Dear Markus

Once again, many thanks for your comments and the detailed reviews from the referees. I am not sure if there is very much of the original paper left! However, I think the paper we are now submitting has benefited greatly from the review process. I hope you will find that the paper is much shorter, reflects the reviewers' comments and now has high quality, informative figures. We also took the advice of one of the reviewers in the last round and asked a colleague to review the paper, and this led to many improvements in the paper and a substantial shortening of the text.

You may find the markup up copy quite difficult to follow as we have changed and moved the text around. Our edits are in red and the response to reviewers are in turquoise. A full marked-up copy of the file can be found after our detailed response to the reviewers.

Below are our responses to your comments and the reviewers' comments which are shown in Red

"Editor Decision: Publish subject to minor revisions (Editor review) (03 Feb 2015) by Markus Hrachowitz

Comments to the Author:

Dear authors,

all reviewers agree that the revised version of your manuscript showed considerable improvements, which is in particular true for the justification/explanation of your approach. However, the reviewers also point out several points that still need some attention (see reviewer reports).

We agree that the review process is leading to a better paper and that more changes have been required.

In addition, I would encourage the authors to rethink/discuss some of the points below:

- Putting wastewater treatment plants and groundwater in the same box may reduce complexity of the model, but also makes it much less useful for water management scenarios.

In the end we decided to leave them together but did make it clear in the text what fraction of the P load arising from the largest WWTP. Apart from that the P and N loads have been estimated from an export model for the whole catchment.

- The fit between modelled and observed discharge is still hard to see (fitting a time series of 7 years into one window doesn't work).

The timeseries graphs (Fig 5) now zoom in to 2 years of data (whilst the model was still calibrated over the entire period). Two years was chosen to reflect different hydrological conditions, one average and one wet year.

- The bad fit between observed and modelled total phosphorus remains. The authors solved this by removing the scatter plot (fig. 5), which showed too much scatter, and refer to the time series plots. Besides the fact that "hiding" unsuitable results (even if it was not done on purpose) does not inspire much confidence, here again, time series of several years compressed into a small window already look well if you have the seasonal cycle right.

- The introduction is still too long and not focussed enough on the motivation for this study.

We have obtained a review from a colleague not involved with the work. The manuscript has been substantially shortened in the Revision. The Introduction in particular has been edited and we have added a clear statement of objectives at the end of the section.

- Several things were explained to the reviewers (for example choices on model complexity), but not in the paper itself. Readers may wonder the same things.

We have incorporated all relevant comments made by the reviewers but admit we did get involved in a debate in our last response. I think the reviewers have now responded to our comments and thus prompted the further changes have now made.

- Overall, the paper does not give the impression of really careful work. The lay-out of the figures could be improved and the text can be more structured and shortened. The parts that were added after the review are not embedded in the original text and little text was removed or shortened.

We have improved the presentation of the figures and removed all boxes. All fonts are the same size, although remember the type-setting process may modify the fonts in the final version of the manuscript. We have also closed up any empty space where practicable. We did have several errors that arose during our last edit. On this occasion we have carried out a very detailed edit.

Reviewer 3

Review of "CRAFT (Catchment Runoff Attenuation Flux Tool), a meso-scale nutrient pollution model that uses a Minimum Information Requirement (MIR) approach" by R. Adams, P.F. Quinn, and M.J. Bowes

The authors describe a model that simulates loads and concentrations of nutrients with a focus on nitrate, total phosphorus, and soluble reactive phosphorus in the River Frome. This is the second time I have reviewed this paper. My principal criticism of the first version of the paper was that model fits to nutrient concentrations were not very strong, which caused me to question the representation of processes and the landscape. I was not very familiar with the Minimum Information Requirement approach that the authors advocate and that is applied in this paper. In this revision, I note that they have done a much better job of providing some background as to how and why this MIR approach has developed. The authors have also added some information as to how this specific model (which they term "CRAFT") was developed. They describe the justification for how more complex process representation was rejected as unnecessary to achieve their aims. I appreciate this background and description of the logical process of model building as it better informs the reader and provides some confidence to the model user.

Although I still struggle a bit with a nutrient model that does not attempt to represent any of the biogeochemical processes that we know affect the concentrations and loads of these nutrients, I am willing to accept that a modeling approach such as this does capture key elements of variation in nutrient concentrations and loads as they are affected by variation in runoff processes. The authors have demonstrated through the Management Intervention Scenario that this model can provide insight to how broad catchment-scale strategies would be expected to affect loads and concentrations. I do question whether this model alone is enough to use as "the" management tool for control of nutrient loss, but the authors do describe that multiple models may be needed to implement appropriate management strategies.

This response reflects the longer debate we had in our responses to the reviewers' comments and our attempt to improve the paper. However this did mean that in the end the previous submission was not shortened! We are confident that we have now addressed the weaknesses this reviewer found in the original paper. Our new edit has focused on shortening the text and removing any repetition.

In short, the paper is much improved and the modeling strategy is more easily understood in this revised version. I do not have any remaining technical concerns with the paper at this point.

Many thanks, this does reflect how useful the review process can be.

Some editorial suggestions:

- Abstract, 2nd P – "Also" at beginning of sentence can be eliminated as it is redundant
- Abstract, 2nd P – sentence that begins with "A management scenario" is incomplete
- Page 2, 2nd P of Intro. – "simulations" should be singular
- Page 2, 2nd P of Intro. – "times" should be singular
- Page 5, 2nd P – move "than this one" directly after "domain"

- Page 11, 2nd P – add on between "based these"
- Page 13, 3rd P – eliminate from in this phrase "that from the HFD"
- Page 13, 3rd P – in the sentence that begins with "A similar analysis", eliminate the second instance of "therefore"
- Page 13, 4th P – change "suggested to that improved" to "suggested that an improved"
- Page 14, 2nd P – change this sentence: "The above discussion led to the following model structure for the CRAFT model being chosen, it representing a MIR representation of a more complex hydrological system." to "The development of the conceptual model discussed above led to an MIR structure for the CRAFT model that represents the complex hydrological system in the simplest manner feasible."
- Page 14, 2nd P – after (ii) and (iii), eliminate "as"
- Page 14, 4th P – change "accounting for controlling ET and the drainage control rate" to "controls ET and the drainage rate"
- Page 15, 1st P – change "reducing" to "reduce"
- Page 16, just above eqn. 11 – change "sin" to "in"
- Page 19, 1st P of Section 2.4 – part of the sentence that begins with "However" seems to be missing
- Page 19, 2nd P of Section 2.4 – in (i), "remove" should be plural
- Page 21, top line – eliminate "in terms"
- Page 21, 2nd P in Section 3.3.1 – change the "model appeared to model" to "the model appeared to simulate"
- Page 22, 1st P of Section 3.3.2 – the sentence that begins with "Visually" appears incomplete
- Page 22, 1st P of Section 3.3.2 – second instance of channel can be eliminated from this phrase "within-channel river channel"
- Page 24, 1st P – add as between the words "such riparian"
- Page 24, 1st P, 1st line – "scale" should be plural
- Page 24, 1st P, 2nd line – "tends" should be singular
- Page 24, 1st P, 4th line – change "in" to "at"
- Page 24, 2nd P, near bottom – change "as a optimizing" to "as a means to optimize"

We have made all the changes suggested by this referee in the Revision and highlighted these in blue in the marked-up copy.

Reviewer 2

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

Some tips:
- Ask a colleague who was not involved to critically evaluate the structure of the text and indicate parts that can be shortened.
- Improve the figures by zooming in to one year, so you can distinguish between the lines for modelled and observed time series. Also, make them look good. Increase the resolution, remove boxes, remove empty space, use the same font sizes, etc.
- Shorten the introduction and end the introduction with a very clear statement why it was necessary to develop this model.

Our Response:

1 We have obtained a review from a colleague not involved with the work. The manuscript has been substantially shortened in the Revision. The Introduction in particular has been edited and we have added a clearer state of objectives at the end of the section.

2 We have improved the presentation of the figures and removed all boxes. All Fonts are the same size, although the type-setting process may modify the fonts in the final version of the manuscript. We have also closed up any empty space where practicable

3 The timeseries graphs (Fig 5) now zoom in to 2 years of data (whilst the model was still calibrated over the entire period). Two years was chosen to reflect different hydrological conditions, one average and one wet year.

Reviewer 1

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

The manuscript has been widely improved in the structure, important information has been added about the model, and the scenarios have been more clearly described and interpreted. The addition of uncertainty estimates is much appreciated.

I still have a few comments and questions which are detailed below, key points are:
-Checking the symbols used for the model variables and parameters which are sometimes varying in the text itself and between the text and the conceptual schemes

We have done this in the Revision

-Moderating the conclusions about impacts of the tested scenario on nitrate load considering the processes not taken into account by the model.
-I am surprised that the model does not use input data, especially for nitrate which is OK in the currently tested scenario which focused on P load reduction, but it limits the range of scenarios for which can be tested using the model. So nitrate is more used here for constraining the deep flow path than to properly test any management strategies about nitrogen (as explained by the authors this would need simulation greater than a decade)

We have made the point that in Fig.7, we clearly show that nitrate loads reduced under the MI scenario. We reduced the nitrate concentrations in the SS component to achieve this. This is also indicated by the adjusted parameter values in Table 4 for nitrate simulation.

-Adding some discussion about the non-simulated peak events on Phosphorus (PP mainly): it is ok that the model does not aim at representing all the variability and will neglect some processes. I fully agree with the authors that a daily time step model won't be able to catch precisely all the variability of storm events, however since events are an important contributor of PP loads, as highlighted by the HFD time series, it is worth to discuss this point as a limitation of the model.

We have addressed this point, however please note that this is not a "limitation" of the model, CRAFT can simulate sub-daily runoff and nutrients if there are observed data available to do so. This is now quite clear in the paper.

Detailed comments:

We have taken most of these on board. We have now made some fairly major edits in the current Revision so some of these changes have been made in any case, hence there is no response below from us.

Abstract line 3: "AND which focusses"? (Suggestion)

Changed

Introduction

p.2, paragraph 1, line 4: Just to point out that if the meso-scale may be the relevant scale for policy making in the UK, and that it is a relevant scale regarding the physical system, it is not always the relevant scale in terms of economic and human systems.

Noted but no change has been made

p. 2, paragr 2, line 7: "complex model simulations are prone to high uncertainty" Ok but I am not sure that it is not the case as well for simpler models.

Changed to "However, these models tend to be too complex for informed end users to use and the simulations are prone to having greater parameter uncertainty than simpler models (McIntyre et al., 2005; Dean et al, 2009)."

i.e. we accept that all models have uncertainty.

p. 3 paragr 4, lines 8-10: I do not understand the last sentence of the introduction

Edited

Section 1.1.

p. 4, (1): Doesn't it depend on the management issue?

Exactly… you must adapt the model to the issue. Here we show an example of how to do this.

p. 4, (2): "how nutrients are lost", what do you mean? which pathway lost which nutrient?

Changed to “Can the MIR model simulate how nutrients are lost from the catchment through the dominant hydrological pathways”

p. 4 paragr. 1, line 5: (suggestion) “meso-scale diffuse MULTIPLE pollution”
Section 1. 3.

We have greatly reduced the length of this section and this text has been deleted.

Methods

p. 9, paragr. 2, line 5-6: weekly time series of nitrate concentrations have also been shown to be sufficient for load estimates. The sentences “The correlation between C and Q … Cassidy and Jordan (2011)” is confusing, it seems to refer to SRP concentration in the middle of the nitrate paragraph. At the end of this paragprah: I assume indeed that nitrate is more concentrated in groundwater flow than in overland flows but is it probable that sewage effluent are rich in nitrate as well?

This section has been heavily modified and reduced in length in the revision. The sentence referring to SRP was in the wrong paragraph as you have pointed out, thanks (it has now been removed). Unfortunately no data on the concentration of nitrate in sewage (WWTP) discharges were available.

p. 9, paragr. 3, lines: 3-4: It is frustrating not having weekly PP data, especially because you argue to use the HFD time series to select a daily time step as relevant, while PP is the most sensitive variable to the monitoring frequency. Also, when I looked at the Figure 2 when it is cited for the first time (p. 8) I was wondering why there were no LFD for the TP, maybe adding the precision of missing data p.8 in the paragr. (1) so that it is known before (just suggestion).

We explained that TP data were not available over this period to tie in with the figure. Regarding PP concentrations in the LTD dataset, unfortunately there was also a gap here in the record, because TDP data were not available either (to estimate PP as TP-TDP).

p. 9, paragr.(i) Type “3” events may be associated with decoupling in PP and SRP transfers… as they are used in the following, it may be worth to properly define “Type 1”, “2”…

This section has been greatly reduced and the remaining text changed and thus we have addressed this issue during the edit.

p. 9, paragr. (iii) However, type “5” events on Fig. 2 seems to be associated with small peak of Q.

That is possible, but was not the case here if you carefully trace a line between the two panels.

p. 12 reference to Figure 2: the model does not reproduce the background NO3 concentration but it does not reproduce any dynamic in base flow at all!! I am not sure that depicting this curve is really useful. In revanche when the authors say that “However analysis also shows the advantage of a constant leachate concentration” I do not see such analysis (or reference to the relevant paper)? What do you mean by constant leachate? Constant within a year or over the entire period?

