Overall Response to Everyone

Monday 12th January 2015

We would like to thank the editor and reviewers once again for their comments. We think that the questions raised by the reviewers were insightful and highlighted several omissions from our paper. We hope that the extensive edits we have carried out will show that we have taken great care to address these concerns. In the document below we have listed all the (numerous) edits we have made to the text, and illustrated the sections we have modified in the main text. We have added 6 new references to support the document. The tone of the comments were complementary to most of the aims of the model but need for detail and clarification have encouraged many modifications. In our response to the reviewers and editor we agreed with them on all points other than how good the simulated 'fit' of the model was. We have thus clarified our evidence to this point in the document and modified several diagrams and removed one. We have introduced one new figure that makes the most of the high frequency dataset (of flow and nutrients) to support the aims of the model and to strike a balance between the degree complexity and physical meaning. We have also added more material about the evolution of the MIR model and how it led to the final model structure, and to state clearly what has not been modelled (such as nutrient cycling, riparian processes and within channel processes) stressing that their effects are being effectively integrated into the MIR structure. The temptation to add more processes is high but we are trying to match the model to the observable data and the needs of the policy makers. We have stressed that more complexity is not justified until more data become available and that sensitivity and uncertainty issues are demonstrated.

.

Reply from Editor (December 2014)

As you have seen, all 3 reviewers appreciate the general intention of your manuscript, but they also stress that it is not yet developed to the point. They point out the fact that, although simplicity is frequently required and even desired in hydrology (in particular for communicating results and interpretations to a wider audience, including policy makers), the model, as it currently formulated, may be too much of a simplification. The reviewers agree that it will be necessary to document the "road" to MIR in a more systematic way, illustrating and discussing which processes can be omitted and why, and which need to be in a model, as at this point MIR is a largely unsubstantiated claim. This could for example be done in a stepwise analysis, by one by one adding simple(!) representations of different biogeochemical (e.g. time-varying concentrations in the stores; more explicit processes linked to P or N, etc.) and hydrological (e.g. wetlands) processes. From this analysis it can then more objectively be inferred what the MIR really is to still allow for valid interpretations.

In addition I would like to encourage the authors to link their work a bit stronger to a wide body of recent literature on the topic, e.g. Benettin (2013; WRR), Botter (2010,WRR; 2011,GRL), Harman (2014,GRL; 2014,WRR), Hrachowitz (2013; HESS); van der Velde (2010,2012; WRR)

Please make sure, in addition, to address *all* reviewers concerns in detail in the revised manuscript and to also provide a point-by-point overview of the responses.

General response to all Reviewers (November 2014)

The onus is on us as the authors to make our choice of model processes and how the model was applied and calibrated to be clearer. Hence we will modify the introduction to stress this. Our attempt to model flow, N, TP and SRP simultaneously is ambitious and we were keen to show the reality of how difficult this task is. Hence we consider the degree of fit we have to the observed data is still good enough that it can act as the basis of a management/decision making tool.

We feel that R1 and R2 and to a lesser degree R3 have perceived that the model did not 'perform' well in terms of reproducing concentration (especially TP). However we were not trying to get a good fit or to over calibrate and optimise the model. We feel this 'curve fitting' exercise is inappropriate to the Meso-scale and to times series of nutrient data with only outlet data is available (weekly or monthly data). We spent some time showing sub-daily concentration data to show that the pattern is very random and that achieving a general fit to storm events, background and to the seasonal shift is possible. We were also explicit about the error in our scatter plots (Fig 5). We agree that we must stress that our calibration is based on our expert judgement and that we are trying to rely a conceptual pattern of runoff based on the observed data and our knowledge gained from research studies. Detailed 'curve fitting' may make the results appear better (which we could have done) but this is not helpful to the policy maker as they want to know that we have not tailored the model to fit the data. We feel that simulating hillslope farming effects at the meso-scale is of value to future management and to the WFD.

We need more than an export coefficient approach as it shows why and when export differ from event to event. A spatial model is not justifiable as this has too many parameter (a point made by all the referees). We could add more processes (for example: river and riparian processes) but the data does not justify it hence these processes were not included which is the MIR modelling philosophy. So we agree that we do need to clarify this.

Some more specific points were raised by the reviewers and this is our response:

- The MIR approach, which has led to the development of the CRAFT, needs to be described more clearly. It is important to show how the process of developing a MIR model is systematic and all 3 reviewers felt that it could be improved in the paper. We structured this paper such that the final MIR model was shown but omitted much detail in terms of eliminating more complex model structures etc. This was mentioned by all reviewers, for example R3 noted that the model structure appeared to be "*chosen at the beginning of the study*". Hence we will add a section covering the provenance of the MIR model using more references to earlier work and versions of the model.

- In more detail, R1 commented on the absence of a riparian process component in the CRAFT, which would simulate sources of sinks of nutrients into the channel system. We argue that although we are aware of such processes that our model can reproduce the typical weekly or monthly time series of nutrient concentrations collected at catchment outlets. Another approach would be to develop a riparian component and then discard it later if it failed to improve on the model results (and add to the number of parameters). We settled on a three-compartment hillslope model, R3 found this surprising but did not consider it to be unacceptable (we argue that the results in terms of predicting flows (Q) validates the viability of this model structure). The paper would be extremely long if all options were discussed explicitly but we will add a paragraph to clarify the model development process. Hence we will acknowledge that these processes are important and explain why we have omitted them.

