Referee comments are in italics, author response is in plain text

Referee 1

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, $p = 0.00$.

Referee 2

This paper uses high-frequency nutrient monitoring data to characterise responses to a major rainfall event across 3 contrasting catchments. I really like the concept of examining hysteresis behaviour to the same storm event across the UK, and it is a really novel use of the detailed data generated by the DTC catchments. This paper is original, of interest to an international audience, and generally well written. However, it would require some corrections and restructuring before I could recommend publication.

The authors thank the reviewer for this comment, we are pleased that you liked the concept of the paper and found the hysteresis comparison to be novel.

General comments

1) *The paper is very overlong and contains much repetition. It needs to be much more focussed and more clearly structured. The introduction in particular needs to be greatly reduced and refocused. Much of the text is justifying the DTC project. This has probably been presented elsewhere in other papers, and should be kept to a minimum here.*

In fact, at present background information on the DTC project hasn't been published elsewhere, hence the overlong description here. Other reviewers made the same comment and so this section has been reduced.

2) *A large section of the introduction describes the storm event from a meteorological perspective, which is not needed to support this study. Much of this is repeated in the Methods anyway, and the text in the Introduction should be removed. The intro contains other descriptions of methodology and site descriptions that should also be omitted. Once the above items have been removed, the remaining Introduction is very weak, and there is a real need to put this study into context. This should be done by including a brief review of similar hysteresis studies. The authors state (probably correctly) that no previous studies have captured hysteresis patterns at high frequency in catchments across the UK. However, there are lots of studies that have captured storms at multiple sites across catchments, and also studies that have studied hysteresis at single sites at high monitoring frequencies. Suggested references include –*

Bowes, M.J., Smith, J.T., Neal, C., 2009. The value of high-resolution nutrient monitoring: A case study of the River Frome, Dorset, UK. J. Hydrol. 378, 82-96.

House, W.A., Warwick, M.S., 1998. Hysteresis of the solute concentration/discharge relationship in rivers during storms. Water Research 32, 2279-2290.

Ide, J., Haga, H., Chiwa, M., Otsuki, K., 2008. Effects of antecedent rain history on particulate phosphorus loss from a small forested watershed of Japanese cypress (Chamaecyparis obtusa). J. Hydrol. 352, 322-335.

McKee, L., Eyre, B., Hossain, S., 2000. Intra- and interannual export of nitrogen and phosphorus in the subtropical Richmond River catchment, Australia. Hydrol. Process. 14, 1787-1809.

Siwek, J., Siwek, J.P., Å˙zelazny, M., 2013. Environmental and land use factors affecting phosphate hysteresis patterns of stream water during flood events (Carpathian Foothills, Poland). Hydrol. Process. 27, 3674-3684.

Stutter, M.I., Langan, S.J., Cooper, R.J., 2008. Spatial contributions of diffuse inputs and within-channel processes to the form of stream water phosphorus over storm events. J. Hydrol. 350, 203-214.

The authors thank the reviewer for this comment. We have now added a paragraph into the Introduction about the value of hysteresis studies (page 3 line 23 – page 4 line 9). The meteorological information has been removed.

2) There is no mention of how the Hach phosphate chemistry and probe data was quality checked. As the study is wholly reliant on this time-series data, its quality is vital. Were these automated data corrected using analysis data from spot samples? How were the probes calibrated? How often?

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 between data collected using bankside analysers and grab samples analysed using standard laboratory procedures which show the measurements are in good agreement (Table 2).

3) The paper structure is very confusing. There is lots of discussion within the Results section, but then many observations are left unexplored. I think the readers (and the authors) would benefit from having a combined results and discussion section, if that is allowed within the format of this journal.

This comment has been taken on board and the interpretive elements have been moved from the Results to the Discussion section.

The conclusion is really just another summary of the manuscript, and adds nothing to the paper. The conclusion should state why this work is worthy of publication.

The conclusion has been edited to reflect the unique contribution of this work.

Specific comments
Title: (and throughout). High resolution could refer to either spatial or temporal resolution. Replace with high temporal resolution or high frequency.

This has been done.

Demonstration Test Catchments should be capitalised.

This was a typesetting error and has been corrected for.