Here we were attempting to show that a 2 store MIR model of NO3 could not reproduce the trend in Fig. 2 (as a discussion point). Our final MIR had 3 stores, and the results in Fig. 4 were therefore improved. Here “constant” is assumed to be constant over the entire simulation period.

p. 12, paragr. 2 “Type 4” events represent 75 % of the 12 identified events, meaning that 25% of the events will not be reproduced by the model. As events are an important contributor of PP export, computed reduction of P load in the mitigation scenario are probably overestimated.

This is a reasonable point but 25% of the event load of PP in the Frome is actually quite a small amount in terms of the overall TP export (due to it having 16% of the TP load originating from WWTPs) and only 1/5 from events (P14 para 1).

p. 13, paragr. 1, line 3: “If the model is ABLE (?) to capture”

This section has been removed.

p. 14, paragr. 2, line 8 (ii) : subsurface component is often supposed to be faster than the deep component due to difference of the hydrological properties in the material, or e.g. a decrease of transmissivity with depth.

We agree, as the transmissivity of soil is higher than the bedrock so therefore the SS component is a faster flow pathway than the DG component. No change necessary.

p. 14, paragr. 2, (2.2.2) flux rates unit should be m.day-1 (point or separation between m and day)

Thanks, we assume the copy editor will check the formatting of the units as spaces are sometimes lost when the documents are converted by the typesetting process.

p. 15 between eqs. 4 and 5: "into the subsurface DS and DG stores" should be "SS and Dg stores" isn't it?

Thanks for pointing out the typo, however we now changed this whole section

p. 15 between eqs. 6 and 7: the flow (QSUB) should be (QSS)?

Done

p. 15 Eq. 7: idem than above and S(t-1) should be SSS(t-1) isn't it?

No, regarding the above 2 points; we attempted to show a general form of the equation for flow in terms of storage (S), hence the terms QSUB, K and S (which represent QGW or QSS, KGW or KSS, SGW or SSS respectively). So the manuscript is correct.

p.16. After Eq. 10: What are the observed data used to calibrate K(N)? Do you have any observed concentration and flow data for the overland flow component?

Here "N" represents a nutrient (either N or P), perhaps this is confusing to readers as it could represent nitrogen, so we have changed it to use vector notation in Eq 11. We calibrated a K(**n**) value for PP, however for nitrate it was set to zero and a constant concentration in overland flow was assumed (discussed in results), which could generate the dilution observed in the HFD (discussed earlier and shown in Fig. 2 and is shown in table 3).

p. 16, Before Eq. 11 "The constant concentrationS In the dynamic"

Thanks

p. 17, Eq. 13: Add parentheses.

We edited the equation in the Word Equation Editor, the parentheses are not required unless the Copy Ed. wishes to add them?

p. 19, paragr. 1: So I understand that the hydrological criteria values did not change (80% and 10% for Nash and VE) otherwise it should be added in Table 3.

Correct

p. 19: Management scenario description is much clearer. Is there any reason to have chosen this scenario? Is it to have a significant response in term of modelled stream water quality? Or is it something actually discussed by the managers?

It represents a typical scenario reflecting a range of current issues in the European policy framework as cited (e.g. WFD).

p. 21, 3.3.3 paragr. 1, "The load from DG... After storm events" What do you mean?

The second part of this sentence describes the flow pathways that can occur immediately after the event (as defined by a period of rain and surface runoff). Note that this load is from the SS not the DG as the reviewer has stated but we feel our description makes it clear.

p. 21 last paragr., Addition of the uncertainty is appreciated!

p. 22, 3.3.2 Fig. 5 seems to show a small dephasing for SRP , could this be explained by some lack in the phasing of modelled QSS?

Hopefully zooming into a two year time period (Fig 5) in the Revision should make this clearer. There is a smaller phase error in the prediction of the flows than the nutrients.

p. 22, paragr. 3 : modelled event load is half of the one estimated from HFD. Could this be due to the non-simulation of event type different from 4? Or due to the time step?

We did calculate a load under prediction of 17% compared to both observed SRP and TP loads as estimated from the HFD (by Bowes et al., 2009a), however we did not discuss this detail as we wanted to complicate the paper. We wanted to just show the magnitude of changes from running the scenario.

p. 24, paragr. 2, line 13: "aN optimizing"?

Thanks for pointing out the typo

p. 25, paragr. 1, lines 3-4-5: Be careful here. The model does not take into account the variability of the inputs and the transfer from one store to the other. In particular it does not take into account the memory effect in the subsoil due to the microporosity. It won't be able to reproduce the past scenario with rising nitrate concentrations since the 1940s.

Interesting point, if longer time series are to be modelled. However our revised paper has stated the assumptions and limitations of the CRAFT. Using CRAFT the user can alter the N concentrations based on this type of knowledge, in fact that is what we are encouraging the user to do.

p. 25, paragr 1, Line 10-11 "The MI scenario…" It is because in the scenario, inputs are supposed to be reduced!!! With same inputs, reduction of overland flow could lead to increase leaching and enrichment of the SS S and DG stores, so should their concentrations increase as well (which is not the case in the model). Maybe it is not the case on this particular catchment as precipitation are high enough to flush all the surface store in a year, but under more limited infiltration context, this could accentuate the leaching of nutrient in deep stores.

Interesting point, here we assumed that the user seeks to reduce the input loads to the catchment (from all three pathways). Pollution swapping has been discussed, where the pathways for diffuse and point pollution can switch in terms of their relevant magnitudes.

Table 2: KSSF should be KSS? Thanks

Figure 4: the scheme still needs some improvements:
-inputs are represented while they are not injected in the model.
-outputs representation is a bit strange; concentrations would be more relevant from my point of view.

From our experience of working with policy makers they usually want a source apportionment representation. Therefore the aim of the bar chart representing output fluxes is to tie these in with the model results, shown in Fig. 7 as fluxes, under the baseline and MI scenarios, so we have not changed this in the Revision.

Marked copy is on the next page

The Catchment Runoff Attenuation Flux Tool, a Minimum Information Requirement N~~n~~utrient P~~p~~ollution M~~m~~odel

R. Adams[1], P.F. Quinn[2], M.J Bowes[3]

1. School of Civil Engineering and Geosciences, Newcastle University, Newcastle u~~U~~pon Tyne, NE1 7RU, UK.
2. School of Civil Engineering and Geosciences, Newcastle University, Newcastle u~~U~~pon Tyne, NE1 7RU, UK.
3 Centre for Ecology and Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, U~~K.~~

[2]Corresponding author

Abstract

A model for simulating runoff pathways and water quality fluxes has been developed using the Minimum Information Requirement (MIR) approach. The model, the Catchment Runoff Attenuation Flux Tool (CRAFT) is applicable to meso-scale catchments which focusses primarily on hydrological pathways that mobilise nutrients. Hence CRAFT can be used investigate the impact of flow pathway management intervention strategies designed to reduce the loads of nutrients into receiving watercourses. The model can help policy makers ~~, for example in Europe,~~ meet water quality targets and consider methods to obtain “good” ecological status.

A case study of the 414 km^2 Frome catchment, Dorset UK, has been described here as an application of the CRAFT model in order to highlight the above issues at the meso scale. The model was primarily calibrated on ten year records~~s~~ of weekly data to reproduce the observed flows and nutrient (nitrate nitrogen - N - and phosphorus - P) concentrations. D~~Also d~~ata from two years ~~of~~ with sub-daily ~~high resolution~~ monitoring at the same site were also analysed. These data highlighted some additional signals in the nutrient flux, particularly of soluble reactive phosphorus, which were not observable in the weekly data. This analysis has prompted the choice of using a daily timestep ~~for this meso-scale modelling study~~ as the minimum information requirement to simulate the processes observed at the meso-scale including the impact of uncertainty. A management intervention scenario

was also run to ~~show~~ demonstrate how the model can support catchment managers ~~to~~ investigat~~ing~~e how reducing the concentrations of N and P in the various flow pathways. This meso-scale ~~appropriate~~ modelling tool can help policy makers consider a range of strategies to meet the European Union (EU) water quality targets for this type of catchment.

Key words:

Hydrological Modelling, diffuse pollution, nitrate, phosphorus, land management

Colours:, Turquoise = Response to last review

1 ~~1~~ Introduction

The meso-scale is classed as catchments that vary between 10km^2 -1000km^2 (Blöschl, 1996). Uhlenbrook et al., (2004), states 'The satisfactory modelling of hydrological processes in meso-scale basins is essential for optimal protection and management of water resources at this scale'. It is therefore important that government policies on pollution abatement ~~must~~ be implemented at this scale. The EU Water Framework Directive (WFD) (European Parliament, 2000) has ~~increasingly~~ required catchments to meet in-stream standards in order to obtain "Good" ecological status. Therefore, all surface water bodies must meet exacting water quality and ecological targets (Withers and Lord, 2002). ~~Hence we require a~~There is a need for a framework that helps inform policy makers and regulators to understand the source of nutrient pollution at the scale of their interest.

Numerous models have been developed to simulate water and nutrient fluxes at ~~a~~ the meso~~catchment~~ -scale (e.g. INCA, Wade et al., 2002, 2006; PSYCHIC, Davison et al., 2008; SWAT, Arnold, 1994). ~~INCA has been used to investigate compliance issues with the WFD in terms of water quality (Whitehead et al., 2013).~~ These models have been used to underpin policy decisions and feed into the decision making processes with regards to the catchment land use~~ in catchments~~, and assess the impacts of any changes ~~to this~~ including source control or modified agricultural practices (Whitehead et al., 2013). However, these models tend to be too complex for ~~average~~ informed end users -to use and the simulations are prone to ~~high~~ having greater parameter uncertainty than simpler models~~ as they have a greater number of parameters~~ (McIntyre et al., 2005; Dean et al, 2009). Conversely export coefficients ~~(Johnes, 1996; Hanrahan et al., 2001)~~ can be an over simplification of reality and omit the role of event driven nutrient losses (Johnes, 1996; Hanrahan et al., 2001). ~~-~~A series of recent catchment scale studies have investigated the role of residence time and its variability in the export of nutrients (particularly nitrate and conservative tracers (e.g. chloride); Botter et al., 2011; Hracowitz et al., 2013;Van der Velde

Commented [RA1]: We inserted this paragraph in response to requests by the Ed. and R2 (Van Der Velde's work) et al., 2010), in small catchments (<10 km^2) to identify travel time distributions within a catchment. These studies focussed on a much smaller scale domain than this onesmall research catchments, with more extensive datasets, including high-resolution DEMs. Moreover, their scope was limited:, for example not onlyfirstly in terms of the number of different nutrients investigated; and secondly in the number of flow pathways; for example as Van Der Velde et al. (2010) only considered a single flow pathway (shallow groundwater) that transported nitrate from the catchment to the stream without any representation of overland flow in their model.

High frequency (defined here as containing sub-daily data) water quality monitoring data sets are becoming increasing available with newly developed auto-analysers and sondes (for example: Cassidy and Jordan 2011; Owen et al., 2012; Wade et al., 2012), and from high frequency samplers (Evans and Johnes, 2004; Bowes et al., 2009a).Here, a meta modelling approach is outlined of using a simple model structure to emulate more complex models and data sets (Fraser et al., 2013). The MIR (Minimum Information Requirement) approach will make a case as to what processes to include or exclude in the model. The question of how accurate a simulation needs to be when working at the meso-scale is raised. The study investigates the information content of the observed time series and thus we justify the use of a minimal model structure at this scale. The model retains sufficient complexity to allow management scenarios to be investigated and visualised at the meso-scale.

It is vital that models should aid policy makerscatchment planners when considering alternative strategies to attain policy objectives the likely consequences of policy needs (Cuttle et al., 2007; DEFRA, 20142015). This study aims to show that modellingmodel must include sufficient processes to reflect nutrient losses from the catchment which must be based primarily on soil and hillslope processes: such as overland flow; subsurface soil flow and slower groundwater dynamics (in temperate catchments). Hence the model must represent both chronic nutrient losses (seasonal fluxes), and acute losses (storm driven fluxes) (these terms were defined by Jordan et al., 2007). To this end we have developed ana MIR modelling approach was developed (Quinn et al., 1999; Quinn, 2004) which: (i): uses the simplest model structure; that achieves the current modelling goals; (ii); that uses process-based parameters that are physically interpretable to the users and so that the impact of any parameter change is can be clearl y interpreted by the end user (Quinn et al., 1999; Quinn, 2004). The CRAFT (Catchment Runoff Attenuation Flux Tool) has been developed to address these goals. Hence the MIR approach leads to is a parsimonious lumped model that capitalises on the mixing effects ofs aggregation and homogenisation of processes observed at the meso-scale.