- The structure of the revised paper should perhaps pose a single research question relating to the MIR concept more strongly. We agree with R2 that the Case Study (currently Sec 2.2) should probably come before the description of the final MIR model (CRAFT) structure (2.1). The current structure was designed to place the description of the MIR upfront, but since its structure is tailored somewhat to fit the catchment characteristics it may make more sense for it to be reversed. Again, we need to improve the description of the CRAFT model as all three reviewers have felt that it lacked enough detail as it stands. Perhaps we focussed too much on the observed data (Sec 2.2) and we will add new text. It is also important to stress that much development of the MIR concept was carried outlined in other papers where earlier versions of this model has previously been published. We also will attempt to explain why using the CRAFT model over more complex tools will add value to studies of nutrient management and mitigation (in response to R2 and R3).

- We showed two observed datasets collected at the mouth of the Frome, one sub-daily one weekly. The use of the sub-daily data was intended to highlight that (particularly for P) there is a lot of “noise” and non-behavioural patterns observed in the time series data that cannot be modelled using any process-based tool or spatially distributed model. The reviewers (R1 and R2) have not commented on this part of the paper. R3 however has made the point that showing “bad” results is sometimes necessary even if these results (when assessed using standard curve fitting methods) would fail most criteria except the ability to reproduce the mean concentration. These data can also reveal a lot of information about what is happening at the meso-scale, e.g. uncontrolled and random spikes in nutrient concentrations due to local farming activities or wastewater treatment processes. Regarding R2’s comments about model errors and not being able to reproduce the observed spikes in concentrations (Cs), it is important to note that in terms of loads (Q x C) these errors are quite small unless the spikes are associated with storm events with high Qs. As we have stated above this is not a paper about getting a ‘good fit’ it is about a model appropriate to this scale and this type of data.

- A more detailed discussion of in-stream processes, particularly relating to P, identifiable from the hi-res dataset can be found in Bowes et al. (2009b). Plotting C vs. Q did not highlight any strong correlation between P (in both SRP, TP and PP forms) and Q that would be expected if runoff events were a major driver of elevated Cs. The Frome catchment itself is dominated by subsurface flow so modelling flow at a daily time step is unlikely to miss many sub-daily storm events. The processes may be there but they are not observable at this scale.

- At least two of the reviewers (R1 & R2) have criticised the model performance in terms of curve fitting (e.g. producing high values of evaluation metrics such as R^2 or Nash & Sutcliffe efficiency (NSE)) C data although R3 thought that it was acceptable in the context of “bad” data (discussed above). We need to stress once again that the point of the modelling exercise has been intended to show how critical processes can be represented by the CRAFT MIR model as stated above. We have shown that poor fit (e.g. in a scatter plot) due to the timing of our recession period but the overall pattern is still visually acceptable.

- A mitigation “scenario” was then used in the paper to demonstrate how the observed loads can be reduced. R3 commented that we did not reduce nitrate loads, however this is not true as the nitrate loads from the SS component were reduced (See Fig 7 for a clear depiction of the decrease in loads). R3 has commented that the “relationship between model outputs and practical management options is not obvious”. We would argue that the model needs to be used in conjunction with terrain analysis to identify “hotspots” of pollution generation in the catchment, where management actions could reduce loads as shown by the model results. We agree that this step is probably missing in the current paper and future work will attempt to link the different components together holistically.

We will address the more detailed and specific points when we have had a reply from the editor.

Our Response to Reviewers 1-3 (in Turn)

Please note that we have used section numbers in the Revision in each response. We do not have the final typeset page numbers at this point in time.

Response to Reviewer 1

We thank this Reviewer (R1) for their constructive comments and have revised the manuscript according to some of the suggestions that they have made (i.e. the "Revision"). They have not made any itemised comments. Our responses to their general comments are below.

C4627 "*The inability of the model to predict variation in nutrient concentrations as shown in Figure 5 is especially troubling.*": We accept that the CRAFT's predictions of concentrations (especially TP) as were shown in Fig. 5 do indicate that there were under (especially TP) and over prediction at times. In the Revision we have (a) added an upper pane to the four timeseries plots of modelled and observed flow and *C* to show the error (observed-modelled) at each data point or timestep (Fig 4); (b) removed the original Fig. 5, since we feel that this was not particularly informative or supportive of the modelling exercise. Showing more detail (i.e. the errors) in Fig. 5 was the best way of improving the presentation of the results.

In any case, we are not aware of any published results from models that we have reviewed (e.g. physically based ones such as SWAT and INCA), i.e. those showing <u>daily</u> timeseries of <u>concentration</u> (*C*) data, that can improve on these based on previous studies using these models. These studies usually assessed model performance by comparing modelled load predictions with observed annual or seasonal loads (in any case the observed "loads" themselves were usually subject to significant errors unless high-frequency nutrient *C* data were available).

C4627 (1st para): "*The model does not formally consider any of the myriad biogeochemical processes such as denitrification and adsorption that are well known to greatly affect the transport of these nutrients.*" The MIR modelling procedure has been discussed in detail in earlier papers on the TOPCAT-NP MIR model (from which CRAFT was derived). These papers (Quinn et al. 1999; Quinn 2004; Quinn et al., 2008) have been cited in the literature and we have amended the Revision to stress the process of generating a MIR model (particularly the "mimicking" of results from physically based nutrient models such as EPIC and SWAT with a simpler MIR), in Section 2.2.1. It was found subsequently that end-users had difficulty in selecting values for some of the physically-based parameters retained in TOPCAT (e.g. the P absorption coefficient). Based partly on this previous experience we do not feel that additional "testing" of more complicated biogeochemical model processes is required in this particular instance, although we would welcome enhancements to the model to simulate these if the observed data were to become available from a different location(s) in the catchment. This has been further explained in Section 2.2.3 following the description of the nutrient processes in the model.

C4627 (2nd para) Additional data would also be required to parameterise variable *C*s in the model stores, in response to R1's suggestion of incorporating this. The comments above (1st para C4627) apply here also.

C4627 (2nd para) We do not feel that the use of additional data on "*velocity, temperature and stream width or depth*" would improve the modelling of the Frome and in any case such data were not available for this study. The comments above (1st para C4627) apply here also.