Abstract Page 15122, line 9. Studying one extreme storm event after a prolonged drought does not indicate the size of the nutrient pools in the catchments (unless they became exhausted). The study would need to extend across multiple events for an extended time period to quantify this pool (i.e. until exhaustion was actually observed). Please remove this statement.

This has been done.

Line 26. How are point sources "increasingly controlled"? Replace with "Improved nutrient removal at WWTPs are reducing point source inputs".

This has been done.

Page 15123, line 24 to 15124, line 28. Delete paragraphs.

This has been done.

15125. Line 5 (and throughout) Change resolution to frequency

This has been done.

Line 6 – 16. Delete (Repeated in Methods)

This has been edited to lead onto the aims of the paper.

Line 19 – 26. Delete (should be in methods)

This briefly outlines the aim of the paper so has not been removed.

Section 2.2. Give brief method descriptions for the nitrate and phosphate analysis. How was the data quality controlled? Is this data raw, or is it corrected against spot sample analysis?

See comment above.

Section 2.3. Greatly reduce word count or delete. The study only requires data on the timing and quantity of rainfall to each catchment, which is given in Figure 3.

This paragraph has been deleted.

Results.
There is lots of discussion of (some of) the results here, and I think a combined Result and Discussion section would be most appropriate.

See comment above.

Section 3.1, line 16. Not true. The Eden responded to the rainfall event on the 12th April.

This sentence has been removed.

Line 23-25. Delete the meteorological causes of the rainfall events.

This has been done.

Page 15129, line 21. Doesn't the dilution of the nitrate concentration imply that there is little diffuse nitrate input to the Avon either from throughflow and overland flow? This result suggests groundwater input domination.

This is discussed in depth in the Results section (page 9 line 8 – 9) and then in the Discussion section (page 13 line 11 – 16; page15 line 29 – page 16 line 14).

line 26. There is no evidence that the nitrate input to the river is by throughflow (although I agree it might be). The authors need to add lots of caveats with their speculations on nutrient sources.

This has been removed from the results and discussed in depth in the discussion (see above comment).

15130, line 2. Delete "during this event".

This has been done.

Line 17. There are 2 major sources that seem to be ignored within this paper; groundwater inputs (for the Avon and Wensum) and storage of phosphate within the bed sediments. These major inputs would explain why nitrate is diluted during storm events i.e. the loops have a negative gradient, and why P and ammonium respond so quickly to produce clockwise hysteresis.

The authors would like to point out that groundwater contribution was already included in the Discussion but has been further discussed with the inclusion of borehole concentration data to strengthen arguments (see comment above for Avon, page 13 line 23 – page 14 line 9 for Wensum). The storage of P within the bed sediments has been added as a potential source in the Avon (page 17 line 18-19) but was already included in the discussion for the Wensum interpretation (page 18 line 6).

15131 line 17. See comments above.

See our response above.

Line 25. To interpret this observation correctly, the reader (and author) needs to know the nitrate concentration of the ground water. Please provide for the Wensum and Avon.

This has been done (page 9 line 9 for the Avon, page 13, line 24 for the Wensum).

15135, line 20. Should it be left of the plot?

Yes! Thank you for this observation!

15137. Line 6. Are there any sewage treatment works in the catchments? If so, please give details in the study area descriptions. If so, how does this change your interpretation of the results?

There are no sewage treatment works as these are headwater streams, however, there are a number of septic tanks in the area which could influence the water quality which were already included in the discussion for the Avon where they are thought to have most impact (page 17 lines 25-28).

15141, line 9-21. Repetition. Delete.

This has been done.

15143. Line 25-26. What ongoing research? Is there a reference? Is it future work within DTC?

Has been edited to make clear this is referring to the wider scientific community.

Line 27. Delete temporal.

This has been done.

Section 5. Delete summary of findings within this section. These are already given previously and in the abstract.

This has been done.

Figure 1. Remove shading for London. It looks like one of the study catchments.

This has been done.

Figure 2. Delete. This meteorological data is not required.

This has been done.