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Commented [RA2]: In response to R3 we have revised this part of the text

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

High resolution monitoring data are becoming increasing available with newly developed auto-analysers and sondesWe are also living in a new era of high resolution datasets. These datasets may become invaluable to research-scale studies but at the meso-scale such detail may be less useful. More data are becoming available from high resolution monitoring using newly developed auto-analysers and sondes (for example: Cassidy and Jordan 2011; Owen et al., 2012; Wade et al., 2012), and from use of high-frequency samplers (Evans and Johnes, 2004; Bowes et al., 2009a). These datasets may become invaluable to research-scale studies but at the meso-scale such detail may be less useful. This study will attempt to show that high-frequency (defined here as containing sub-daily water quality data) data sets at this scale can help to justify the choice of a simpler MIR model. However these high frequency measurements may be prone to localised "noise" can introduce errors to the observations (Bowes et al., 2009a). Unravelling trends, seasonality and "noise" may require signal processing techniques to extract meaningful time series data and perform trend analysis (e.g. Kirchner and Neal, 2013). A case study will be shown that includes a both sub-daily and weekly time series of water quality, collected at the River Frome catchment in the Dorset (Marsh and Hannaford, 2008; Bowes et al., 2011;).

Formatted: Not Highlight

1.1

The MIR approach to modelling methodology

Here, a meta modelling approach is outlined of using a simple model structure to emulate more complex models and data sets (Fraser et al., 2013). The MIR (Minimum Information Requirement) approach will make a case as to what processes to include or exclude in the model. The question of how accurate a simulation needs to be when working at the meso-scale is raised. The study investigates the information content of the observed time series and thus we justify the use of a minimal model structure at this scale. The model retains sufficient complexity to allow management scenarios to be investigated and visualised at the meso-scale.

Formatted: Not Highlight

Formatted: Not Highlight

The MIR approach was developed partly as a response to a perceived excessive number of parameters in the established water quality and sediment transport models (Quinn et al., 1999; Quinn, 2004), and partly to address the issue of excessive model complexity to end user needs. The In principle s of MIR models are based on how much information can be gained from localised and experimental studies on nutrient loss, so that the most pertinent process components can be retained in the model and be easily manipulated and assessed by an end user.

Models derived through the MIR approach models must be suitable for use in the decision-making process in order to become a valuable tool. In this approach the issues that require addressing include: (i) the complexity of the model, (ii) linking nutrient losses and hydrological flow pathways and (iii) the ability to simulate both acute and chronic nutrient fluxes.~~ The use of such an approach leads to the following research questions:~~

~~1. How complicated does a MIR model need to be in order to address catchment management issues?~~

Formatted: Normal, Add space between paragraphs of the same style, No bullets or numbering

~~2. How important is it that the MIR model represents how nutrients are lost from the catchment, through the dominant hydrological pathways?~~

~~3. Does the model reflect the importance of acute losses of nutrients from the catchment during storm events and chronic losses during inter-event periods (and also any non-agricultural components)?~~

In the MIR approach, the modelling of runoff is ~~also~~ kept as simple as possible~~ to avoid excessive computation~~, although key runoff processes that influence nutrient and sediment loads are retained (Quinn, 2004). By creating a meta model of more complex process based models, a minimum number of processes are retained in the model structure that are required to satisfy a model goal: in this case the simulation of meso-catchment scale diffuse pollution. A series of simple equations are implemented in MIR models with a parsimonious number of parameters. The TOPCAT MIR family of models (Quinn, 2004, Quinn et al., 2008) were developed using this approach to simulate various sources of sediments and nutrients. Heathwaite et al. (2003) developed a simple spatial index model for estimating diffuse P losses from arable lands into waterways called the PIT (Phosphorus Indicators Tool). A series of Decision Support System (DSS)-based models were developed in Australia: commencing with E2 (Argent et al., 2009), then WaterCAST and finally SourceCatchments (Storr et al., 2011; Bartley et al., 2012). These have similar features of a MIR including: a daily simulation timestep to predict sediment and nutrient concentrations (*C*); and fluxes (i.e. *C* x daily flow); containing only two flow and nutrient pathways termed "event mean" i.e. storm flow, and "dry weather" i.e. baseflow, both assigned fixed *C* values for each sediment and nutrient simulated.

Formatted: Not Highlight

Formatted: Not Highlight

It is important that models are seen as useful in terms of the decision making process and its relationship to land use through a feedback mechanism between the regulators (DEFRA, 2015) and the land owners (e.g. farmers as in Cuttle et al., 2007) or holders of discharge consents into receiving watercourses (e.g. water companies) (Whitehead et al., 2013). ~~Hence, there is a need to re-interpret broad scale planning decisions and assess their likely impact on a single farmer or farming community. The key research question arising from this process relates to how large scale catchment management decisions impact~~

Formatted: Not Highlight

nutrient concentrations and fluxes at the scale of assessment. Modelling can highlight any potential problems such as changes in nutrient form, known as pollution swapping (Stephens and Quinton, 2009). In essence, the model shows how catchment management decisions impact nutrient concentrations and fluxes at the scale of assessment. In this study *pollution swapping* could show for example that SRP increases due to the mitigation measures that have reduced the concentration (and loads) of particulate P.

In this particular study we assess whether a particular water body is likely to become compliant within key regulations such as the WFD, although any other water quality standards could be used. CRAFT is meant to be just one of many tools that can be used to aid the planning process and address several catchment management issues. The key research question arising from this process relates to how large scale catchment management decisions impact nutrient concentrations and fluxes at the scale of assessment. If the aim of the modelling study is to determine a total export of nutrients from the catchment outlet then simulating all the processes within the catchment may not be required and an export coefficient model (e.g. Johnes, 1996; Hanrahan et al., 2001) may be useful. However, the provenances of the fluxes still need to be linked back to local sources, pathway and nutrient loading factors. The modelling process seeks to link science and process knowledge gained at the local 'research scale' (1 m^2–10 km^2) with a larger (meso-scale) catchment (100 – 500 km^2) 'applied science' scale (Haygarth et al., 2005). Hence, the astute choice of model structure and timestep allow a scale appropriate MIR model to be set up. …..

At larger catchment scales mixing processes may dominate the final observations at the outlet , and the choice of sampling frequency will still be important if load estimates are required (Johnes, 2007). Tand the temporal fluctuations in runoff and water quality observed in headwater research catchments may not necessarily be observed at the outlet of the larger catchment area (Haygarth et al., 2005, 2012; Storr et al., 2011). As a rule therefore, the smaller the catchment the more detail is required in the model to define processes, but as the catchment size increases then in-stream processes associated with channel routing and the effect of point sources (especially of P) will tend to take over from nutrient generation processes in influencing the signal observed at the outlet of a larger catchment (Haygarth et al., 2005, 2012).

A series of recent catchment scale studies have investigated the role of residence time and its variability in the export of nutrients (particularly nitrate and conservative tracers (e.g. chloride); Botter et al., 2011; Hracowitz et al., 2013;Van der Velde et al., 2010), in small catchments (<10 km^2) to identify travel time distributions within a catchment. These studies focussed on a much smaller scale domain with more extensive datasets, including high resolution DEMs, than this one. Moreover, their scope was limited, for example not only in terms of the number of nutrients investigated as Van Der Velde et al.

Formatted: Highlight

Formatted: Highlight

(2010) only considered a single flow pathway (shallow groundwater) that transported nitrate from the catchment to the stream without any representation of overland flow in their model.

The goal here is to develop a model that contains a useful and parsimonious set of parameters resulting in a “visual thinking tool” that can provide a semi-quantitative risk-based assessment of management decisions. The CRAFT model described below is written in a MS Excel spreadsheet and the results, graphs and load calculations update instantaneously; hence the consequence of changing the parameter values on all the outputs (e.g. runoff and nutrient load) can be seen immediately. Instead of expecting the end user to perform an explicit uncertainty analysis, they are encouraged to investigate the sensitivity of the output fluxes to a wide range of parameter values. Hence, the onus is on the user to think through the meaning of the parameters and the implications of changing their values.

Commented [RA3]: R2: Here we are justifying and explaining the goal of the MIR approach

1.3 The Spatial and Temporal Scales of the Data

High-frequency water quality monitoring has become achievable over the last decade, firstly with the availability of automatic water samplers (Bowes et al., 2009a) enabling several measurements per day to be taken (e.g. sub-daily measurements of concentrations of sediments and nutrients). Recent examples of long term monitoring at a high temporal frequency include the DTC (Demonstration Test Catchments edendtc.org.uk/) study in the UK, based in the Eden catchment in Cumbria (Owen et al., 2012), the monitoring in the Blackwater catchment, Ireland (Cassidy and Jordan, 2011), and the Irish Agricultural Catchments project (www.teagasc.ie/agcatchments/), and the monitoring of the Enborne and Kennet subcatchments of the Thames by Wade et al. (2012). These studies were made possible by the development of bankside nutrient auto-analysers (Jordan et al., 2007) which have allowed very high-frequency (hourly / sub-hourly) data sets to be assembled. These data have enabled better estimation of nutrient export from catchments to be made for the first time (Bowes et al., 2009a; Johnes, 2007). The growth of these data sets allows us to pose an additional research question as to what is the value of collecting high-frequency data to parameterize models at the medium-large catchment scale (100-500 km^2). However these high frequency measurements may be prone to localised “noise” can introduce errors to the observations (Bowes et al., 2009a). Unravelling trends, seasonality and “noise” may require signal processing techniques to extract meaningful time series data and perform trend analysis (e.g. Kirchner and Neal, 2013)).

The modelling process seeks to link science and process knowledge gained at the local ‘research scale’ (1 m^2-10 km^2) with a larger (meso-scale) catchment (100 -500 km^2) ‘applied science’ scale (Haygarth et al., 2005). Hence, the astute choice of model structure and timestep allow a scale appropriate MIR model to be set up.

Commented [pfq4]: We need to check what references have been dropped

At larger catchment scales mixing processes may dominate the final observations at the outlet, and the choice of sampling frequency will still be important if load estimates are required (Johnes, 2007). The

~~temporal fluctuations in runoff and water quality observed in headwater research catchments may not necessarily be observed at the outlet of the larger catchment area (Haygarth et al., 2005, 2012; Storr et al., 2011). As a rule therefore, the smaller the catchment the more detail is required in the model to define processes, but as the catchment size increases then in-stream processes associated with channel routing and the effect of point sources (especially of P) will tend to take over from nutrient generation processes in influencing the signal observed at the outlet of a larger catchment (Haygarth et al., 2005, 2012).~~

~~Thi~~Thiss paper aims to ~~to~~ show how a parsimonious model, driven primarily by hydrological processes, can help ~~reflect~~predict nutrient fluxes at the meso-scale. End users can manipulate such a tool to help underpin ~~there~~their decision making process related to local land management and evaluate the ~~oustputs~~outputs likely to occur at that scale. The paper:~~To achieve this aim the paper~~

1. ~~Reviews a~~Introduces meso-scale case study where both weekly and sub daily nutrient data was available. The information content of the flow and nutrient data is evaluated and hence recommendations are made as to what phenomena can be realistically modelled ~~are made~~.

2. The CRAFT model is described and a justification is made as to what processes and parameters need to be included in order to simulate both event and seasonal phenomena.

3. The sensitivity of the model outputs to hydrological process is ~~made~~analysed ~~and thus~~with an example of a land use change scenarios ~~are~~ demonstrated.

~~This~~Hence the paper ~~will suggest~~identifies that there are useful simplifications ~~i~~that can be made tos model structure ~~that can be made that can~~in order to simulate observed fluxes at ~~this scale~~the meso-scale. ~~.~~ This is particularly important for policy makers who must comply with the relevant~~meet~~ regulations at this scale.

2 Methods

2.1 ~~Case Study~~Catchment Description

The case study focusses on the 414.4 km^2 River Frome catchment (Fig. 1) which drains into Poole Harbour with its headwaters in the North Dorset Downs (Bowes et al., 2011; Marsh and Hannaford, 2008; Hanrahan et al., 2001). Nearly 50% of the catchment area is underlain by permeable Chalk bedrock, the remainder consists of sedimentary formations such as tertiary deposits along the valleys of the principal watercourses (including sand, clay and gravels). There are some areas of clay soils in the lower portion of the catchment. However, most of the soils overlaying the chalk bedrock are shallow and well drained. The land use breakdown is dominated by improved grassland (*ca*. 37%, comprising

Commented [RA5]: Here, the order of 2.1 and 2.2 has been swapped over in the Revision

Formatted: Not Highlight

hay meadows, areas grazed by livestock and areas cut for garden turf production), and *ca*. 47% tilled (i.e. arable crops primarily cereals) usage (Hanrahan et al., 2001). The major urban area in the catchment is the town of Dorchester (2006 population over 26000, Bowes et al., 2009b) otherwise the catchment is predominantly rural in nature.

The mean annual catchment rainfall was 1020 mm and mean runoff 487 mm from 1965 to 2005 (Marsh and Hannaford, 2008). ~~The major urban area in the catchment is the town of Dorchester (2006 population over 26000, Bowes et al., 2009b) otherwise the catchment is predominantly rural in nature.~~ At East Stoke the UK Environment Agency (EA) has recorded flows since 1965. The Centre for Ecology and Hydrology (CEH) and Freshwater Biological Association have collected water quality samples at this same location at a weekly interval from 1965 until 2009 (Fig. 1) (Bowes et al., 2011), see 2.1.2 below.