C4628 (1st Para) "*Representation of the riparian area, even as a simple additional box in the model seems necessary* " Regarding their comments on the representation of processes in the model, particularly the absence of a "Riparian Zone", we feel that R1 may have not fully agree with the MIR model simplification. In a riparian zone sources and sinks could include: releases of SRP from bed sediments; uptake of N&P by aquatic vegetation and denitrification (removing N from the water column). However, we argue that the inclusion of these processes in our model is not required and have attempted to show this in the Revision. This has been further explained in Section 2.2.3 following the description of the nutrient processes in the model. The fact that what is essentially a hillslope model (as pointed out by R1, p C4627) is able to reproduce a time series of concentrations reasonably well at the outlet is significant finding of the study, and presumably can be partly explained by the mixing processes occurring down the main river between the hillslope and outlet. It could be demonstrated that, whilst there may be sources of sinks of both N and P in the riparian zone these will have little net effect at the catchment outlet as they will tend to cancel each other out. It is at the outlet where the (sub-daily (HFD) and weekly (LTD)) observations of water quality are available in the Frome catchment. Limited nutrient concentration data have previously been collected at a monthly interval at upstream locations in the catchment. Additional data collected at various upstream points on the main channel reaches (a weekly interval may be sufficient during low –flow periods but not runoff events) would be required to parameterise a riparian zone model, although the processes could be described as "simple" (if they are constant fluxes or first-order kinetic equations), they may create problems of equifinality given the introduction of additional model parameters to the calibration process. For an overview of existing datasets and previous modelling/monitoring studies in the catchment with an interpretation of physio-chemical processes taking place in the River Frome, R1 is referred to these publications cited in our paper (Bowes et al, 2005,2009b,2011; Casey (1993), Smith et al. (2010)).

C4628 (2nd para): "*However, I am not convinced that these simulations provide much additional insight than could be gained through application of a simple export coefficient approach*" We feel that to compare our model predictions to an "export coefficient approach" neglects the major limitations of export coefficients, principally that they have to be derived using observed flow (Q) and C data (again often subject to limited C data). This makes the approach (a) unsuitable for modelling the effects of changes to flow pathways for contaminant transfer, aimed at mitigating high concentrations (as we have done with the Management Intervention Scenario) since they require observed data in the first place; (b) the export coefficients are usually only able to provide annual or at best seasonal load estimates. The data shown in Fig 4 indicate that, on an inter-annual and seasonal basis in this catchment, N&P concentrations in the Frome were highly variable and were reproduced fairly well by the model. This figure also shows that the model predicted the patterns of seasonality in the nutrient C data if not the absolute values, hence the errors in the scatter plots in Fig 5 (in the original version of the paper) and the upper panes in the revised Fig. 4, showing model error (in the Revision).

Response to Reviewer 2

We thank this Reviewer (R2) for their constructive comments and have revised the manuscript taking into account some of the suggestions. This is subsequently referred to as the "Revision". We note that they are generally supportive of the CRAFT model and its findings and address their technical suggestions in turn. The numbers below refer to the points numbered on C4776-:

1. Agreed
2. Agreed
3. Agreed
4. We accept the comments that the Methods section of the paper could be improved and have attempted to do so. We also checked that all Equations are in the correct order and that all the terms are defined correctly on first appearance in the Revision. A few minor changes were made as a result. These should be clear to the Editor when reading the Revision (Sections 2.2.2 and 2.2.3).
5. There are several criticisms of the paper here. Firstly regarding "*For example, it is hard to see how the points match in Figure 4*" , we addressed the model errors in predicting nutrient concentration (*C*) data in our response to Reviewer 1. We felt that the model was able to reproduce the seasonal patterns observed in nutrient *C*s reasonably well when viewed as time series plots (Fig. 4), however for reasons not totally clear this did not translate to a particularly good scatter fit in Fig. 5 except for SRP, and Fig. 5 has been subsequently removed from the Revision (as discussed above).

   (#5. 1st para) "*In addition (if I understand correctly), there is no model Validation*". Validation was not seen as necessary in this modelling exercise as we decided to use the longest time series of flow and concentration data available to us to investigate seasonal and inter-annual trends in the data without splitting the record up. We did not feel that a detailed validation exercise would add anything to the study or this paper.

   (#5. 2nd para) "*For example, how realistic is it that flow from WWTPs is proportional to groundwater flow (73-13)*?" A forerunner of CRAFT (TOPCAT-NP model; Quinn et al., 2008) generated flow (*Q*) from deep groundwater and WWTPs together, and combined these into a single constant term (*viz*. QGW in the CRAFT) with a constant nutrient C (referred to as QBACK and CBACK respectively). This is described in Section 2.2.1. In the Frome dataset it was not possible to distinguish between these sources during dry weather, given the available weekly monitoring data at the outlet (with no "observed" data on WWTP *Q* or *C* values) so these two fluxes are combined into a single source term (referred to as "DG" in Fig 1) which varies according to Eq. 7. This is an enhancement of the original model's constant QBACK. We appreciate that this may be seen as a limitation of our model, however in terms of mitigating SRP loads it is convenient to reduce the concentration of SRP (CGW) in the Management Intervention Scenario to represent a reduction in SRP loads from either or both agriculture and WWTP sources as they are lumped into a single term (QGW) at the outlet of the catchment.
6. C4778*"…there is some additional information missing".* Here we feel that R2 has misunderstood how the model is formulated. In our opinion the Discussion paper adequately described how the catchment was modelled (i.e. in Excel; as stated in Section 1.2 (2.2 in the Revision)) as a single, lumped set of processes (i.e. "buckets") . A "*discretization scheme*", i.e. numerical scheme usually requiring iteration

to solve differential equations, was not used here (equations are solved in sequence at each timestep). Fluxes such as PERC are derived using the storage value (SRZ) at the previous timestep, after which a water balance updates SRZ (Eqs. 1-9). In response to the second half of point #6, we have checked that relevant sections (2.2.2 and 2.2.3) are clearer on how Eqs. 1-9 are used and for any sequencing mistakes in their order.