Referee 3

A Summary of the Manuscript's Content

The manuscript describes the monitoring of three streams in England during April 2012. The three streams are the Wylye, Blackwater Drain and the Morland and these drain catchments that contrast in terms of location, climate, soil type, geology and land cover. The data presented describe the changes in the flow and streamwater nitrate, ammonium, total phosphorus and total reactive phosphorus concentrations during a major storm event coming at the end of a major drought. The chemical data were collected using instrumentation deployed on the stream bank and therefore issues with sample degradation were avoided. The measurements were made every 30 minutes and are part of a much larger dataset being collected as part of the Demonstration Test Catchment programme. The streamwater concentrations were set in the context of data collected over one hydrological year by plotting the flows and concentrations on flow-duration and concentration duration plots. The results of a storm event hysteresis analysis were used to explore transport pathways and sources of inorganic nitrogen and total and total reactive phosphorus during the storm event, and the analysis was also used to indicate whether the diffuse nutrient sources were source or transport limited. It was concluded that source limitation did not occur in these agricultural systems when the frontal system passed in April 2012.

Evaluation

The article presents important datasets and provides interesting and useful insights into the hydrochemical dynamics during a storm. The subject of the paper is of definite interest to the readers of HESS and a broad international audience, and these new data add to our understanding of catchment hydrochemical functioning. Before the paper can be accepted, I feel there are a number of major issues that need to be addressed by the authors regarding the work, and therefore this review focuses on points of clarification, rather than an in-depth discussion of the results.

The authors thank the reviewer for these comments and we are pleased that you feel this paper is relevant for publication in HESS and that the data add to understanding of catchment functioning.

General Points

1. The title reflects the contents, though it seems unnecessary to include 'demonstration test catchments' at the end. If this is to be kept, then capital letters should be used: 'Demonstration Test Catchments'.

This was a typesetting error and has been corrected.

2. The abstract alone is sufficiently informative, even when read in isolation, however there is an (over)emphasis on describing the supporting research programme which is perhaps unnecessary and makes the abstract too long. The same issue occurs in the introduction.

We have removed superfluous information about the research programme from the Abstract and Introduction.

3. One of my main concerns is that I do not feel the overall aim of the paper is made clear. What do the authors hope to gain by studying the effects of the same frontal system moving over the three catchments that could not be done by looking at all the extreme events in the data collected since

2011? If it is because of the highly unusual nature of the April 2012 event (coming at the end of prolonged drought), then the analysis needs to account for the different antecedent conditions, reported on page 15128, line 20 to page 15219, line 18. There appears to be an assumption that all three systems are at a similar baseline at the onset of the rainfall event on the 25 March 2012, but I am not convinced by this. For the Wensum, higher rainfall values were reported for March, 2012 than at the other two study sites. The authors note 'The Wensum DTC, by contrast, was already exhibiting relatively high flows before the first event (5.9% exceedance), due to heavy rainfall at the end of March and continued wet conditions in April 2012.' (page 15129, lines 10-12). The different baseline conditions need to be accounted for in the analysis and these differences make comparison difficult and must have an effect on the interpretation of the nutrient transport behavior in the dry-to-wet transition.

The authors agree that some of the more general sentences about the drought condition across the country give the impression that the catchments were all experiencing identical antecedent conditions before the storm event. These sentences have been removed. We did however include a discussion of antecedent conditions before the event with the flow duration curves. We have now, partly in response to other comments made by this reviewer and others, included the time-series data from the hydrological year 2011-2012 (Figure 2) and as part of this we have used an Antecedent Precipitation Index (API) as an indication of antecedent catchment 'wetness' throughout the water year, including a discussion of the API values before the storm (see section 3.1 and 3.2). It is correct that the Wensum had experienced wetter conditions in the run up to the storm and had also experienced a large event in March which we have pointed out as a possible explanation for the quick phosphorus exhaustion in the hysteresis discussion. Reference to antecedent conditions affecting storm response can be found: page 15 line 30 –page 16 line 2, page 17 line 3- page 18 line 1 for the Avon; page 16 line 28 -33, page 18 line 12-16 for the Wensum; page19 line 21-24 for the Eden.

4. As only one event is described in detail, it is difficult to know if the behavior reported in each of these systems is typical, or not. The flow-duration and concentration-duration curves show that the flow and chemical response was extreme within the year, however there is no sense of what the flow and streamwater concentration dynamics are like in general. Given the variation of chemical response expected and the extent of the dataset with measurements made since 2011, it appears a major limitation not to have explored the full dataset to comment on the range of storm responses to provide a comparison with this extreme event (see specific comment 9).