~~*Perhaps the next section should come in the discussion when we compare modelled with observed loads?*~~

Hanrahan et al. (2001) ~~presented~~ calculated both export coefficients for diffuse sources of TP, and load estimates for diffuse and point sources (comprising: WWTPs (serving Dorchester plus other towns); septic systems; and animal wastes). The total annual TP (total phosphorus) export from diffuse sources in the catchment was estimated to be 16.4 tonnes P yr^{-1}, a yield of 0.4 kg P ha^{-1} yr^{-1}. Point source loads from WWTPs, septic systems and animals added an extra 11.5 tonnes P yr^{-1} (from the data in Table 2 in Hanrahan et al. (2001)) to the catchment export, giving a total load of 27.9 tonnes P yr^{-1}. Nitrogen (as nitrate) export from the catchment in the mid-1980s was estimated by Casey et al. (1993) to be 21.6 kg N ha^{-1} yr^{-1}, with 7% of this originating from point sources in the catchment.

2.1.1 ~~Hydrological~~ Meteorological Data

Forcing data (precipitation) was supplied by the EA for the period 1997 to 2006 which was therefore chosen as the modelling period. A single raingauge, Kingston Maurwood (ST718912) located ca. 4 km downstream of Dorchester, was used for the modelling as this gauge had the most complete record and was centrally located in the catchment. Daily mean and 15-minute interval flow data were also provided from East Stoke gauging station for the same time period. Potential Evapotranspiration (PET) was derived using an algorithm developed to calculate a daily PET based on monthly temperature patterns, in order to obtain a daily PET time series which when totalled for the year would match the estimated annual PET (465 mmyr^{-1}). Given the dominance of winter runoff in the Frome catchment the model predictions are unlikely to be sensitive to input values of PET.

Commented [pfq6]: I think this is charaterisation but it could be shortened

Formatted: Font: Italic

Formatted: Not Highlight

Commented [RA7]: In response to R2

~~Daily rain gauge data was obtained from Kingston Maurwood (ST718912) located ca. 4 km downstream of Dorchester. Earlier studies have noted some spatial variation in precipitation across the catchment (Bowes et al., 2011), and Smith et al. (2010) reported that between 1993 and 2008 there were 3-5 gauges operational in the catchment (albeit with missing data). We understand that model errors sourced from rainfall are likely to be significant and may influence predictions of overland flow (where rainfall is an important factor) and the associated nutrient transport by this pathway. However, we did not feel that it was appropriate to develop a spatially distributed model of the Frome catchment (incorporating multiple rainfall timeseries inputs) given the focus was on predicting *Q* and associated nutrient fluxes at the catchment outlet only.~~

2.1.2 Monitoring Datasets

Two sets of water quality monitoring data were used in this study ~~(Table 1 below shows the statistics relating to long term concentrations) along~~ with daily flows recorded by the Environment Agency at East Stoke gauging station. The data were compared and analysed so that the MIR model could be defined. The attributes of the data are described in Table ~~0~~1 and long term statistics relating to nutrient concentrations are ~~found~~listed in Table ~~*~~2. ~~To summarise~~

~~(1) T firstly~~The first is the CEH/Freshwater Biological Association long-term dataset (LTD) of water quality for the River Frome (Bowes et al., 2011; Casey, 1975; open access via gateway.ceh.ac.uk) ~~was collected from 1965 to 2009 at a near-continuous weekly interval (average number of observations per year = 48) and thus represents one of the longest (relatively) high frequency datasets on water quality in existence from the UK. In this study we analysed their nitrate-N (nitrate) from 1997 to 2006, and their TP and SRP data between 1997 and 2002~~~~was analysed from 1997 to 2006~~.~~.~~ After March 2002 the introduction of P-stripping measures at Dorchester WWTP produced a step reduction in SRP concentrations and reduced SRP loads by up to 40%, according to the analysis of Bowes et al. (2009b~~, 2011~~)~~, which produced a step reduction in stream SRP concentrations~~. ~~The statistics for the periods of analyses are shown in Table 1.~~

~~(2) A~~~~Secondly~~The second dataset (Table 1) is~~,~~ a high frequency data set (HFD) described in Bowes et al. (2009a) which ~~,~~ was also collected at East Stoke overt a shorter period ~~between 1/2/2005 and 31/1/2006,~~ using a stratified sampling approach and EPIC™ water samplers (Salford, UK)~~ (Table 0)~~. High resolution measurements may be prone to localised "noise" that can introduce errors to the observations (Bowes et al., 2009a). Unravelling trends, seasonality and "noise" may require signal processing techniques to extract meaningful time series data and perform trend analysis (e.g. Kirchner and Neal, 2013).

Formatted: Not Highlight

Formatted: Not Highlight

Commented [RA8]: R2 mentioned additional raingauges, here we have simplified the discussion to state what was actually used

Formatted: Not Highlight

Commented [RA9]: It needs to come somewhere, perhaps move it down to the section below where Fig 3 now is, as we discuss loads here.)

Commented [pfq10]: Do we need this??

Commented [pfq11]: Should we cut this??

2.1.3 Temporal Runoff and Nutrient Behaviour in the Frome Catchment (LTD and HFD)~~The statistics related to nutrient concentrations are shown in Table 1. The frequency of the water samples varied between two to four times daily during dry periods with up to eight samples per day during rainfall events. The average number of samples was 3.7 per day. Also in the dataset were river flow (Q) values taken from the Environment Agency 15 minute interval flow data. In this study we used the Q, TON, TP and SRP data. A more detailed discussion of the two datasets follows in order to justify several MIR simplification assumptions.~~

~~Firstly, the~~ The flow timeseries of the LTD (daily mean flows; DMF) and HFD (sub-daily) flows were compared over the HFD ~~course of the high resolution~~ monitoring period ~~described in Bowes et al. (2009a)~~ and both time series of flows are shown in Fig. 2a along with the residuals. For most of the period both sets of flows closely matched (ρ = 0.98) except perhaps during runoff events of less than a day where the HFD flows were sometimes higher as indicated by the positive residuals. The analysis suggests that, for modelling purposes including load estimation, that a daily timestep can capture the variability in the observed data without the need to use an hourly timestep.

For nitrate it is assumed that nitrite concentrations were negligible in the LTD dataset (Bowes et al., 2011) so that TON concentrations (equivalent to nitrate plus nitrite) were effectively equal to nitrate. This allows the HFD TON data to be directly compared against the observed (weekly LTD) nitrate data. The patterns observed visually (i.e. locations of the peak *Cs*) in the weekly and high ~~resolution~~ frequency nitrate/TON timeseries were very similar indicating that the weekly monitoring data were probably sufficient to estimate the range of nitrate/TON concentrations in the catchment, in order to assess compliance with EU WFD quality standards (in this case ensuring that $C \leq 11.9$ mgL^{-1} N). ~~The monitored periods overlapped (Fig 2b)~~ In Fig. 2b it can be seen that~~and~~ there were a few spikes in the HFD above concentrations measured by the LTD, with those measured during recession spells in the flows~~,~~ generally being less than 1 mgL^{-1} N in magnitude. ~~The correlation between C and Q was weak (in the HFD ρ = 0.12), due to the complex SRP concentration / flow relationships caused by point source dilutions at low flows and increasing diffuse inputs at higher flows (Bowes 2009b). Therefore, it would not be possible to develop a Q vs. C rating curve to estimate loads from this dataset using the methods used by Cassidy and Jordan (2011).~~ There was also no evidence that high flows would generate correspondingly high nitrate concentrations~~. In~~ and in fact, in Fig. 2b a dilution effect can be clearly observed during several events in autumn 2005 (indicated by "1", and the dashed blue line linking the concentration timeseries to the corresponding events in the hydrograph in Fig 2a), with lower concentrations ~~lasting~~ persisting in some cases for several days ~~in some cases during the subsequent period of high baseflow~~after the event. This indicates that concentrations of nitrate in the combined slower baseflow / sewage effluent must ~~be~~ have been higher than concentrations in rapid overland flow.

For phosphorus the HFD SRP data were compared visually with the LTD SRP data in Fig. 2c and again the patterns in both datasets were broadly similar, with increasing concentrations during the summer

Formatted: English (United States)

Commented [RA12]: I've added a new heading here for the analysis section, I moved the load calculations up to this section and shortened the description to the minimum

Formatted: Font: Bold

Formatted: English (United States)

Formatted: Font: Not Italic, English (United States)

Formatted: English (United States)

Formatted: Font: Not Italic, English (United States)

Formatted: English (United States)

Formatted: Not Highlight

Commented [RA13]: In response to R2

Formatted: Not Highlight

period between May and November 2005. HFD TP concentrations are also shown in Fig 2c by the red line. Between November 2004 and March 2006 there was a gap in the LTD TP data for operational reasons discussed in Bowes et.al (2011). ~~Flow data from the upper panel (Fig. 2a) will be used to illustrate s~~Several key points arising from the HFD data are:

(i) Some of the spikes in TP concentration, for example in February and mid-December 2005, were during the falling limb or low-flow periods of the hydrograph and were not associated with significant storm runoff events. Corresponding spikes in SRP concentration were not usually prominent at these times except for one in January 2006. (Examples are indicated by “2” on Fig. 2c). Some spikes were also observed during medium flow periods on several occasions in summer 2005, without corresponding SRP spikes but during a period where SRP concentrations were increasing. (Examples are indicated by “3” on Fig. 2c).

Three events between November 2005 and 1st January 2006 did generate high concentrations in PP that coincided with the storm peak in the flow hydrograph (>1 mg/L P). This could indicate a faster mobilisation of PP into the channel system during wet conditions in autumn-winter 2005 compared to summer storms. Haygarth et al. (2012) have observed similar peaks in PP in smaller headwater catchments due to sheet flow events. (Examples are indicated by “4” on Fig. 2c). Some s~~S~~maller “Type 4” events were also observed between February and April 2005.

(ii)

(iii) Some SRP concentration spikes were not simultaneously observed in the TP concentrations, these may have been due to WWTP discharges or leaky septic tanks (the high sampling frequency ~~sampling methods~~ permitted this to be observed; Bowes et al. (2009a)). Examples of ~~this~~ these are indicated by “5” on Fig. 2c.

SRP concentrations during the summer months tended to increase by approximately 0.07 mgL^{-1} P indicating chronic sources of nutrients in the catchment whereas acute sources tended to be associated with runoff events or other events in the catchment not associated with high flows. Bowes et al. (2011) also observed this phenomenon in the LTD dataset and suggested that the probable cause was a combination of lower flows with less dilution of SRP in the river originating from point sources (WWTPs) in the catchment. Jordan et al. (2007) attributed acute sources of TP in their 5 km^2 agricultural catchment in Northern Ireland to applications of slurry and inorganic P during periods of low rainfall (with no associated runoff events).

Of the 12 runoff events observed between February 2005 and Feb 2006, 9 were classified as “Type 4” events in terms of TP, where a corresponding increase in TP *C* was also observed (Fig 2c). The total

Commented [RA14]: Reviewer 3 had commented on missing TP data, it is explained here.

Formatted: List Paragraph, Numbered + Level: 1 + Numbering Style: i, ii, iii, … + Start at: 1 + Alignment: Left + Aligned at: 0.63 cm + Indent at: 1.9 cm

Formatted: List Paragraph, Numbered + Level: 1 + Numbering Style: i, ii, iii, … + Start at: 1 + Alignment: Left + Aligned at: 0.63 cm + Indent at: 1.9 cm annual loads (1/2/2005-31/1/2006) of TP and SRP were estimated from the HFD using simple baseflow separation and load analysis techniques as carried out by Haygarth et al. (2005) and Sharpley et al. (2008) in order to estimate the percentage of the annual TP load generated by events. These loads (with the % contributed from the 9 runoff events in brackets) were estimated to be 27.8 tonnes TP (20.0 %) and 13.1 tonnes SRP (17.7 %) respectively.

Figure 3 goes around here

The total annual TP loads are shown in Fig. 3 as a pie chart that indicates the percentages due to event and non-event sources. The percentage of the SRP load from point sources (mostly WWTPs) was estimated to be 34% based on Bowes et al. (2011) and is indicated by the dashed segment (i.e. 4.5 t P). Making the further assumption that PP = TP-SRP allowed the PP load to be estimated as well (here the "PP" load estimate will probably include a component of unreactive, organic P, so it will be an overestimate) to be 14.8 tonnes PP (22.1 % from events).

The correlation between C and Q was weak (in the HFD ρ = 0.12), due to the complex SRP concentration / flow relationships caused by point source dilutions at low flows and increasing diffuse inputs at higher flows (Bowes 2009b).

The HFD dataset shows the range of concentrations that are seen in reality which are often missed in weekly and monthly datasets. These data also show the problem of noise and incidental events that are not correlated to storms. Hence the meso-scale model requires a structure that can address the identifiable seasonal and event driven patterns but equally should not be expected to exhibit high goodness of fit metrics. Any calibration therefore, should be logical and not misleading to the user and an acceptance that the uncertainty is high must remain. However, the impact of any manipulation of input parameters should be observable and self-explanatory to an informed user.