7. "*The paper would improve if the authors add a paragraph on why it was necessary to develop this model*". We have added several sentences to the Methods (last part of 2.2.3), Results (3.4) and Discussion (4, first para) to make our "selling point(s)" for the CRAFT model stronger. The model is intended to eventually become a web-based tool for assessing mitigation options at a meso-scale catchment scale. We appreciate that there are many existing water balance/hydrological models out there that are capable of producing similar results, at least in terms of flows, but not in terms of nutrient concentrations.
8. "*Does the title clearly reflect the contents of the paper?"* As we pointed out in our response to R1 the MIR framework on which CRAFT is based was developed and tested through a series of models (TOPCAT) and published some years ago. We have added a section describing more fully how the MIR development process works in the Introduction (Section 1). We have also added more detail to Section 2 (Methods)."*I think it would be a good idea to include "CRAFT" in the title*". We have included "CRAFT" in the title of the Revision as suggested, which is now "*CRAFT (Catchment Runoff Attenuation Flux Tool), a meso-scale nutrient pollution model that uses a Minimum Information Requirement (MIR) approach*."
9. C4779: *Abstract* Thank you for your comments regarding the Abstract, we have actioned these.
10. "*Is the overall presentation well structured and clear*" We have improved the presentation of the results in the Figures as suggested by R2. We have added the cultivated layer that was missing (from the lower panel). We have addressed R2's comments regarding Section 2 (Methods) in #4 and #8 above and have improved this section. We have made some deletions to the Introduction as suggested by R2, partly to offset additional information on the MIR model development and theory in Section 2.2.1. We have expanded the Figure and Table captions where appropriate to improve their clarity. We have replaced Fig. 5 (a scatter plot of modelled vs. observed data) with the additional upper panes in Fig. 5 (formerly Fig. 4) showing a timeseries of the modelled error, as this method of depicting the results helps to explain model performance more clearly (This was also in response to R1, see above).
11. Thanks
12. We are not aware of any missing abbreviations in the table that R2 is referring to (Ed to comment?). Eq. 1 contained a typo as the term "QCSR" refers to a runoff process that has subsequently been removed from the final version of the CRAFT, and this has been expunged in the Revision.
13. We have acted on this comments above
14. We would prefer not to increase the length of the Revision by reviewing additional models but have taken into account the comment made by R2 and also the Editor in this instance. We have added a citation and brief discussion on the model studies of Van Der Velde (2010) and others (suggested by the Ed.). From reading these papers it appears that their scopes are not obviously directly comparable to our study for several reasons: (a) the catchments under investigation (e.g. Huppel Brook, in Van Der Velde's study) are small scale (e.g. 6.6 km^2) not meso-scale (b) only one nutrient was investigated (nitrate), apparently (c) with only one flow pathway (shallow

groundwater discharging into a dense surface water drainage network); (d) these papers in general (they appear to have all cited one another) focussed on the themes of travel time distribution and source particle tracking, which while important to the discipline, is completely unrelated to the aims of our modelling study with the CRAFT and its application to the Frome case study.

15. Our aim is to make the Excel spreadsheet containing a simple version of the CRAFT available on the internet for use in evaluating catchment management scenarios, in the near future, when practicable.

Response to Specific Comments (C4780-1)

"*This model requires actual evapotranspiration data*" and "*The water balance for the Frome catchment does not close*". The observed P and Q annual totals in Sec. 2.1.1 were obtained from the UK's annual hydrometric database (Marsh and Hannaford, 2009) for the entire period of the record (1965-2005), not from summing up the daily observed (forcing) data used in the model. This publication does not contain PET estimates for the catchment, and these have been derived from regional average values (as explained in 2.1.1). The model is not particularly sensitive to PET input data at least in this case study probably because: (a) PET is low compared to rainfall in the UK; (b) most runoff is generated by elevated groundwater flow (DGW) in the winter months corresponding to recharge periods. "*Can you justify that there is no evapotranspiration reduction*". Yes because introducing such a reduction factor essentially adds another (calibrated) parameter to the model that mostly compensates for errors in the observed forcing data, i.e. over or underestimates of runoff. *"Can you explain where this water goes*?" There was a difference between initial and final storages in the model over the simulation period of the order of +- 100mm when the three stores were summed together. "*Is the lower boundary well sealed*", yes, there is no leakage term in this version of the CRAFT model as the schematic diagram (Fig. 4 clearly shows).

75-6: "*Why did you choose to keep SPLIT fixed"* Introducing additional variability into (fixed) parameters without any evidence (e.g. soil moisture or recharge data) may cause problems during the calibration procedure, such as equifinality, that could prevent convergence towards an optimal SPLIT value. In general, the model should be first run with SPLIT=0.5 to apportion the recharge fluxes evenly between the DG and SS stores.

76-14: We will rewrite Eq. 10 as it should read:

$$\mathrm{COF}\,(N) = \mathrm{MAX}\,(K(N) \cdot \mathrm{QOF}, \mathrm{COFMIN}\,(N)) \qquad (10)$$

The use of COFMIN (N) thus prevents the calculated value of COF(N) falling below this value, as at low flows given the linear relationship (termed a "rating curve") it is possible that this could happen as the intercept is zero. This was a typo which we appreciate R2 spotting.

79-4: Regarding the point made by R2 about the 3-5 raingauges in the catchment, we did carry out trial simulations using data from an alternative gauge (in reality only 2 gauges had reasonably complete daily records covering the modelling period), but the results were not particularly sensitive to the choice of gauge. As with the PET/AET data these issues were not particularly interesting or relevant to this particular case study. This is reported in Section 2.1.1.