As mentioned above we have taken this comment on board by including time-series data for rainfall, discharge, API, nitrate and TP concentration data for the water year 2011-2012 (Figure 2). The storm period discussed in more detail is highlighted in grey so that the reader can see how it fits in to the context of the temporal dataset for each catchment. We feel this provides the reader with a good deal of information about the general hydrological regimes of the three catchments (section 3.1 and 4.1) and the antecedent conditions that may account for certain storm responses observed within the dataset (section 3.2 as well as comments mentioned above). We discuss this figure in Discussion (section 4.1), along with other relevant information for interpretation of the data such as geology, groundwater nitrate concentrations etc. that build a conceptual model of catchment functioning for

each site. We feel the inclusion of these data has strengthened the paper so we thank the reviewer for this helpful feedback.

5. I am not sure why the authors have modelled the rainfall event when they have tipping bucket gauges in each catchment. The data from the gauges do not appear to have been used other than in Figure 2.

This section has been removed.

6. The catchment descriptions are too brief. There is no overview of how the hydrology or hydrochemistry compares across the three study sites. This is important given the paper is about comparing the flow and chemistry response to a single storm event across the three different sites. Without a discussion of the normal hydrological regime of the individual catchments it is difficult to distinguish differences in storm response between the catchments.

We have increased the amount of information regarding geology and soils in section 2.1. We also believe that these concerns have been addressed with the inclusion of Figure 2 and relevant discussion in section 4.1, described above.

Much is made of the importance of fertiliser applications controlling the concentration changes, but there is only a limited description of the land use and no data on fertiliser application rates within the catchments.

This has been included in section 2.1 and Table 1.

There is no detail on point sources; the work seems to focus on diffuse sources only.
Does effluent, either from sewage treatment works or septic tanks, influence the observed chemical dynamics?

As these sites are on headwater streams there are no sewage treatment works. Septic tanks are only considered to be a significant source of nutrients in the Avon and this has already been included in discussion.

The cause of the nitrate rich baseflow in the Hampshire Avon is not really discussed. Is the resuspension of bed material important for explaining the phosphorus dynamics?

The authors would like to point out that groundwater contribution was already included in the Discussion but has been further discussed with the inclusion of borehole concentration data to strengthen arguments (page 9 line 8 – 9) and then in the Discussion section (page 13 line 11 – 16; page15 line 29 – page 16 line 14 for the Avon, page 13 line 23 – page 14 line 9 for the Wensum). The storage of P within the bed sediments has been added as a potential source in the Avon (page 17 line 18-19) but was already included in the discussion for the Wensum interpretation (page 18 line 6).

7. There seems to be some confusion as to whether the Blackwater, a tributary of the Wensum, is the study site or the Wensum itself. This is important because the geology seems to be quite different in the Blackwater from the rest of the Wensum. The references to the 'Wensum DTC', 'Eden DTC' and 'Hampshire-Avon DTC' rather than the name of the study site causes confusion because it is unclear if the authors are referring to the larger catchments of the Wensum, Eden and Hampshire-Avon or the smaller study areas where the sub-daily measurements have been made.

The authors thank the reviewer for this comment. The manuscript has now been edited to reflect the fact that the monitoring sites discussed are located in the headwater tributaries of the Avon, Wensum and Eden wider catchments and have been changed to the Wylye, the Blackwater and the Newby Beck.

8. I would recommend that a section is included on the quality assurance of the sub-daily data. How robust are these data compared to laboratory based analyses. Were corrections made for temperature effects or drift?

We have now included a section on QA/QC procedures (section 2.3) which we agree are essential for validating the high resolution data. 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.

9. The storm event responses are described in great detail but I'm not sure if all this detail is relevant. Could the paper be made shorter and more focused by reducing the level of detail in the results section and drawing out the salient points, and then placing the results in the context of other storm events measured since 2011? This would be very interesting.

We thank the reviewer for this comment and have taken this on board. We have shortened the description on the storm response and have, as mentioned above, included Figure 2 which shows the hydrological and hydrochemical regime for the three sites for a full water year in sections 3.1 and 4.1.

10. Given the focus of the paper is a comparison of the response across the three sites, it might be better to cut down the description of the individual storm responses and focus more on this comparison. In addition, I'm unsure what new has been learnt about storm event responses in these catchments, or how the findings fit into the context of the broader literature (e.g. Ferrant et al., 2012; Melland et al., 2012, Mellander et al, 2012a, b and the list of references given the paper). Please can the authors put the results into context? It would be excellent to see if these results confirm or contrast with other studies.