2.2 Model Description

2.2.1 Developing the CRAFT model using the MIR approach

The justification for including some processes and omitting others is a difficult task in modelling. Hence it is worth firstly reviewing the MIR process to date. CRAFT has evolved from the model TOPCAT-NP (Quinn et al., 2008). In terms of the hydrology, TOPCAT-NP contained a dynamic store model as used in TOPMODEL and a constant (flow and concentration) groundwater term, however the Topographic Wetness Index was removed. TOPCAT-NP also contained a time varying soil leaching model for N and SRP (with an associated soil adsorption term for SRP).

Commented [RA15]: In response to Greg's comment I moved discussion of loads out of the model development section

Formatted: Not Highlight

Formatted: Not Highlight

Commented [RA16]: Now comes after the Case Study Description

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

In terms of nutrient process modelling (in TOPCAT-NP), a meta-modelling exercise of the physically based model EPIC (simulating flow, SS, N and P) (Williams, 1995) and the N-loss model SLIM (Solute Leaching Intermediate Model) (Addiscott and Whitmore, 1991) was~~ere~~ carried out and are published in Quinn et al. (1999). Herein a case wa~~i~~s made to reduce many of the soil hydrological and chemical processes. Multiple simulation of EPIC showed that both the annual exports and the daily losses could be readily simulated by a leaching function and knowledge of how much N or P was being applied and available for mobilisation. Based on these earlier studies, the final version of TOPCAT could simulate flow, N and P at a number of research locations (hence the suffix "-NP"). It included a leaching model; hence a soil nutrient store and a leaching term based on a soil type parameter were required to determine the flux into the store.

Essentially the MIR formulation is thus a series of mass balance equations that sum the flux of nutrients $F=C.Q$ from each store over time to obtain a nutrient load. In order to study nutrient pools and/or explicit soil flux processes then a physically based model is required (e.g. Arnold (1995); Van der Velde (2010); Hracowitz et al., 2013). The HFD dataset (Section 2.1.2) described above is used to estimate the likely origin and magnitude of nutrient fluxes in the catchment and help inform our choice of model structure in terms of processes and stores. The second simplest form of a MIR water quality model (other than merely using a constant concentration of nutrients in all the stores) is the EMC/DWC formulation (Argent et al., 2009) with two stores: (i) "Dry Weather", i.e. baseflow; (ii) "Event Mean", i.e. overland flow events in this case. Each store is represented by a single, constant *C* value, i.e. DWC and EMC respectively.

The results of modelling nitrate using ~~this~~a two-store MIR model can be seen in Fig 2b by the green line. ~~The modelled period corresponds to the HFD data period.~~ The two *C* parameters are respectively 6.5 mg/L N (DWC) and 2 mg/L N (EMC). Here, the "flow" component of the MIR is able to reproduce events (here with lower nitrate *C*) reasonably well, but the background nitrate *C* is not reproduced well during the summer-autumn period since the model overpredicts it between July-November 2005. A similar phenomenon could be demonstrated using the SRP dataset with this structure of MIR model. The modelling of the Frome catchment using a CRAFT MIR will be revisited later, but this exercise neatly illustrates how an MIR model can be too simple to represent all the phenomena that are detectable in the observations. Thus TOPCAT-NP's ~~the~~ constant (flux and *C*) groundwater term ~~of TOPCAT-NP model is~~ was hence too simple for this study. ~~However, analysis also shows the advantage of using a constant leachate concentration and thus the soil leaching model of TOPCAT-NP was replaced.~~

The signals observed in the HFD dataset are examined slightly more deeply, in order to further develop the conceptual MIR model processes (particularly for P). ~~A caveat here is that this analysis is fairly crude and intended to illustrate the MIR model development strategy only.~~ ~~Firstly, it is necessary to~~

Formatted: Not Highlight

Commented [RA17]: I have removed observed data and loads to the previous section and trimmed down this paragraph to focus on model development only with no loads or other derived data here

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

make some assumptions about runoff “events”, such as they are defined from the flow hydrograph as an increase of >0.5 mm/d in the observed flow, and there are 12 such events identifiable in the flow timeseries (Fig 2a). Of the 12 runoff events observed between February 2005 and Feb 2006, 9Nine of the twelve events discussed above were classified as “Type 4” events in terms of TP, where a corresponding increase in the TP *C* was also observed (Fig 2c). These should be incorporated in a MIR model, if it is to be a useful predictive tool for modelling P event fluxes and TP loads due to events., Performing baseflow separation enabled the flow hydrograph and load timeseries to be split into event and baseflow components, as carried out by Haygarth et al. (2005) and Sharpley et al. (2008) in order to estimate the percentage of the annual TP load generated by the events. The total annual loads (1/2/2005-31/1/2006) of TP and SRP were estimated from the HFD to be (with the % contributed from the 9 runoff events in brackets):

TP = 27.8t P (20.0 %)
SRP = 13.1 t P (17.7 %)

*Figure 3 goes around here*The total annual TP loads are shown in Fig. 3 by as a pie chart that indicates the percentages due to event and non-event sources. The percentage of the SRP load from point sources (mostly WWTPs) was estimated to be 34% based on Bowes et al. (2011) and is indicated by the dashed segment (i.e. 4.5 t P).

Commented [RA18]: Many changes were made to this section so we have highlighted the entire text

Formatted: Not Highlight

Making the assumption that PP = TP-SRP allowed the PP load to be estimated as well (here the “PP” load estimate will probably include a component of unreactive, organic P, so it will be an overestimate):

PP= 14.8 t P (22.1 %)

The fact that (according to the HFD data suggest that) one fifth of the total P load over a year in the Frome catchment was generated by events (Fig. 3), mostly due to elevated PP fluxes indicates that including a process in the final MIR model that can generateby generating TP (as PP) from runoff events will be important, if the model is to captureso that the model is able to reproduce the observed TP dynamics and accurately estimate the TP loads. The fraction of PP estimated from the load analysis to have been exported during events was 3.27t which equated to 12% of the overall TP load. The HFDIn data shown in Fig 2c also it was indicated that the TP Cs during “Type 4” events were quite variable (it was highest in late autumn-winter 2005) so that using a constant C value in the overland flow/surface process store in a MIR model would be an oversimplification.

The Type 2 and 3 events discussed above generated spikes of relatively high TP *C*s and Type 5 events generated spikes of SRP *C*s that were not associated with significant catchment rainfall, or flow events observed at the outlet (Fig 2c). Therefore, in terms of total annual P loads the Type 2 and 3 events contributed a very small percentage% of the total (mainly due to the low flows at the time of occurrence,

Formatted: Not Highlight

~~), which on about 6 occasions during the 12 month period only accounted for an additional load of approximately 30-200 kg/day of TP relative to the baseflow loads of TP. These are grouped together into "Other events" in Fig. 3~~ and may have been generated by incidental losses.

In Fig 2b it was shown ~~that~~ with ~~from~~ the HFD TON signal ~~indicated~~ that many of the runoff events ~~were~~ were categorised as "Type 1" ~~for TON,~~ where dilution of the TON, presumably due to overland flow, was observed. A similar analysis to that carried out with the TP data ~~wa~~~~is~~was ~~therefore~~ not appropriate as it was~~is~~ clear ~~therefore~~ that the TON *C* ~~of TON~~ in overland flow during events must ~~be~~ have been lower than the ~~C~~ observed *C* in the baseflow in order to have caused the dilution ~~signal~~patterns. ~~It is thus important that~~Thus the MIR model ~~can~~ should capture: (i) a dilution signal; (ii) the observed variations in TON *C*s, particularly the decrease observed between later winter and summer (i.e. in the winter 2005-6 period from ca. 7 mg/L to ca. 4 mg/L followed by a recovery back up to 7 mg/L).

The two store ~~(e.g. EMC/DWC)~~ MIR model ~~discussed above~~shown in Fig. 2c was unable to reproduce any seasonal patterns at all in the observed TON HFD data.

Therefore, it was decided that an additional flux term (and store) was required in the model to represent a time-varying baseflow component from deeper groundwater (GW). This modification also had a similar beneficial effect on the modelling of the SRP concentrations. The shape of the flow hydrograph and some background information on the catchment physical characteristics (Casey et al., 1993; Marsh & Hannaford, 2008) suggested ~~to~~ that an improved representation of the subsurface flow processes was important in the Frome catchment.

In meso-scale catchments such as ~~the Frome~~this a physically-based leaching function (~~described above~~ as used in TOPCAT-NP; Quinn et al., 2008) thus also becomes redundant— as the 'minimum requirement' is to know the concentration of the nutrients at the outlet and it is assumed that fluxes of N and ~~SR~~P are being generated at some location in the catchment throughout the year, due to the (assumed uniform) spatial distribution of intensive agricultural land uses. These fluxes are thus incorporated into a soil flux store in the final MIR with this flux assigned constant *C*s of SRP and N.

The development of the conceptual model discussed above led to an MIR structure for the CRAFT model that represents the complex hydrological system in the simplest manner feasible~~The above discussion led to for the CRAFT model being chosen, it representing a MIR representation of a more complex hydrological system~~. The upper pane of Fig. 4 shows that the model comprises three dynamic storages and the associated flow and transport pathways (or fluxes). The lower pane in Fig. 4 shows the flow and nutrient transport pathways that exist in a catchment such as the Frome using a conceptual cross-section of a hillslope. Here, inputs and outputs of N and P in the catchment are shown

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

diagrammatically. There are three flow pathways shown: (i) an overland flow component which also represents processes in the cultivated near surface layer (down to several centimetres depth); (ii) a faster subsurface component encapsulating agricultural soils that may have been degraded by anthropogenic activities and perhaps enhanced flow connectivity (e.g. through field drains); (iii) a slower groundwater component encapsulating any background flow in the catchment due to: deeper flow pathways; Wastewater Treatment Plants (WWTP) discharges (assumed constant); and other non-rainfall driven constant fluxes including any generated within either the channel or the riparian areas. We will refer below to the pathways as: (i) overland flow (OF); (ii) ~~as~~ fast subsurface soil flow (SS); and (iii) as the slow, deeper groundwater flow pathway (DG) respectively. It has been argued above that the composition of SRP and nitrate fluxes must be dominated by the DG and SS pathways. The TP flux includes a PP component that is generated by the OF pathway in the model (as discussed above)~~.~~.

2.2.2 Water Flow Pathways

There are six parameters that require estimation or calibration to control the water flow pathways. Their values are shown in Table ~~2~~ 3 below.

The uppermost dynamic surface store (DSS) is conceptualized to permit both crop management and runoff connectivity options to be examined. The DSS store is split into two halves with the upper half representing a cultivation (tillage) layer that generates overland flow, and the lower half ~~accounting for controlling~~controls the ET and the drainage ~~control~~ rate to the lower stores. Firstly, a water balance updates the storage (SS) and then computes the overland flow from the surface store (QOF) through the following equations, where *R* is rainfall, *D* drainage to the lower half of the store. Note that all stores are in units of length (e.g. m) and all flux rates (e.g. *R*, *D*, QOF) are in units of length per time step (e.g. m day^{-1})

$$SS(t) = SS(t-1) + R(t) - QOF(t-1) - D(t-1) \quad (1)$$

$$D(t) = \mathrm{Min}(SDMAX, SS(t)) \quad (2)$$

$$QOF(t) = (SS(t) - D(t)) \cdot KSURF \quad (3)$$

The parameter SDMAX can be used to deliberately partition excess water between surface and subsurface flows which is crucial for investigating connectivity options and possible pollution swapping effects. The lower half of the SCS represents the soil layer (below the cultivated layer) and also accounts for ET in the model. The parameter limiting the size of the store is called SRZMAX. The storage of water in the store (SRZ) at each time step is updated by the following mass balance:

$$SRZ(t) = SRZ(t-1) + D(t) - ET(t) \quad (4)$$

Formatted: Not Highlight

Commented [RA19]: Reviewer 2: Stated in response that the other sources are constant

Formatted: Not Highlight

Formatted: Not Highlight

Commented [RA20]: Order swapped over in response to R2

Formatted: Not Highlight

Commented [RA21]: This section was shortened with the equations used to describe the model structure rather than text

Any excess water present in the store above SRZMAX will form percolation (PERC) which then cascades into the subsurface ~~DS~~ SS and DG stores. SRZ is then reset to SRZMAX:

$$PERC\ (t) = MAX\ (0, (SRZ\ (t) - SRZMAX)) \quad (5)$$

Both the SS and DG stores are dynamically time varying and generate fast (QSS) and slow groundwater flows to the outlet (QGW) respectively. A dimensionless parameter SPLIT (0,1) apportions active drainage from the lower surface store towards either store, i.e. a water balance for the storage (SSS) in the SS store can be written as

$$SSS\ (t) = SSS\ (t - 1) - QSS\ (t - 1) + PERC(t) \cdot SPLIT \quad (6)$$

The equation for the storage in the DG store (SGW) is identical except that (1 - SPLIT) is substituted for SPLIT.