80-10: Regarding the correlation between flow data (both daily (DMF) and instantaneous values). The DMF time series contains the daily averages of the "instantaneous" flows, which were measured at 15 minute intervals. We calculated the coefficient between the instantaneous flows recorded at 12pm and the DMFs for the same day. This is discussed in Section 2.1.2.

80-18: *"observed graphically"* The statistics shown in Table 1 indicate that means and σs of the LTD and HFD nitrogen time series are very similar, adding this to the sentence formerly referenced at 80-18 should explain this observation.

85: Choice of parameter values. Referring to the choice of a value for QUICK (final para on P C4781), it is implied that a user requires some "*expert*" knowledge of the amount of surface runoff generated in the headwaters of the meso-scale catchment by the "OF" flow pathway (see Fig. 4), and how this may be reduced volumetrically by Management intervention Scenarios. We suggest that the user considers how this may be achieved in their catchment, and then by what percentage (e.g. 25%) QUICK would then have to be reduced to represent the improvement. The length of the surface flow pathway will not directly affect this value unless there is some re-infiltration or storage (leading to evaporation). A further development of the model may incorporate an attenuation term to add a delay to the OF pathway to represent the effect of mitigation features at attenuating water in the headwater catchment (the "A" in CRAFT), referred to in the Discussion. Note also (Ed.) that two parameters have been renamed in the Revision, QUICK to KSURF and KSURF (original usage) to SDMAX.

Response to Technical Corrections (C4782)

We welcome these corrections and the Revision has incorporated most of these (where we feel this is an improvement) as well as the minor changes referred to above. The suggestions listed below have been actioned on unless stated otherwise.

67-23: "will show": you don't know that yet in the introduction.

We have changed the tense here (to "aims to show") to address this comment (P3 Para 1).

69-6: Itemizing these research questions draws the attention, which is good. However, these are not the main research questions of the paper (If I understand correctly), so this attention may not be helpful here.

We agree that this was not the correct place in the paper to ask these questions and the text have been removed. Some key questions are asked in section 1.1. under a relevant heading of "The MIR Modelling Methodology".

75-7: Write out both equations for subsurface flow. Introducing Qsub just to combine equations is a little confusing.

We don't feel this to be necessary. (Question for Editor?)

75-25: You could split Sec. 2.1 into 2.1.1. Water fluxes and 2.1.2 Nutrients.

This has been carried out in the Revision. These sections are now numbered 2.2.2. and 2.2.3.respectively following a discussion on the MIR theory (extended into Section 2.2.1).

77-6: Remove the outer brackets in the numerator.

We have added a note to the Typesetters to make this change as the brackets were added in the process of preparing the HESSD typeset manuscript not by ourselves.

77-23: Refer to Section 2.2.2 where you explain what kind of samples you took.
We did not take any samples, hopefully the ordering in the Revision makes it clear how the two "observed" water quality datasets were collected and when?

*78-27 Replace the word "known" with e.g. "previously measured", as you neven (*sic*) really know the ET.* It is "estimated" PET and has been described as such in the Revision.

83-13: Move Eq. 14 after "simulation i"

91-14 "as a" ! "for"

91-18 "scenarios" ! "scenario"

93-4: "uncertainty" ! "uncertainty analysis"

Table 1: Why did you include both mean and median? They give the same information here.

We originally included the means and medians of each concentration in Table 1, in case the time series of the different nutrients exhibited a log-normal distribution (*in which case the mean and median may differ significantly).* We have therefore removed the medians from Table 1, and stated that the two measures are very similar, as R2 has pointed this out.

Figure 2: Adding catchment boundaries and/or topographic map might be useful.

Regarding Fig. 2 (now Fig. 1) we do not have either a DEM or digitized catchment boundaries available in order to add the catchment boundary. One of the points we are striving to make is that the CRAFT only needs timeseries of rainfall, PET and observed flow (and nutrients) to run. We suggest that the reader (e.g. R2) uses Google Earth or similar tools to determine the attributes and elevation range of the catchment, based on the information shown in Fig. 2, which includes the lat-long grid coordinates of the catchment.

Response to Reviewer 3

Specific comments

Abstract line 1: “water pathways” instead of “runoff pathways”?

The term “runoff pathways” is used extensively in the hydrological modelling community and we wish to retain it?

P 10367 lines 19-22 are these questions answered in the manuscript?

These questions have been removed, key questions are asked in Section 1.1: “The MIR Modelling Methodology” which we feel are answered later in the Results and Discussion.

What does “more transparent” mean? More transparent on the model assumption, on the model running process for non-modeller, on the uncertainty on the model outputs?

We have removed this particular sentence as it was clearly ambiguous.

P 10367 line 24 “this model”: which one?

We have clarified that “this model” refers to a “new model”, described subsequently in the Revision.

P 10637 line 25 Are we sure that the presented model reflect the dominant processes at the end?

It is not clear what is meant by the “end”, if R3 is referring to the catchment outlet then we are confident that our model has represented the signals that are observable in the monitoring data (time series of flow and concentration i.e. *C*). If it is the end of the paper then we have addressed these issues in the Discussion and moved some of the statements from the Introduction that the Reviewers were not happy with.

The authors assume that main drivers are hydrological, this is actually their assumption and should be explicitly written. Soil and hillslope processes describe indeed dominant flow pathways in humid temperate climates: would these assumptions be the same in an arid Mediterranean catchment or in a North American snowmelt driven catchment or in specific geological conditions such as in karst system?

Here we have added a caveat that limits the applicability of this particular model (as described for the Frome case study) to temperate catchments (in Section 2.2). Additional processes could be incorporated into the CRAFT framework (e.g. snowmelt) if the data are there to support their inclusion.