We thank the reviewer for this comment and have now cut down the individual descriptions as suggested and put the findings into the context of the broader literature throughout the discussion (page 15 line 1-19, page 16 line 14-18, page 17 line 8 – 12, page 18 line 1-3, page 18 line 22-25, page19 line24 – 27).

11. I feel the conclusions require further consideration as I do not believe that all the potential nutrient sources have been fully evaluated in the interpretation of the results. For example, there is little consideration of effluent from sewage sources or septic tanks, or of inputs of groundwater enriched by nitrate in the Chalk dominated catchments.

See comments above.

12. Table 3 is not necessary. Loads are presented in this table, but loads are not considered in the aims, discussion or conclusions.

The loads are in fact discussed in what was section 4.1, now section 4.3 and form an important part of the Discussion.

13. Figure 8. The results for total phosphorus in the Hampshire-Avon look odd in that the maximum concentration appears to be exactly 1 mg P l^{-1} and the measurements seem to remain constant at this concentration even though the flow decreases from approximately 1.7 to 0.7 m^3s^{-1}. Is this a data transcription error or is it suggesting exhaustion of supply (see specific comment 22)?

These points occur due to the 1 mg P L^{-1} limit set on the Phosphax instrument at this time. This has been made clear in the manuscript and the points have been removed from the hysteresis plot.

Specific Points

1. 15214, 23-28. Is the reference to the examples from Australia and Ireland relevant? The details regarding the Demonstration Test Catchments add to the length and it is unclear why these are needed in the context of this article.

This has been removed.

2. 15215, 3. 'Numerous authors' is rather vague. Please could you be more specific?

This sentence has now been removed.

3. 15215, 19. Did all three catchments experience drought stress and then flood conditions? What were the hydrological conditions in all three catchments prior to, and after, the storm?

See comment above regarding the inclusion of the time-series data.

4. 15217, 27. What does 'ARW' stand for? Please could you define acronyms on first use?

This whole section has now been removed.

5. 15128, 4. Please don't use colloquial terms such as 'nudging towards'. What is meant here – bias correction?

This whole section has now been removed.

6. 15128, 12. What is the definition of 'winter 2011-12'? Is this December 2011 to February 2012? I thought that March 2012 was also dry in some catchments. Please could you clarify the period that you are considering?

This sentence has been removed.

7. 15128, 20 to 15219, 18. These lines suggest that there were differences between the antecedent conditions in the three study sites. How did this affect the results? Does it really matter that you have captured the same weather front as this will manifest itself in different ways in the different catchments. The key is measuring the transition from very dry to wet conditions in a short space of time, but what will this tell us that other analyses of storm event responses haven't done already? Please could the authors explain?

See comments above.

8. 15219, 23-26. Could the high nitrate concentrations in the Wensum be due to the transport of groundwater, enriched with nitrate, from Chalk?
9. 15130, 15-29. Again I would suggest that nitrate enriched groundwater is an important source of nitrate in the Hampshire-Avon and Wensum, which are both predominantly Chalk catchments.

See comments above.

10. 15131. Be careful with precision here. Can you really quote runoff to 0.01 mm per hour?

The runoff values have now been reduced to 1 decimal place.

11. 15134, 4. Change 'NRL' and 'NFL' to 'N_{RL}' and 'N_{FL}'.

This has been done.

12. 15134, 6. What is meant be 'more hysteretic'? Do you mean a greater difference between the concentrations on the rising and falling limbs of the hydrograph?

Yes, this has been clarified (page8 line17-19).

13. 15135, 3. Are the dilutions caused by relatively clean water entering the system and diluting effluent inputs? The authors acknowledge that sewage will affect the ammonium concentrations, but why is sewage not important for nitrate and phosphorus?

This section of the text has actually been removed by means of reducing the detail on the individual storm responses. However, the dilution is largely of the nitrate-rich baseflow, as demonstrated by the similarity in concentration of the chalk groundwater and the stream baseflow concentrations. There are septic tanks in the catchment but no sewage works, the contribution of which have been acknowledged for P transfer (see comments above).