The flow (QSUB) from either subsurface store is described by Eq. (7) where K is a recession rate constant (d^{-1}) and S is the storage (in m). Therefore QSUB at time *t*, is given by

$$QSUB\ (t) = K \cdot S(t-1) \quad (7)$$

In the DG store the initial storage SGW0 is set by the user ~~by specifying an initial value of the resulting flow (QGW0, where we are using the suffix "GW0" to denote initial value of slow groundwater flow) rather than explicitly defining the storage (which is difficult to estimate in a complex catchment). It is convenient to commence the model simulation during a dry spell, where the slow groundwater component is usually relatively constant and most of the runoff consists of this flow. Therefore, rearranging~~by rearranging Eq. (7) in terms of the groundwater discharge at the start of the simulation (assumed to be equal to the observed flow in a dry spell) QGW0~~to invert its terms gives~~

$$SGW0 = QGW0/KGW \quad (8)$$

Where QGW0 ≡ Observed runoff on first day of simulation (m d^{-1}), following the assumption above

Lastly, the total modelled runoff at each timestep, at the outlet is calculated (QMOD)

$$QMOD = QOF + QSS + QGW \quad (9)$$

2.2.3 Nutrient Fluxes

Commented [RA22]: New heading in response to R2, splitting "flow" and "nutrient" model description

The ~~'informed'~~ user must now add a sensible range of input nutrient concentrations to the model in order to simulate loads (i.e. *C x Q*). They are encouraged to set and alter these values and see the impact instantaneously. The nutrient transport processes are conservative and the user is encouraged to understand the link between land use management and the level of nutrient loading assuming that they have a working knowledge of the relevant terms and processes.

In general nutrients are modelled in the CRAFT by either a constant concentration assigned to each flow pathway or by using an uptake factor (or “rating curve”) approach (e.g. Cassidy and Jordan (2011); Krueger et al., (2009)), where the concentration is directly proportional to the overland flow rate (Eq. (10)). A conceptual model of the flow and transport pathways in the catchment that are incorporated in the CRAFT is shown in the lower part of Fig. 4.

In the uptake factor approach, the concentration vector (units mg L^{-1}) ~~of~~ ~~a~~ of different nutrients (**n**) in overland flow (COF) is given by

$$\mathrm{COF}(\mathbf{n}) = \mathrm{MAX}(\mathrm{K}(\mathbf{n}) \cdot \mathrm{QOF}, \mathrm{COFMIN}(\mathbf{n})) \quad (10)$$

Where: QOF is the overland flow; K(**n**) represents the slope of the relationship between flow and nutrient (**n**) concentration in the observed data (i.e. uptake factor) and COFMIN(**n**) is the minimum concentration. This is included in Eq. (10) to prevent unrealistically low concentrations being used in the model during low flow periods, i.e. below the measurable limit. Krueger et al. (2009) used this type of equation to model TP concentrations in high flows generated by enrichment of sediment with P.

The daily nutrient load is calculated by the mixing model described by Eq. (11), where L(~~N~~**n**) is the vector of the nutrient loads (NO_3, SRP and TP, denoted by **n**)~~load~~, CSS and CGW are the constant concentrations ~~s~~in the dynamic soil and dynamic groundwater zones respectively

$$\mathrm{L}(\mathbf{n}) = \mathrm{COF}(\mathbf{n}) \cdot \mathrm{QOF} + \mathrm{CSS}(\mathbf{n}) \cdot \mathrm{QSS} + \mathrm{CGW}(\mathbf{n}) \cdot \mathrm{QGW} \quad (11)$$

The concentration vector of the nutrients in the catchment outflow (C(**n**)) can be calculated directly from the vector L(**n**) using Eq. (12)

$$\mathrm{C}(\mathbf{n}) = \mathrm{L}(\mathbf{n}) / \mathrm{QMOD} \quad (12)$$

Nitrate and SRP concentrations are calculated at each timestep using Eqs. (11) and (12). The TP concentration is calculated by Eq. (13)

$$\mathrm{C(TP)} = \frac{\mathrm{L(SRP)} + \mathrm{L(PP)}}{\mathrm{QMOD}} \quad (13)$$

CRAFT can thus capture the mixing effects of N and P losses associated with several hydrological flow pathways at the meso-scale. The above equations that remain in the MIR for CRAFT do not contain:-

Formatted: Highlight

Formatted: Not Highlight

Commented [RA23]: Note to reviewers, vector notation introduced for the nutrients. Please check my terminology is correct?

Formatted: Not Highlight

Commented [RA24]: Eq 10 modified in response to Reviewers and explanation improved

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Justified

Commented [RA25]: Note to typesetters, please do not use outer brackets on the numerator unless they are needed to comply with the rules? Note to R3, we changed the formatting of Eq 13 as requested

Formatted: Not Highlight

i) The myriad of nutrient cycling processes occurring in the N and P cycles. Section 2.1.2 shows the observable processes at the catchment outlet and Figure 3 the nutrient apportionment at this scale. However, the MIR captures the integrated effect of the processes and how these might change over time.

ii) Riparian processes are not explicitly included in the model. However, it is argued the impact of these processes is not observable at the outlet. The net effect of riparian processes are integrated into the soil and groundwater concentration values.

iii) Within channel processes such as plant uptake and the bioavailability of nutrient from bed sediments. Again, the impacts of these processes are not identifiable in the HFD time series. Unless the evidence of impact is clear they are not included in the MIR process.

Commented [RA26]: New section justifying the choice of model processes and summarizing the equations above it.

Formatted: Not Highlight

2.3 Modelling and Calibration

Flow and nutrients were simulated with the CRAFT for a ten year baseline period, 1 January 1997 to 31 December 2006 using a daily timestep. ~~The model parameters were assumed to be constant over space and time.~~ A comparison of the model performance at predicting the SRP and TP concentrations was curtailed at the end of February 2002. However, for nitrate the model performance over the full 10 yr period was assessed. ~~The daily timestep was used in the CRAFT for reasons discussed above.~~

The performance of the calibrated CRAFT model at reproducing observed stream flow at the catchment outlet~~s~~ was assessed by a combination of visual inspection of the modelled against observed runoff and the use of the Nash-Sutcliffe Efficiency (NSE) evaluation metric. The hydrological model calibration aimed originally to maximise the value of the NSE whilst ensuring that the MBE (mass balance error) was less than 10%. ~~The visual comparison was necessary to retain the overland flow process in the final, calibrated model (discussed in Section 3 "Results" below).~~ The parameters KSURF, KGW, KSS, SPLIT, SRMAX and SDMAX were adjusted iteratively to enable this and ~~create~~ obtain a single "expert" parameter set ~~and~~for the ~~a~~ baseline simulation (values shown in Table 3). The calibration strategy involved firstly obtaining an acceptable simulation of overland flow. In order of process representation: KSURF and SDMAX control the generation of overland flow (SDMAX must be adjusted to less than the maximum rainfall rate to initiate overland flow, and then KSURF controls the flow volume); SPLIT is then used to proportion recharge to the two subsurface stores; SRMAX controls the timing and volume of recharge events; and finally KGW and KSS are adjusted to reproduce the observed recession curves in the hydrographs (KSS being the more sensitive of the two). The sensitivity of the model was then assessed by running a Monte Carlo analysis of 100000 simulations, where the six parameters were randomly sampled from a uniform distribution (the upper and lower bounds are

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Commented [RA27]: Additional information highlighted in green above was added to describe the calibration procedure in response to Reviewers shown in Table ~~2~~3). ~~The performance metric used to compute a likelihood function (Beven, 2009); the Sum of Square of Errors (SSE) was chosen here, in order to identify which simulations were "behavioural".~~

Simulations with a MBE greater than 10% were ~~also~~ rejected. The top 1% of simulations meeting both criteria were thus chosen as "behavioural" and a normalised likelihood function ($L(Q)_i$) was calculated using Eq. (14) with the SSE values determined above for each simulation i.

$$L(Q)_i = \mathrm{SSE}_i / \sum \mathrm{SSE} \qquad (14)$$

Lastly, weights were assigned to the behavioural flows based on the likelihood of each simulation. These weighted flows were then used to compute the upper and lower bounds (here the 5th and 95th percentile flows were chosen) applied to the modelled flows (QMOD).

The NSE metric is suitable for assessing flow simulation performance but is less suitable for nutrient concentrations due to the occurrence of~~Dean et al. (2009) found that the NSE metric when used to assess the performance of the INCA-P model usually resulted in~~ negative NSE values, partly as a result of calculating variance terms using sparse observed data (where the sample mean is unlikely to reflect the true mean). Therefore, the nutrient model parameters were calibrated by assessing the performance of the model against the weekly concentration data in the LTD, using the following metrics to determine an "expert" parameter set:

- Visually comparing the time series of nitrate, SRP and TP against the observed data and adjusting the ~~most sensitive~~ nutrient model parameters to obtain a best fit between modelled and observed time series.

- Optimising the errors between modelled and observed mean and 90th percentile concentrations with the aim of reducing these below 10% if possible. The mean and 90th percentile concentrations were chosen as these represent the concentrations over the range of flows (mean) and events (90th percentile), and therefore allow the model performance under all flow regimes to be assessed. ~~This should be carried out alongside the previous step.~~

~~If satisfactory nutrient model outputs were not obtained by adjusting the nutrient parameters in the first step then it was necessary to adjust the hydrology model parameters, particularly KSURF and SPLIT, to increase or decrease the proportions of the different flow pathways,. KSURF~~

A further sensitivity analysis was then performed using the flows from the behavioural hydrology simulations (discussed above) and re-running the nutrient model (without adjusting the "expert"

Formatted: Not Highlight

Commented [RA28]: Here we are defining what is "acceptable" in terms of WQ model performance

Formatted: Not Highlight

parameter values for the nutrients) to determine a set of upper and lower bounds (5th and 95th percentile values) to the predicted concentrations and their associated loads ($Q \cdot C$).

2.4 Management Intervention Scenario

For a model to be effective at the management level it needs to be to demonstrate the impacts of changes in local scale in land managementable to link back to processes at the local scale. The creators of the model are thus conveying their key findings to catchment managers to inform them of the consequences of local scale changes at the catchment scale. Here the local land use change is assumed to occur at all locations. Nevertheless, the CRAFT model can show the magnitude and proportion of the nutrients lost by each hydrological flow pathway. Equally it is possible to show the concentration of each nutrient at each time step as this helps educate the end user. However, for simplicity, here a combination of land use changes and express the output as the change in export loads for each pathway at the outlet will be shown.

In order to demonstrate the impact of a catchment management intervention strategy, the following changes were made to the catchment as a runoff and nutrient management intervention (MI) scenario. For simplicity a combination of land use changes were applied and the output expressed as the changes in export loads for each pathway at the outlet, shown below:

(i) : (i) The modelled overland flow was reduced by reducing the value of the KSURF parameter to 0.012, representing a management intervention that removes or disconnects the agricultural pollution "hotspots"

(i) ; (ii) Nutrient loads in the rapid subsurface zone were reduced by reducing the values of CSS(SRP) and CSS(NO_3) by 50% (i.e. halving the impact of diffuse sources linked to the outlet by this flow pathway) to represent improved land management with reduced fertilizer loads. No change to the DG nitrate concentration was made as firstly, any changes in land management may take decades to be observed in the deeper groundwater (Smith et al., 2010); and secondly, recent improvements to WWTPs have only targeted reducing SRP loads and not nitrate loads (Bowes et al, 2009b, 2011).

Commented [RA29]: Note to R3 that nitrate loads were reduced in the catchment under the MI Scenario

(ii)

(ii)(iii) (iii) Background loads of SRP in the catchment are reduced by lowering CGW(SRP) to represent the reduction in deeper groundwater concentration caused by both lower leaching rates from the soil store and making further improvements to WWTPs in the catchment to reduce SRP loads. Bowes et al. (2009b) found that a 52% reduction in the SRP export from point sources had taken place since 2001 in the catchment (up to 70% of the SRP loads from each improved WWTP is assumed to be stripped out). In terms of the total (point and diffuse)

SRP load, Bowes et al. (2011) estimated that between 2000 and mid 2009 it had been reduced by 58%, which was due to further improvements to the smaller WWTPs in the catchment as well as a reduction in diffuse sources of up to 0.1 kg P $ha^{-1}yr^{-1}$. Figure 3 shows that point sources (in 2005-6) were thus estimated to contribute 16% of the annual TP load.

Formatted: Not Highlight

3 Results

~~Essentially we can compare the modelled and observed data sets and the core statistics (Table 1) or by visually assessing model performance firstly from the "expert" calibration.~~The baseline model results are shown in Fig. 5 ~~shows as~~the time series plots of modelled and observed flow at East Stoke along with the modelled and observed ~~("expert" calibration) and observed~~ nitrate, TP and SRP concentrations for a selected two year period. The years chosen have average followed by wet hydrological conditions. To further illustrate the model performance ~~in terms of~~at predicting flow and concentrations, the upper panes in Fig. 5 ~~–~~show a corresponding timeseries plot of the ~~errors~~absolute error (i.e. Observed flow or concentration – Modelled flow or concentration).