P 10368 line 6 “Aggregation and homogenisation” do not mean that there is no geochemical process. Once again here the authors make an assumption which is that at their scale of interest there is no effect, no control of the spatial variability of land use, and management practices on the nutrient fluxes. This hypothesis should be discussed at the end of the manuscript when interpreting the periods or scenarios the model fails to reproduce.

The text in the Introduction has been modified slightly to state that we are using a parsimonious, lumped mixing model. Therefore, the CRAFT model does not require spatial patterns of input data (particularly of nutrient fluxes) to simulate nutrient. Later in the Revision we state clearly that these processes are not observable at the outlet of the meso-scale catchment, but their effects are being captured in the model parameters (in Section 2.1.2).

P 10368 lines 18-22 Looking at the meso-scale, the use of a hillslope model is quite unexpected while the riparian and in stream processes are expected to be taken into account as their relative contribution to Nutrient dynamics is known to increase with the order of the catchment. So far as I understand, the original assumption of the authors is that even at meso-scale catchment, hydrological flow paths are sufficient to represent both the seasonal and the event dynamics of nutrients. But it is not clearly explained and quite in contradiction with what the reader would expect at this point therefore it should be more explicit.

This point is related to the above point, we stressed which processes were no longer being represented in the model. Their effects however are still being captured in a mixing model where the hillslope and channel are in fact assumed to be integrated (Section 2.2).

P 10369 line 6 this is a very relevant question but the limit of management interventions which can be designed using such tools have to be precisely defined.

We feel that they are "precisely defined", a future version of the tool(s) on the internet will have clearly-stated limitations and caveats on the use of the CRAFT model to investigate different scenarios which basically require the end-user to think about the values of the parameters being used and how these should be changed given high uncertainty. The CRAFT is developed in such a way to highlight the sensitivity of the parameters not mask it.

Regarding the assumption that the spatial variability is not required in the model the relationship between model outputs and practical management actions is not obvious.

We have improved our description of the MI Scenario (Section 3.4) in terms of how "practical management actions" at the local scale can have an impact on water quality outputs at the modelled scale.

P 10369 line 8: At this stage, we would expect a proper comparison of the MIR approach with and without taking into account geochemical transformations.

Based partly on this previous experience we do not feel that additional "testing" of more complicated biogeochemical model processes is required in this particular instance, although we would welcome enhancements to the model to simulate these if the observed data were to become available from a different location(s) in the catchment. This has been further explained in Section 2.2.3 following the description of the nutrient processes in the model.

*P 10369 line 26 please define **DSS**-based model?*

The abbreviation "**DSS**" has been expanded in the text in the Revision.

P 10373 The use of a Minimal Information Approach is not clear in the choice of the model which seems to be already chosen at the beginning of the study. Some explanations about the conceptual choice would be nice, e.g. what is the link between WWTP and deep groundwater flow? In many cases, WWTP release the water directly into the hydrographic network and would be associated with rapid surface flows in this case.

We have explained a simplification of the processes in the CRAFT in our response to Reviewer 2. The "signal" of deep groundwater inflows (i.e. "baseflow") is not observable at low flows if the WWTP releases also contribute a significant proportion of the total (and base) flow. What is observed at the outlet is a mixing of the two sources during these periods of low flow.

Is this conceptual model similar in practice with other models? It seems that the partition of Overland flow vs Base flow is similar to what is described in classical physically based models such as Topmodel for instance, with a second step of partitioning within base flow a subsurface and a deep flow...

We have expanded our description of the model development. Most "lumped" hydrological models contain some form of storage-discharge relationship with multiple stores representing overland flow, baseflow etc. A forerunner of the CRAFT, called TOPCAT was a MIR model developed from Topmodel and retained some of its components, however the relationship between topographical index and overland flow (i.e. spatial component) was not retained in the MIR structure (See section 2.2.1). These models are cited in the literature.

Eq. 1: QCSR is not defined, is it?

The term has been removed as it is no longer included in this version of the CRAFT and had been erroneously left in Eq. 1.

Figure 1 the conceptual scheme is not clear: e.g. it seems that nutrient outputs only come from DG.

The offending arrows have been removed, deep groundwater flow was not what was intended to be depicted by them in Fig. 1 (now Fig. 4 in the Revision).

P 10376 lines 4-7: What would be the implications of using the uptake factor if flow is not well simulated? Could this explain the need of a minimal concentration threshold?

Any errors in flow (i.e. Obs-Mod; *ERR(Q))* will be directly translated into errors in concentration (above the value COFMIN) by a linear function (Eq. 10) (*ERR(C)*), which will further add to the load error (i.e. increase it by the absolute value of *ERR(Q) x ERR(C)*). Setting a minimum concentration is required because if QOF (overland flow) is relatively small, then the corresponding COF calculated would be below what is normally detectable in samples taken from the rivers in the catchment, due to limitations in analytical techniques and the near impossibility of observing a zero concentration. In the Frome catchment it appears (both from the shape of the observed flow hydrograph/ flow duration curve and previous studies (Marsh & Hannaford, 2008) that estimated the baseflow index (BFI) to be over 0.7) that overland flow is relatively uncommon, so that errors in predicting COF are not significant in terms of the annual nutrient loads of TP.

Case study description: maybe this section should come before the model description so that it would be easier to understand some choices in the model structure.

We have swapped over Sec 2.2 (Case Study Description) with 2.1 (Model Description) as suggested by R3.

P 10379 line 27 Table 2 is cited after table 1.

The above reordering has meant that Tables 1 and 2 are now cited in the correct order, if this is what R3 was referring to here.

P 10380 lines 12-14: What the analysis suggests is not so obvious for the reader: some information about the mean duration of the flood would help to justify the choice of the daily time step.