14. 15135, 4. Are the increased concentrations caused by the flushing of mineral N from the upper soil layers, or is it the input of groundwater enriched with nitrate?

This sentence has now been removed.

15. 15141, 1-13. Was the pre-storm period drier than 1997, or wetter?

Unfortunately we are not able to say as this information was not included in the Jarvie et al. 2002 paper.

16. 15142, 7-10. How can you be sure that you are separating the effects of transport- and source limitation? Please could a note be added to the methods section? Would it be useful to refer to the paper by Evans and Tranter (1998)?

This sentence has been removed.

17. 15143, 7-8. A hydrological year is not very long in terms of capturing a good representation of the full range of hydrological conditions. If you look at the series of storms in one hydrological year you will gain a better understanding of the dynamics, though you won't capture the full range of conditions.

We agree with the reviewer that one hydrological year will not represent the full range of hydrological conditions. However, as made clear in Figure 2 the high resolution bank side analysis in the three DTCs only commenced in 2011-2012 which means that we don't have historical records for all of these sites.

18. 15144. 25-27. I would suggest that to understand sub-hourly data you need long-term, low frequency data to determine the context of the sub-daily dynamics.

We agree with the reviewer, but, as mentioned above, but such data are often not available, especially if working in headwater streams which are not included in longer-term national monitoring studies.

19. 15145, 16. What is the evidence for transport limitation?

This sentence has been removed.

20. 15145, 17. Wensum – is it drainage, or groundwater nitrate?

This has now been more clearly explained in section 4.1 (page 13 line 23 – page 14 line 9). The groundwater concentration of the chalk in this part of the catchment is extremely low (<1 mg N L^{-1}) so cannot explain the peaks experienced during storms. We have provided concentration data for the tile drainage in the upper part of the catchment, some of the deeper drains flow all year round with summer concentrations of up to 10 mg N L^{-1}, which is a more likely explanation for the rise of

nitrate on the recession limbs of storms when the catchment is experiencing greater levels of saturation (demonstrated by the API values).

21. 15145, 26. No nutrient exhaustion, but TP in the Hampshire Avon maxed out at 1 mg P l^{-1}, does this indicate that there might be some exhaustion of supply?

This has been explained above.

22. Table 1. The geology is described, but not the soils.

This has now been added.

Is the elevation the mid-point of the catchment? Please could the elevation be specified as a range?

This is the elevation of the sampling point, which has now been specified in the table.

23. Figure 1. The catchment areas of the Eden, Wensum and Hampshire-Avon are shown on the map, but these are different from the study areas, which are much smaller.

We have now edited this figure to show a close up of each catchment, with the specific monitoring point and tributary discussed in this paper highlighted.

24. Figure 4. It would be better to present the flow-duration and concentration-duration plots on a normalized scale, so that the extremes can be seen more clearly?

Effectively, the x-axis is normalised so that the exceedences can be compared. The authors think it is more useful to keep the y-axis as it is to show the differences between the catchments in terms of flow or concentration values.

References

Evans, C. D. and Tranter, M. D. 1998. Causes of concentration/discharge hysteresis and its potential as a tool for analysis of episode hydrochemistry, Water Resour. Res., 34, 129–137, doi:10.1029/97WR01881, 1998

Ferrant, S., Laplanche, C., Durbe, G., Probst, A., Dugast, P., Durand, P., Sanchez-Perez, J .M., and Probst, J. L. 2012. Continuous measurement of nitrate concentration in a highly event responsive agricultural catchment in south-west of France: is the gain of information useful?, Hydrol. Process., doi:10.1002/hyp.9324.

Melland, A. R., Mellander, P.-E., Murphy, P. N. C., Wall, D. P., Mechan, S., Shine, O., Shortle, G., and Jordan, P. 2012. Stream water quality in intensive cereal cropping catchments with regulated nutrient management, Environ. Sci. Policy, doi.10.1016/j.envsci.2012.06.006.

Mellander, P.-E., Jordan, P., Wall, D. P., Melland, A. R., Meehan, R., Kelly, C., and Shortle, G. 2012a. Delivery and impact bypass in a karst aquifer with high phosphorus source and pathway potential, Water Res., 46, 2225–2236.