Formatted: Not Highlight

Formatted: Not Highlight

3.1 ~~Expert Calibration~~Baseline Simulation

The hydrology model parameters ~~from~~ used by the ~~final "expert" calibration~~baseline simulation are shown in Table ~~2~~3. ~~Hence we are suggesting that the user has a level of knowledge and experience in nutrient inputs and outputs.~~ The model results from the CRAFT were as follows: The NSE for the baseline hydrology simulation was 0.80; the mass balance error was over predicted by ~~+~~1.0%~~% (over prediction), less than the 10% limit that is considered acceptable for assessing the model performance as "satisfactory"~~. In the Frome catchment the percentage of overland flow (which includes surface runoff and near-surface runoff through the ploughed layer) according to the calibrated model was very small (2.2 % of the annual total runoff of 516 mm yr^{-1}). This value may be low but as stressed before it is difficult to see the overland flow signal at the meso-scale. ~~.~~ Here, an overland flow component has been retained (by setting KSURF and KSR to the values shown in Tables 3 and 4) due to an assumption that P is being lost via this process i.e. from the knowledge arising from research studies (e.g. Owen et al., 2012; Bowes et al. 2009a; Heathwaite et al., 2005). Values for the parameters KSR(PP) and KSR(SRP) were ~~set~~ determined in the ~~"expert" calibration~~baseline simulation based on some events (as suggested in figure 2 and 3) where ~~both~~ runoff ~~and~~ driven TP spikes were observed. ~~Such spikes were also observed in the HFD dataset and classified as "Type 4" events (Ref Section 2.1), although unfortunately the modelled period (for TP and SRP) did not overlap with the HFD monitoring period.~~

Commented [RA30]: R3 suggested a more complicated calibration technique using P in overland flow.

3.2 Runoff

It is possible ~~of course~~ to optimise the ~~model~~ parameter values in the model~~s~~ to generate either a smaller mass balance error or a larger value of the NSE metric (over 0.8 is possible with this model and data, as evidenced by the Monte Carlo simulation results). Here a compromise was sought between both these metrics, and to in terms retainretaining the overland flow process (discussed above) and a good visual fit with the observed flows.

The behavioural flows from the Monte Carlo simulation are shown in Fig. 6 as dotted lines representing the upper (95th percentile) and lower (5th percentiles) prediction bounds. There were 511 simulations classed as "behavioural". The envelope of the predicted flows indicates that most of the observed flows during the ten year period of data could be reproduced, supporting the choice of runoff processes represented in the CRAFT for this particular catchment. Some events may have been either missed or over predicted which could be due to limitations with using a single rain gauge in the forcing data for the model. Table 65 shows the minimum, median and maximum flows extracted from these timeseries. The table shows that -the model outputs are sensitive to the- parameter valuesparameters and the end user needs to retain this fact. .

3.3 Nutrients

3.3.1 Nitrate

The HFD observed nitrate concentrations in Fig. 2b indicated that concentrations of nitrate in overland flow are much smaller than concentrations in baseflow, and the model parameter COFMIN(NO_3) (see Eq. 10) was set to 0.4 mgL^{-1} N (Table 4). *In the baseline scenario* the proportion of nitrate loads generated by overland flow was thus fairly negligible (<1%) and the nitrate loads were split fairly evenly between the SS and DG pathways according to the model. The load from the DG contributed around 31% of the total load, compared to 43% of the modelled runoff originating from this pathway. This implies that a significant proportion of nitrate drains from the shallow subsurface (SS) immediately after storm events, probably through either enhanced connectivity due to agricultural drains or recharge into the underlying chalk aquifer (Bowes et al., 2005). The DG component includes nitrate loads from the WWTPs in the catchment which were estimated to contribute around 7% (1.5 kg N ha^{-1} yr^{-1}) of the total load based on monitoring data from the mid-1980s (Casey et al., 1993), and 14% of the modelled DG load..

In terms of the sensitivity of the nitrate results to the flow model parameters, SPLIT was important since it controlled the proportions of the slow and fast nitrate in the total runoff. Overall, the CRAFT model reproduced a moving average of the observed nitrate LTD concentrations reasonably well and mean concentrations were within 10% of the observed (Table 45). The fit between modelled and observed nitrate in terms of absolute errors (Fig. 54b lowerupper pane) was not so good probably due to timing errors in predicting the onset of dilution, although visually (Fig. 54b upperlower pane) the model appeared to model simulate the seasonal patterns of nitrate fairly well. Table 5 6 shows the

Commented [RA31]: This ties in to an earlier comment to R3 about model processes, here KSURF and KSR may not be set to their optimum (NSE) value but to a value that generates QOF (in Response)

Commented [RA32]: R3: We stated in our Response that the model will generate runoff when sufficient rain falls, if the rain is reported as zero there is therefore no runoff, so an event could be missed!

Formatted: Not Highlight

Commented [RA33]: State in R3 response that the nitrate "observed" WWTP load is assumed to be the same as Casey's value from the mid-1980s and is included in DG

Formatted: Not Highlight

Commented [RA34]: We modelled the pattern fairly well but the timing errors resulted in model errors in terms of Cs

Formatted: Not Highlight

Formatted: Not Highlight

uncertainty in nitrate loss arising from the hydrological model in terms of the 5th, 95th percentiles and medians of modelled concentrations and yields~~.~~

3.3.2 Phosphorus

Bowes et al. (2009b) estimated that between 1991 and 2003, SRP provided 65% of the TP load in the Frome catchment. In the *baseline scenario*, the DG component in the model generated almost four times the load of SRP than the SS component (Fig. 7). This seems plausible as the DG component also included the SRP loads from the WWTPs, in addition to the SRP originating from springs and seeps from shallow groundwater. Again, the SPLIT parameter in the flow model had a large influence on SRP loads, by adjusting the ratio between the SS and DG components of these. The model errors, identifiable from the panels above the timeseries plots (Fig. 5) may have been caused by timing issues leading to periods of overprediction and underprediction of SRP concentrations. Visually, the SRP concentrations ~~were fitted well~~showed a close match, ~~using on average~~ and the seasonal patterns and trends were simulated (Fig 5c). Any spikes in the observed data which were not reproduced by the model appear not to have been caused by actual hydrological runoff events (as seen in Fig. 2 and discussed above). Modelled concentrations (on sample days only) were within 10% of the observed SRP concentrations for both the mean and 90th percentile values but underpredicted the mean and 90th percentile TP concentrations by around 50% (Table 5). This may be due to additional source(s) of P not being accounted for in the model (e.g. within-channel river ~~channel~~ dynamics and/or conversion of SRP to entrained particulate forms of P as suggested by Bowes et al. (2009a)). Table 6 shows the uncertainty in the TP and SRP losses arising from the hydrological model in terms of the 5th, 95th percentiles and medians of modelled concentrations and yields.

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

These results however showed that high concentrations of TP associated with the transport of PP during runoff events were predicted by the Monte-Carlo and expert simulations (over 1.9 mg/L P), which was similar to the "Type 2" events identified in the HFD dataset where TP concentrations reached 1.75 mg/L P in late 2005. The LTD dataset did not contain many spikes of this magnitude in the TP concentrations, however the HFD data did measure occasional high concentrations of TP associated with runoff events (e.g. those indicated by a "4" on Fig. 2c). Figure 2c, and the model results in Fig. 5, show that the issue of fitting TP at the meso-scale is problematical and is unlikely to be improved by having a more complex model ~~.~~

Formatted: Not Highlight

Formatted: Not Highlight

In the baseline scenario the modelled proportion of TP (i.e. PP) generated by overland flow was about 11% which ~~was~~ is quite high considering that only 1.2% of the modelled runoff ~~was~~ is generated via this pathway. ~~However, this was only half of the percentage event load estimated from the HFD data in~~

Formatted: Not Highlight

2005-6 (probably due to additional sources of P being included in this figure). The PP concentrations generated by the model were calibrated by adjusting the value of the KSR(PP) parameter (Table 34).

We also calculated theThe export yields (load per unit area) for each nutrient to show the impact of the flow pathways at transporting nutrients were also calculated (see Fig. 7 and Table 56). This aggregation lends itself to comparisons with previous studies. The baseline simulation predicted a TP export of 0.69 kg P $ha^{-1}yr^{-1}$ which is slightly more than both the export rate estimated by Hanrahan et.al (2001) for diffuse and point sources in the catchment of 0.62 kg P $ha^{-1}yr^{-1}$ (for calendar year 1998). SRP loads were modelled by Bowes et al. (2009b) and the SRP export was predicted to be 0.44 kg P $ha^{-1}yr^{-1}$ between 1996-2000 (of which WWTP discharges accounted for 49%), compared to the CRAFT modelled baseline SRP export of 0.62 kg P $ha^{-1}yr^{-1}$ (between 1997 and February 2002). Similar historical estimates for nitrate export were not available, to compare with the model estimate of 32.8 kg N $ha^{-1}yr^{-1}$ over the period 1996-2005, except a single year from the HFD dataset where the TON export was estimated to be 20.2 kg N $ha^{-1}yr^{-1}$ (Bowes et al. (2009a)). Table 5 6 shows the uncertainty in terms of the 5th, 95th percentiles and medians of modelled concentrations and yields.

3.4 Management Intervention (MI) Scenario

The yields of nitrate and TP are summarised by the use of bar charts in Fig. 7 which illustrate the fluxes under the baseline conditions (left bars) and the MI scenario (right bar), and the relative contribution of each of the three flow pathways to these, which provides valuable source apportionment information for policy makers.

Formatted: Highlight

The results show that the amount of PP generated by the overland flow pathway (denoted by the blue rectangle in the baseline scenario bar in Fig. 7) has reduced to almost zero due to the reduction in overland flow, and the difference between TP and SRP export is negligible as a result. This indicates that a limited amount of "pollution swapping" is predicted so that the proportions of PP and SRP comprising TP have changed from 8.8% and 92.2% to 0% and 100% respectively under the MI scenario. Nitrate and TP loads are predicted to decrease by 34.4% and 65.0% respectively. Under the MI scenario, the nitrate concentration in the DG flow component (which includes point sources) was not reduced (it was assumed that WWTP improvements targeted P and not N). Both nitrate and SRP loads in overland flow were negligible (< 0.1%) under the baseline scenario and have been reduced to effectively zero by drastically reducing the amount of overland flow generated. SRP loads due to point sources are included in the DG component, the predicted load from this component reduced by 63%. The export of SRP via the faster SS component also reduced by 55% (to 0.045 kg P $ha^{-1}yr^{-1}$) under the MI scenario. These reductions in the SRP loads from different components compare well to the overall reductions since the 1990s in point and diffuse sources in the catchment (Bowes et al., 2009b, 2011). Clearly the need for the end user to understand and interpret these phenomena is a pre-requisite for the CRAFT model to have meaning. The build-up of knowledge and experience should ideally already exist and form part of an improved understanding of catchment management (Cuttle et al., 2007, DEFRA, 2014).

4 Discussion and Conclusions

This paper has attempted to exploreexplored the role of scale appropriateMIR modelling methods at the meso-scale. Specifically, iIt has explored the information content of flow and nutrient data within a case study, that helps justify the choice of model structure and timestep. The MIR approach to modelling is thus the minimal parametric representation to model phenomena at the meso-scale as a means to aid catchment planning/decision making at that scale. The approach is based on either a simplification of a more complex model or is based on observations made in research studies in the Frome catchment. The MIR model that was developed, CRAFT, thus focussed on key hydrological flow pathways which are observed at the hillslope scale. The nutrient components were kept very simple ignoring neglecting all nutrient cycling aspects. The astute choice of a daily timestep also reduced the burden to route flows through the system. The CRAFT model deliberately avoids a spatial representation of local land use in this particlarparticular case study. This implies that the lumping process is appropriate for circumstances where the local variability disappears is lost when aggregated. The model can be used in a semi-distributed form if obviousthe land use patterns justify thesuch a new model structure and itthis form may help to justifyidentify the sources of the fluxes in the overall model for some applications. Future developments of the CRAFT will also permit the investigation of many features such as riparian fluxes and also the impact of attenuation on sediments and nutrient fluxes when routed through ponds and wetlands.

High resolution frequency data (such as the HFD) for all nutrient parameters is desirable at all scales locations if it were affordable. However, it is shown here that at the meso-scale these data tends to reflect the "noise", incidental losses and within-channel diurnal cycling in the system that have a limited effect on the overall signal and loads hence a lower sampling may be suitable in at this scale. For the Frome case study a daily timestep in the CRAFT model could simulate the dominant seasonal and storm driven nutrient flux patterns and thus aid the user policy maker in considering a variety of policy decisions. It is stressed that collecting the longest possible high resolution frequency dataset particularly for all forms of nutrients is still of the utmost importance for effective water quality monitoring and identifying the full range of observed concentrations including incidental losses (see Fig 2c). There may be some evidence here that collecting higher resolution data for nutrients helps to explain the distribution values and addresses the issues of "noise" and diurnal variability (e.g. the fluctuations in P

Commented [pfq35]: This now matches the objective at the start

Formatted: Not Highlight

Commented [RA36]: I've made some edits to this paragraph: note to check it makes sense before submitting!!

