been re-organised as suggested.

In addition, the authors should scrutinize the entire manuscript and remove superfluous details. For example, details on storm tracks and characteristics are irrelevant to the interpretation of stream flow and nutrient concentration patterns. See the line-by-line comments for more examples.

This feedback was also echoed by other reviewers so any storm descriptions have been removed/reduced accordingly.

One other thing stands out: The authors make the point that the drought-ending event is unprecedented in magnitude and spatial coverage, which leads me to ask: If the event is so anomalous, what is the value of study? One way to address that question and bolster the relevance of the findings to science, policy, and management would be to discuss how these types of events

may be more common in the future, if indeed that is consistent with projections of future climate for the region.

Although antecedent conditions are not 'equal' in all three catchments before this event, it is interesting because the conditions were similar due to the wider-scale national drought conditions. Given the nature of the storm that proceeded, an evaluation of behaviour across catchments over a wide geographical area was afforded because of the high frequency monitoring infrastructure in place, so although this cannot be considered a replicate controlled event it did, however, have some desirable characteristics to allow an evaluation of responses. We agree that such transition periods from drought to flood conditions may become more frequent in the future given climate predictions and have added some text to this affect in the conclusions (page 23 line 14 – 23). However, the main thrust of the paper is that these unusual conditions allowed for an analysis of the three catchment responses acting under relatively similar conditions, the variety of responses detected indicating the scale of the challenge to environmental managers in tackling nitrogen and phosphorus pollution in rural catchments.

Title:
There are far many more nutrients than N and P. The particular nutrients of this study need to be listed in the title.

As this was the only comment of this kind from all four reviewers and nitrogen and phosphorus are commonly known as the main nutrients of interest in catchment science due to their importance in controlling eutrophication we feel no need to change this.

Also, shouldn't "Demonstration Test Catchments" be written with capitalization in the title?

Yes, this was a typesetting error and has been corrected.

15121.19-24: This sentence is nearly identical to the following sentence. Repetitious information should be removed.

This sentence has been removed.

15121.16: The case for three different, small, unreplicated research catchments serving as representative "of a national scale" is not supported and this characterization does not seem to be relevant to the themes of the manuscript. Perhaps, the authors could write, "at several locations across the UK."

This sentence has been removed.

15124.17: The sentence starting on line 17 is superfluous. The entire paragraph could be modified/deleted to remove the tangential information about the consortium – that information is irrelevant to the presentation and interpretation of the data. If any of it is needed, it would be better suited for discussions, not the introduction.

In line with this and the other reviewer's comments, this whole paragraph has been removed.

15125.6: Here also, I am not convinced of the premise that three study sites are representative of a national scale.

This has been changed to 'multiple sites across England'

15125.12: The air temperature information seems irrelevant.

This section has been removed.

15125.17 and onward: Much of this information is not introductory information and much of it would be better placed in the site description and methods sections.

The authors disagree – this section sets out the aims and rationale of this manuscript

15127.6: I have reason to believe that this sentence is incorrect. Given the information in following sentences, discharge was calculated, not measured. The authors should scrutinize this section to verify that measurements and calculations have been properly described.

Changed to 'calculated'. An extra QA/QC section has been added which should help in this respect (section 2.3, Table 2).

15127.8: Since data were logged every 15 or 30 minutes, the monitoring was "fixed interval," not "continuous."

Changed to 'fixed interval'.

15127.11: Describe what the Doppler flow meters were used to measure.

This has been done.

15127: There was no mention of measurements to validate stream stage or chemistry sensors. This flaw is a considerable shortcoming that must be addressed. Without validation or a description of data QA/QC, the sensor data are NOT PUBLISHABLE.

As mentioned above, a new paragraph on QA/QC procedures has been added which shows the bankside data to be reliable (section 2.3, Table 2).

Furthermore, the figures appear to show stream flow data that were calculated from uncorrected stage data that contained errors. For example, there are unexplained increases in stream flow without any rainfall and abnormal drops in stream flow that are not consistent with reasonable expectations of stream flow recession.