We have stated that there is little statistical difference between the timeseries of daily mean flows vs. instantaneous flows in Section 2.1.2. There was no evidence from either timeseries that the travel time of a flood peak to the gauge at the Frome outlet was either significantly less than one day (i.e. a sub-daily timestep was required) or more (i.e. where a routing function would be required, in order to lag runoff and nutrient loads from the catchment to the outlet). Therefore, a daily timestep was appropriate.

P 10384 lines 3-6 How far have the hydrological parameter been modified? What is used for the depicted results? The use of Nitrate as nutrient associated to deep flow paths and Phosphorus, in particular PP associated rather to surface flow paths could be improved in order to constraint the model, even the hydrological parts if all the parameters were calibrated simultaneously or using maybe specific criteria to assess the performance on base flow and on peak flow associated to storm events.

This is an interesting point but outside the scope of what can be identified from the weekly water quality data (at least in terms of the partitioning of TP into PP and other sources). It is apparent in Fig. 5 (Fig. 4 in the Revision) that the model can predict the *range* of elevated TP concentrations (due to PP) observed during flood events, however very few such events have been sampled in the LTD to verify this. In terms of nitrate, the SPLIT parameter was adjusted slightly to obtain a better fit to the baseflow nitrate concentration as it adjusts the proportion of flow from SS and DG (baseflow) stores. In Fig.6 the envelope of modelled flows has an upper bound showing Q95 (95^{th} percentile flows on each day). Correspondingly higher PP concentrations are predicted by these flows (not shown graphically but included in Table 4). A higher frequency of sampling would be necessary to permit using the nutrient data in the manner suggested to calibrate the model, and we are exploring this in ongoing research in a different catchment with hourly nutrient data.

Tables 2 and 3: Would it be possible to describe the parameters? It is not clear, especially for the "nutrient" parameters, which one are calibrated and which one are estimated from the text, maybe the table would be the right place to precise (sic).

In our opinion the Methods section (2.2.2 & 2.2.3) is the best place to describe the parameters through the use of equations. We do not feel that there is sufficient space in the Tables to do this. In the Revision we have improved the text to better explain how the calibration procedure was carried out.

P 10384-10385 Management Interventions. At the end the model seems more dedicated to help the end-users understanding how their catchment behaves. The concern of policy makers is generally not only to know that nutrient inputs have to be decreased, but rather how, when, and where? If they are, how much will be the cost for farming production or the benefit for other services?

We have added 2 references to the Revision : (i) cited a report by Cuttle et al. (2007) that refers to farm uptake of farm scale policy directives; (ii) and a recent report by the UK Government (DEFRA, 2015) that have discussed catchment scale approaches, and we have discussed these in the Introduction, Sections 1.2. , 3.4, and the Conclusions (4). In any case the end-users should be interested in how their catchment behaves at the local scale since they have a responsibility under the WFD and other directives to prevent contamination of surface and groundwater bodies in their area. Since water quality standards are set using concentrations not input loads a modelling tool can be used to estimate the changes to input loads required to attain these standards (as discussed in Sec. 3.4.).

P 10386 lines 12-14 So despite that the MIR procedure would lead to consider only base flow, the knowledge on P transfer lead to keep the additional overland flow component. This illustrates my comment on the use of P and N concentrations as additional variables to constraint the model. Moreover it highlights that something is missing either in the calibration *criteria to give some weight to parameter sets able to represent simultaneously flow, NO3 and P dynamics or in the model itself.*

This is a good suggestion as the MIR procedure attempts to retain any processes observed in the data (e.g. spikes in TP caused by PP transport via overland flow). However, the weekly sampling rate is just not good enough to be sued. We aimed to develop a model that could simulate the overland flow generation of PP, and thus enable catchment managers (through a MIR approach) to address reducing the concentration of PP in overland flow to reduce the PP and TP loads. The Frome is dominated by baseflow so obtaining a "good fit" to the flow data was possible by focussing the calibration on those parameters that control baseflow in the model, i.e. SPLIT and KGW / KSS.

P 10387 lines 2-3 miss and over prediction are attributed to the limitations of weather data, could this be related to missing processes in the model?!

In terms of runoff this was unlikely as we have reproduced the observed volume within 5% which is less than the generally accepted error in catchment rainfall of 10%. In terms of TP C we have admitted that there may be a soluble source of P (other than the insoluble PP) that could explain the difference between the observed SRP and TP concentrations during low flows, that the model is currently unable to reproduce, however Bowes et al. (2005) by performing a mass balance analysis on the river, determined that SRP comprised 90-95% of the total dissolved P load over a 14 month period (during the same time period that was modelled here).

P 10387 lines 19 Again there is a little lack of clarity about which parameters are calibrated and which are estimated from previous knowledge, e.g. How is determined the contribution of WWTP?

Unfortunately we only had an estimate of the DWF (Dry Weather Flow) from the major WWTPs in the catchment and no measured effluent concentrations. Earlier studies cited in the study (e.g. Casey et al., 1993 for nitrate and Bowes et al. (2005, 2009b, 2011) for SRP) have attempted to quantify the WWTP loads, in the case of SRP by a model (LAM) that estimated the contribution from point and diffuse sources in the catchment. The reviewer is encouraged to read these papers (Bowes et al., 2009b 2011). We have expanded the results section to indicate the proportion of the SRP load that is expected to have originated from the WWTP (during dry weather) based on this modelling and the HFD in Section 2.2.1 and Fig 3. During wet weather there may have been overflows (e.g. from leaky septic tanks or farm slurry tanks) leading to elevated SRP concentrations and loads, but data on these (locations, loads etc.) were not available (as pointed out by Bowes et al. (2005)).

P 10387 line 27 what is the impact of errors on flow simulation on the simulated concentrations?