Mellander, P.-E., Melland, A. R., Jordan, P.,Wall, D. P., Murphy, P., and Shortle, G. 2012b. Quantifying phosphorus and nitrogen transfer pathways in agricultural catchments using high time resolution data, Environ. Sci. Policy, doi:10.1016/j.envsci.2012.06.004.

Referee 4

This manuscript reports the detailed information on N and P concentrations collected from three DTC catchments in England using on-site monitoring equipment, during the high flows in late April 2012 that followed an extended period of drought conditions. The data collected are of high quality and demonstrate a number of interesting features of the catchment responses. However, having read the ms I was left uncertain as to the main purpose of the manuscript and whether the data presented merited publication in a high level inter-national journal in its current form. My primary concerns are listed below:

We thank the reviewer for this comment and we are pleased that you think the data are of high quality which demonstrates interesting features of catchment responses.

(1)The introduction contains a lot of superfluous information on the background to the DTC catchments. I would see this as being of limited relevance to the main thrust of the study. The key point would seem to be the existence of three catchments instrumented with similar equipment and representative of rather different landscape types.

This information has been substantially reduced.

(2)The modelling of nationwide rainfall patterns for the study period would seem to be of limited value/relevance. Representative rainfall records from each of the catchments would seem to be all that is required. By focussing on a short period as distinct from a longer time-series, the rainfall data presented obscure the important difference in antecedent conditions between the Wensum and the other 2 catchments (see below). I would like to see rainfall records covering a longer antecedent period.

The modelled rainfall section has been removed. We have now, in response to this comment and comments elsewhere, included time-series data from all three catchments for rainfall, discharge, nitrate and TP concentrations for the hydrological year 2011-2012 (Figure 2). We have also plotted up on the same graph Antecedent Precipitation Index (API) values for each catchment throughout the year as an indication of antecedent catchment 'wetness' throughout the water year, including a discussion of the API values before the storm. We feel this helps the reader understand how antecedent conditions effect the hydrological and hydrochemical response to rainfall throughout the entire year and not just for the storm period discussed in detail. This information can be found in sections 3.1, 3.2 and 4.1.

(3)The reason why attention focussed on the nutrient response of the catchments at the end of a drought period and the importance of the findings is not made clear. Is there a suggestion that such events are critical in terms of ecological impact or that changing weather patterns in the UK could make such conditions more common and that an improved understanding is therefore required?

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.

(4)The basis for the comparison presented is not entirely convincing. The absence of N data for the Eden would seem to be a major problem, if the intention is to compare the response of different landscape types. It is important to know how the N concentrations in the Eden catchment responded to complete the story. This gap represents a large hole in the study.

Because of the nature of the bankside monitoring there are down periods for various instruments – it is certainly a challenge to maintain the equipment. The authors chose to study this particular storm because we felt the transition from dry to wet conditions (albeit with different antecedent conditions) was an interesting basis for comparison across the three catchments, but unfortunately there was no nitrate data for the Eden and no TRP data for the Avon at this time. However, the inclusion of the new Figure 2 show the more continuous nitrate data record for the Eden which is actually complete for the remainder of the year. We have included in the discussion the general hydrological and hydrochemical trends revealed by the high frequency data (sections 3.1 and 4.1) which we believe complete the understanding for the three catchments even though there are some gaps for the storm period.

I was also not convinced that the antecedent conditions in the three catchments were very similar. Fig. 4 indicates that the antecedent flows for the Wensum were flows equalled or exceeded <10% of the time, whereas the equivalent values for the Eden and Avon were 60% and 80% respectively. This latter discrepancy makes the comparison dubious.

The authors agree that some of the more general sentences about the drought condition across the country give the impression that the catchments were all experiencing identical antecedent conditions before the storm event. These sentences have been removed. As mentioned before, the addition of the API values to the time-series in Figure 2 and further discussion of API values right before the storm acknowledges the fact that the antecedent conditions were different (section 3.2), which is then used to further interpret the storm responses. For example, the Wensum had experienced wetter conditions in the run up to the storm and had also experienced a large event in March which we have pointed out as a possible explanation for the quick phosphorus exhaustion in the storm event studied in detail here. Reference to antecedent conditions affecting storm response can be found: page 15 line 30 –page 16 line 2, page 17 line 3- page 18 line 1 for the Avon; page 16 line 28 -33, page 18 line 12-16 for the Wensum; page19 line 21-24 for the Eden.