This has been explained in section 3.1 (page 8 line 26 – page 9 line 1) for the Avon as a result of stream support by the local water agency.

Section 2.3: While this information may be somewhat related, it is not necessary. The associated figure, especially, is not needed.
Section 3.1: The authors should consider removing information on meteorological conditions that are not directly relevant. For example, the mention of "low pressure systems and their associated fronts" really has no bearing on stream flow and solute responses to the storm, which are the foci of the manuscript. The same holds for "secondary depressions" and "unsettled conditions."

Section 2.3 and accompanying figure have been removed. The meteorological descriptive sentences have been removed.

Section 3.1: Why haven't the authors described or plotted ammonium or TRP duration

The authors felt the manuscript has a lot of figures and including duration curves for flow, nitrate and TP were sufficient at getting our main point across.

15129.19-29: Including the ranges of nutrient responses during events would be informative to readers.

This has been done.

15130.3: Does "underwent an extreme change" simply mean "showed considerably more variation in concentration than nitrate?" Again, providing concentration variation ranges would help.

We have already stated the concentration change and pre-event and peak event exceedence for all three catchments so find this to be sufficiently explained.

15130.17-18: The authors have no way of elucidating specific hydrological flowpaths from flow duration curves or any other data that are presented in the manuscript. Overall, the topic area of hydrological flowpaths seems to be beyond the scope of interpretation unless the authors can reference other relevant studies on these catchments for the same events, or they present data that address flow and solute concentrations along those specific hydrological flowpaths. Also, this type of supposition, even if appropriately supported by data, belongs in the discussion section, not the results.

See comments above

15132.18-19: The interpretations of sources and hydrological flowpaths belong in the discussion section, but, similar to my previous comment, only if supported by actual documentation of flow and concentrations along those flowpaths.
15134.26/27-15135.7: These methods are inappropriately placed in the results sections.
Section 3.4.1: Interpretations, if supported by observations made in this study, need to be placed in the discussion section.

Re-arrangement of the manuscript now means this has been included in the Discussion.

15135.13-14: No data have been presented on N profiles in soils. Therefore, the supposition is unsupported and hysteresis patterns alone are not sufficient to make definitive statements about specific areas from which N could be flushed. Any attribution to unmeasured sources or source areas is unfounded unless supported by solid evidence with presentation of the data in the manuscript.

This sentence has been removed.

15137.19: How is the citation relevant? To me the wording, "in surface soils immediately adjacent to the sampling location" implies some direct link between the citation and data collected for this study – something such as coordinated sampling or collocated sampling between two different studies. However, the citation is from 2010, and the drought mentioned in this study occurred 2011-2012. How relevant is the citation to the particular conditions during the drought and recovery from drought?

This sentence has been removed.

Figure 1: A larger font size is needed.

This figure has been changed.

Figure 2: This figure is not needed.

This figure has been removed.

Figure 3: Why does flow in Hampshire Avon increase during a period of no rainfall, between 11 and 16 April?
Figure 5: There appear to be irregularities in the stream flow data. Hampshire Avon: There are incomprehensible spikes in stream flow without rainfall and step shifts (a drop in particular) in streamflow between 30 Apr to May 1.

This is because of stream support from the local water agency, description included in section 3.1 (page 8 line 26 – page 9 line 1). On the new time-series graph added (Figure 2) this has also been included so that it is clear to the reader what effect it has on discharge.

Wensum: There is another incomprehensible drop in stream flow between 26 and 27 Apr. Why does stream flow oscillate during high flow between 29 and 30 Apr? These irregularities relate directly to my concerns about the lack of a description of validation of sensor data.

The authors agree that there were some irregular drops in flow on the hydrograph. Since the submission of this manuscript the flow data from the Wensum tributary have been smoothed moving a moving average window of five, and a sentence to this effect has been added to the QA/QC section (page 7 line 28 -30). The smoothed flow measurements from Doppler flow meters

have been shown to have good agreement with gauged discharge throughout the range of flow conditions, Pearson correlation 0.98, $p = 0.00$.