A sensitivity analysis was performed (as described in the paper in Section 2.3) using Monte Carlo simulation to randomly sample, from wide bounds, those parameters controlling runoff generation and predict nutrient *C*s with each set (nutrient parameters were not adjusted). Further analysis subsequently found that the nutrients (as assessed using the root-mean square error, RMSE between modelled and observed *C*s) were not sensitive to the model's performance at predicting flows. It was also found that there was no systematic relationship between the error in predicated runoff and the error in predicted nutrient concentration that would indicate a major structural problem with the model. The errors in predicting flow and concentrations are now shown in Fig 4 (upper panels of each timeseries plot). Results also indicated that the model's performance at simulating SRP and nitrate with the "Expert"ly calibrated values of the runoff parameters was in the upper 25th percentile of the RMSE performance metrics from all the Monte-Carlo simulations. Results for TP were not as encouraging probably due to the underprediction already discussed. We also discovered that the "best" runoff parameter values for simulating SRP (i.e. with the lowest RMSE) were not necessarily also the "best" values for simulating TP.

Is there any transformation processes which could explain part of the error (denitrification, uptake)?
These were not observable at the outlet scale (see above point) and in fact nitrate concentrations do not exhibit a noticeable drop during summer months that could be attributable to denitrification in the River Frome. Refer to the timeseries of nitrate *C* shown in Figs. 2b and 5 (in Revision), as these indicated that the lowest nitrate concentrations were usually observed around mid-autumn (i.e. October) each year. The highest nitrate concentrations were observed during a winter-spring period of elevated baseflow, which the model predicted, albeit with a modest timing error.

Could the use of daily time step explain some discrepancies for storm events simulation (which could last less than one day)?

We have discussed the typical duration of events in the Frome catchment above and also in response to Reviewer 2. These "discrepancies" (in terms of nutrient dynamics during storms) are unlikely to be observed in this catchment with a weekly timeseries collected at the outlet as events were missed. In any case, the sub-daily rainfall data that would be required for a smaller time step often has significant errors (sometimes from processing) and/or gaps (and was unavailable for this study). Fig 2a shows the DM and instantaneous flows and the reader can clearly see that there are no significant sub-daily runoff events that were not identifiable by using the DM flows.

*P 10388 line 14 idem (*sic?) *impact of adsorption-desorption processes?*
We discussed the role of nutrient cycling in terms of soil and in-channel processing in the modelling Section (2.3.)

How well are peak flows reproduced?
In the Discussion paper, a scatter plot (Fig. 5) of observed vs. modelled flows indicated that peak flows were generally predicted quite well with some overprediction by the model. We have revised the figures in order to show the error (Obs-Mod) as a timeseries instead of a scatter plot above each timeseries plot of *Q*s and *C*s. A separate analysis plotting Obs vs. Mod flows as a scatter plot (not shown for brevity) revealed the same pattern and the Nash-Sutcliffe efficiency metric was 0.8.

P 10389 line 3 The spikes which are observed but not modelled could be actual events C4902 that the hydrological model is not able to reproduce (e.g. local runoff on limited saturate areas, or low hortonian runoff in dry season?)

This is unlikely as there was no significant rainfall on the day(s) of these spikes to generate runoff through any mechanism (e.g. IE or SE). In any case "local runoff" in the Frome catchment will probably not be observable at the outlet gauge. In Fig. 2 and accompanying text the plausible mechanisms generating the spikes are explained clearly already (and these have been previously discussed in earlier studies: Bowes et al. (2005, 2009a)).

P 10389 line 9 Might the model be too simple to properly simulate the dynamics of Phosphorus?

The above point regarding a "missing" component of TP in the model and the results (at predicting TP concentrations) relates to this query. We did not set out to develop a physically-based P model that incorporates all the processes in the P (or indeed N) cycles.

In any case the model has simulated SRP reasonably well (Fig. 4, Table 3) with errors (based on mean and 90th percentile concentrations) of less than 10%.

P 10391 line 4, It is not obvious that it will be possible to catch incidental losses with a low sampling frequency.

We agree with R3 on this point, and have modified this sentence accordingly to clarify that the observed "noise" and related signals (in the HFD) "have a limited effect on the overall signal and loads". Identifying these signals is one benefit of having high-resolution monitoring data, albeit at significant extra cost to the monitoring agencies.

P 10391 line 11 Are channel processes really noise? The addition of in stream processes to hillslope processes and then mixing processes make indeed the deconvolution of the stream signal more and more difficult however these processes are real, and they may be relevant in terms of management strategies (actions to increase the natural cleaning capacity for instance)

There were no data available on hillslope (i.e. <1km^2 scale) runoff or nutrient loads to enable such a "*deconvolution of the stream signal"* to be modelled in the Frome catchment. We consider that it is appropriate to consider the signal observed at the outlet in the HFD nutrient data as containing some "noise" and this has been discussed in Bowes et al. (2009a), the first paper to publish the high-resolution data set. It should be stressed once more that we are demonstrating a meso-scale modelling study with a simplified model, not a hillslope scale study of an experimental catchment with highly detailed (in time and space) observation networks. The comments we made above on the scope of the modelling study apply here too. The concept of "cleaning" the channel is not clear to us as a "management option" and outside the scope of the study.

P 10392 lines 6-10 The authors did not integrate management aiming at reducing nitrate loads in their scenarios which is Ok even it could be clarified in the introduction, and however they implemented a scenario with nitrate concentration decrease in the subsurface flow. This leads to some ambiguities regarding the author's objectives.

R3 has not fully understood what is written in the Discussion paper describing the MI Scenario. Nitrate loads in the SS component were also reduced by lowering the CSS(NO3) parameter (see Section 3.4).

P 10392-10393 Discussion about the added value of the use of the CRAFT model comparing to others tools would be worth at this stage.

We agree with this point, also made by the other reviewers, and sentences to this end have been added to several sections (see above).