(5)If the purpose of the study was to provide an improved understanding of post drought catchment response, there would seem to be a need to compare this response with the more standard storm period response at different times of the year. Looking at the post drought period essentially in

isolation would seem to be of limited value. I would also like to see detailed records of flow and nutrient concentration for the entire 2012 year presented in order to demonstrate precisely how the post drought response fitted into the longer term response. The 'duration curve' data are of limited value in this context. How does the hysteretic behaviour of the post-drought periods compare with that at other times of the year? The three catchments would seem to offer far more scope for comparison than that attempted in the study reported.

We thank the reviewer for this constructive comment, which, as already mentioned, was part of the motivation for including Figure 2 and wider discussion around how the post drought response fitted into the longer term response. We felt that it was unnecessary to include hysteresis analysis for other periods as we were using the hysteresis as a tool to explore the different catchment responses to this particular storm period and that the time-series data provide a wealth of information for the wider functioning of the catchments at other times of the year.

(6)The relevance of the analysis of hysteretic behaviour to the overall purpose of the study is not entirely clear and needs to be explicitly stated. If it is going to be included, there would appear to be a need to relate the post-drought behaviour to the hysteretic behaviour at other times.

See comment above.

(7)Towards the end of the ms attention shifts to emphasising the value of the detailed data provided by the onsite monitoring equipment. This introduces a different topic which needs to be treated more fully if it is important. Is the equipment uses really so novel? Details of its calibration and the accuracy and precision of the measurements and the data recovery success need to be reported. The lack of N data for the Eden during the events under consideration could suggest that there were large gaps elsewhere in the record.

In response to this comment and the other reviewer's feedback we have now included a paragraph on the QA/QC procedures to validate the bankside analysis with more standard analysis of grab samples in the laboratory, which show good agreement. We believe that this type of monitoring is advanced as there are only a handful of other research groups which have bankside analysis of nitrate, TP and TRP, particularly at three distinctly different catchment typologies. As mentioned, the lack of nitrate data for the Eden during this storm is unfortunate but the time-series data for the remainder of the hydrological data show that this is a good data set. We appreciate that this was not necessarily obvious to the reader before the inclusion of the time-series.

(8)There is a great deal of rather vague inference regarding flow paths etc and comparisons between the different catchments. This is based primarily on analysis of the hysteretic behaviour and it is well known that such analysis is far from definitive. More rigorous analysis is required to confirm the interpretations presented. To be meaningful such inferences ideally need to be confirmed by analysis of the behaviour of other geochemical parameters and particularly isotopic tracers capable of distinguishing old and new water etc.

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.

This raises issues of the purpose of the DTCs. If they aim to quantify the effectiveness of mitigation measures in terms of % change etc the instrumentation installed should provide useful results, but if the aim is to understand the response of the different catchments a more comprehensive monitoring strategy is arguably required.

The reviewer noted earlier that there was superfluous information on the DTC project in the introduction and they are correct that this paper is not aimed at explaining the entire purpose of the wider DTC project. We haven't outlined how we intend to detect improvements in water quality in response to mitigation measures as it is beyond the scope of this paper. We therefore find no reason to respond to this comment further.

(9)As indicated above, there is much subjective inference regarding contrasts in the response of the different catchments to the post-drought wetting. It was far from clear how the suggestion that the TP response of the Eden catchment was transport controlled was justified. If much of the P was PP derived from surface sources it would seem almost certain that this would be supply limited rather than transport limited. Lack of exhaustion is not in itself evidence of lack of a supply limitation.

These unclear sentences have been removed.

(10)I was expecting some discussion of the extent to which the results presented conformed to other studies or existing understanding and thus their wider importance/ significance.

This has now been done.

(11)I am not familiar with the geology of the study catchments, but I found the reference to a clay layer underlying the Chalk and Greensand unclear. Are you referring to these rocks being underlain by clay strata or simply the existence of a clay layer at their base? More generally, I would see a need for a fuller description of the three catchments so that readers can link the results to the catchment characteritics more fully.

The geology of the Avon catchment has been clarified in the manuscript (section 2.1). A more detailed description of each catchment has been added including soil information in Table 1. Figure 2 also shows the more general hydrological and hydrochemical regimes for each of the catchments which should help the reader make distinctions between the three catchments.