Commented [pfq37]: We need to stress that our model can be semi-distributed if we want it to be

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

concentrations observed in the River Enborne by Wade et al., 2012 and Halliday et al., 2014) in the datasets. Even so, it may still be beneficial to aggregate sub-daily data to daily data ~~as a optimising~~as a means to optimise the capabilities of a process based model, such as the CRAFT, and ~~using~~ make use of all the ~~policy~~ relevant information actually contained in high frequency monitoring~~the HFD~~ data.

The Frome case study revealed a number of interesting factors, leading to the exploration of a management intervention (MI) scenario. The mean annual SRP concentration that has to be attained in order to comply with the WFD standards for P is 0.06 mgL^{-1} P, which was achieved by the MI scenario (modelled mean = 0.053 mgL^{-1} P) by reducing the ~~appropriate~~ SRP concentrations in the model's flow pathways to reduce the modelled SRP load by 61.7%. There are no explicitly defined guidelines for nitrate, except that the maximum concentration must not exceed 11.9 mgL^{-1} N, which is imposed on all surface waters in the EU under the terms of the 1991 Nitrates Directive. In terms of nitrate management in the Frome catchment, the observed data from 1997 to 2006 indicated that concentrations (at least in surface water) were below the limit without any reductions due to nutrient and/or runoff management. The CRAFT model was able to reproduce the seasonality in the observed nitrate concentrations and also make predictions of the likely reductions in concentrations and yields, due to improved management of diffuse sources in the catchment. This MI scenario reduced mean concentrations from 6 mgL^{-1} N to 4.3 mgL^{-1} N at the outlet of the Frome. Recent studies of long term trends (Smith et al., 2010; Bowes et al., 2011) showed that nitrate concentrations were observed to be rising in the Frome since the 1940s, however over the simulation period the rate of increase has slowed down and the CRAFT model could predict the weekly time series reasonably well as a result. The MI scenario shows that interventions to reduce concentrations of nitrate in rapid subsurface flow can have a significant impact at reducing the total nitrate load by 34% although this may occur at the expense of pollution swapping leading to increased nitrate fluxes to deep groundwater. Interventions to reduce the concentration of nitrate in flows originating from deeper groundwater were not investigated as these improvements could take decades to be observable at the monitoring point at the catchment outlet (Smith et al, 2010).

The results of this case study may best be viewed as event driven export coefficients when the origin of the nutrient is tied to the pathway that generated it. This informs the ~~end~~ user as to the aggregate effect of local policy changes and the importance of storm size and frequency. Whilst we have shown that those impacts are still uncertain it could perhaps encourage more intervention in order to guarantee the success of new policy (Cuttle et al., 2007). Equally, locally observed environmental problems caused by high nutrient concentrations may well be lost due to mixing effect at the meso-scale (i.e. catchment outlet).

Formatted: Not Highlight

Commented [RA38]: In response to R3 (earlier revision)

Formatted: Not Highlight

Formatted: Not Highlight

The sensitivity and uncertainty analysis carried out on the hydrological model showed the impact on the resultant nutrient fluxes. The CRAFT model is intended to be just one of many required for setting policy at the meso-scale. Equally, despite the uncertainty in the model, the outputs should encourage the user in that a range of local scale polices can have a large impact on the final nutrient flux at the meso-scale. Clearly the need for the end user to understand and interpret these phenomena is a pre-requisite for the CRAFT model to have meaning. The build-up of knowledge and experience should ideally already exist and form part of an improved understanding of catchment management (Cuttle et al., 2007, DEFRA, 2014)

Commented [pfq39]: Moved a paragraph to new location below

The CRAFT model has been shown to fit the dominant seasonal and event driven phenomena. This The benefits of using the CRAFT are thus firstly that it is a useful tool which conveys the mixed effect of land use and hydrological process at the meso-scale for policy makers. The modelling process assumes that the policy maker or informed end user will then manipulate the model to see the likely impacts of regulations. The burden is still on the user to translate policy into the likely local impact, for example: reduction in N and P loading; more efficient use of N and P in soils and the acute loss of P from well-connected flow pathways. Once the parameters are changed, the net effect at the meso- -scale can then be seen instantaneously. The user is encouraged to try many scenarios and to explore the parameters space. Secondly, its Excel interactive graphical user interface (hence having transparency) that allows an instantaneous view of the changes made to the model parameters, which in itself is educationalinformative. The range of the fluxes seen can inform the user about the uncertainty of the model when mtaking decisions and can alert them to unexpected outcomes such as pollution swapping.

The sensitivity and uncertainty analysis carried out on the hydrological model showed the impact on the resultant nutrient fluxes. The CRAFT model is intended to be just one of many required for setting policy at the meso-scale. Equally, despite the uncertainty in the model, the outputs should encourage the user in that a range of local scale polices can have a large impact on the final nutrient flux at the meso-scale. When used with other model tools and observed data the CRAFT meso-scale model can play a key role in evaluating land use change and the need to conform to WFD targets.

The sensitivity and uncertainty analysis carried out on the hydrological model show the impact on the resultant nutrient fluxes. The output of the does suggest that the 'expert' choice of a (hydrology and nutrients) model parameter set not unreasonable. The interactive nature of the tool allows the user to explore ideas and gain confidence in using the tool for scenario testing. This tool is intended to be just one of many required for setting policy at the meso-scale. Equally, despite the uncertainty in the model, the outputs should encourage the user that a range of local scale polices can have a large impact on the final nutrient flux at the meso-scale. The underlying message that lowering nutrient mobilisation risk flow connectivity and improving WWTPs are all beneficial at the meso-scale.

Nomenclature

CEH Centre for Ecology and Hydrology

CRAFT Catchment Runoff Attenuation Flux Tool

DTC Demonstration Test Catchments

DWC Dry Weather Concentration (i.e. in baseflow)

EMC Event Mean Concentration (i.e. in overland flow)

HFD High Frequency data set of nitrogen and phosphorus, recorded several times per day in the River Frome.

LTD Long term data set of weekly nitrogen and phosphorus measurements also in the River Frome, modelled by the baseline scenario.

MBE Mass balance error

MIR Minimum Information Required

n Vector of nutrients simulated by the model (e.g. N and P).

NSE Nash – Sutcliffe Efficiency (model performance metric)

PP Particulate phosphorus (i.e. the insoluble fraction)

SRP Soluble reactive phosphorus (from samples filtered using 0.45 µm paper)

TON Total oxidised nitrogen (nitrate + nitrite).

TP Total phosphorus (soluble + insoluble forms)

WFD Water Framework Directive

WWTP Wastewater Treatment Plant (Sewage Treatment Works)

Acknowledgements

The collection of both the long term and high ~~resolution~~ frequency nutrient datasets was funded by the Natural Environment Research Council.

References

Addiscott, T.M., Whitmore A.P.: Simulation of solute leaching in soils with different permeabilities. Soil Use and Management 7: 94–102 1991.

Argent, R.M., Perraud, J-M., Rahman, J.M., Grayson, R.B., and Podger, G.M.: A new approach to water quality modelling and environmental decision support systems. Env. Model. Software. 24:809-818, 2009.

Arnold, J. G.: SWAT (Soil and Water Assessment Tool). Grassland, Soil and Water Research Laboratory, USDA, Agricultural Research Service, Temple, TX. (USA). 1994.

Bartley, R., Speirs, W.-J., Ellis T.W., and Waters, D.K.: A review of sediment and nutrient concentration data from Australia for use in catchment water quality models. Marine pollution bulletin 65(4):101-116, 2012.

Beven, ,K.: Environmental modelling: an uncertain future?: An introduction to techniques for uncertainty estimation in environmental prediction. Routledge, Abingdon, Oxon (UK). 2009.

Blöschl, G.: Scale and scaling in hydrology. Wiener Mitteilungen, Wasser-Abwasser-Gewässer, 132, TU Wien, Wein, (Österreich). 1996.

Botter, G., Bertuzzo, E., and Rinaldo, A.: Catchment residence and travel time distributions: the master equation, Geophys. Res.Lett., 38, L11403, doi:10.1029/2011GL047666, 2011.

Bowes, M.J., Leach, D.V., and House, W.A.: Seasonal nutrient dynamics in a chalk stream: the River Frome, Dorset, UK. Sci. Tot. Env., 336(1-3): 225-241, 2005.

Bowes, M.J., Smith, J.T., and Neal, C.: The value of high-resolution nutrient monitoring: A case study of the River Frome, Dorset, UK. J. Hydrol., 378(1-2): 82-96, 2009a.

Bowes, M.J., Smith, J.T., Jarvie, H.P., Neal, C., and Barden, R.: Changes in point and diffuse source phosphorus inputs to the River Frome (Dorset, UK) from 1966 to 2006. Sci. Tot. Env, 407(6): 1954-1966, 2009b.

Bowes, M.J., Smith, J.T, Neal, C., Leach, D.V., Scarlett, P.M., Wickham, H.D, Harman, S.A., Armstrong, L.K., Davy-Bowker, J., Haft, M., and Davies, H.N.: Changes in water quality of the River Frome (UK) from 1965 to 2009: Is phosphorus mitigation finally working? Sci. Tot. Env, 409(18): 3418-3430, 2011.

Casey, H.: Variation in chemical composition of River Frome, England, from 1965 to 1972. Freshwater Biol., 5(6): 507-514, 1975.

Casey, H., Clarke, R.T., and Smith, S.M.: Increases in nitrate concentrations in the River Frome (Dorset) catchment related to changes in land use, fertiliser applications and sewage input. Chem. Ecol., 8: 105-117, 1993.

Cassidy, R. and Jordan, P.: Limitations of instantaneous water quality sampling in surface-water catchments: Comparison with near-continuous phosphorus time-series data. J. Hydrol., 405: 182-193, 2011.

Cuttle, S. P., C. J. A. Macleod, D. R. Chadwick, D. Scholefield, P. M. Haygarth, P. Newell-Price, D. Harris, M. A. Shepherd, B. J. Chambers, and R. Humphrey.: An inventory of methods to control diffuse water pollution from agriculture (DWPA). User Manual (DEFRA Project ES0203), UK, 113pp, 2007.

Davison, P.S, Withers, P.J.A., Lord, E.I., Betson, M.J. and Strömqvist, J.: PSYCHIC - A process-based model of phosphorus and sediment mobilisation and delivery within agricultural catchments. Part 1: Model description and parameterisation. J. Hyd. 350: 290-302, 2008.

Deasy, C., Brazier, R.E., Heathwaite, A.L., and Hodgkinson, R.: Pathways of runoff and sediment transfer in small agricultural catchments. Hydrological Processes, 23: 1349-1358. doi:10.1002/hyp.7257, 2009.

Formatted: Font: Italic

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Dean, S., Freer, J., Beven, K., Wade, A.J., and Butterfield, D.: Uncertainty assessment of a process-based integrated catchment model of phosphorus Stoch. Environ. Res. Risk. Assess. 23:991-1010, 2009.

Defra, Catchment Based Approach: Improving the Quality of our Water Environment https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/204231/pb13934-water-environment-catchment-based-approach.pdf, 2015.

Edwards, A.C., Kay, D., McDonald, A.T., Francis, C., Watkins, J., Wilkinson, J.R. and Wyer, M.D.: Farmyards, an overlooked source for highly contaminated runoff. J. Enviro. Manag., 87: 551-559, 2008.

Environment Agency (EA): Making Information Available for Integrated Catchment Management Science Report - SC060035/SR2007. Bristol (, UK).114pp, 2007.

European Parliament: Common Implementation strategy for the water framework directive (2000/60/EC). Guidance document no. 5. Transitional and coastal waters: typology, reference conditions and classification systems. Brussels (Belgium), 2000.

Evans D.J., and Johnes P.: Physico-chemical controls on phosphorus cycling in two lowland streams. Part 1 - the water column. Science of the Total Environment; 329: 145-163, 2004.

Fraser, C. E., McIntyre, N., Jackson, B.M., and Wheater, H.S.: Upscaling hydrological processes and land management change impacts using a metamodeling procedure, Wat. Resour. Res., 49: 5817–5833, doi:10.1002/wrcr.20432. 2013.

Halliday S., Skeffington R, Bowes M, Gozzard E, Newman J, Loewenthal M, Palmer-Felgate, E. J., Jarvie, H. P., and Wade A. J.et al.: The Water Quality of the River Enborne, UK: Observations from High-Frequency Monitoring in a Rural, Lowland River System. Water; 6: 150-180, 2014.

Hanrahan, G., Gledhill, M., House, W.A., and Worsfold, P.J.: Phosphorus loading in the Frome Catchment, UK: Seasonal Refinement of the coefficient modelling approach, J. Env. Qual. 30(5), 1738-1746, 2001.

Haygarth, P.M., Wood, F.L., Heathwaite, A.L., and Butler, P.J.: Phosphorus dynamics observed through increasing scales in a nested headwater-to-river channel study. Sci. Tot. Env, 344: 83-106, 2005.

Haygarth, P. M., Page, T. J. C., Beven, K. J., Freer, J., Joynes, A., Butler, P., and Owens, P. N: Scaling up the phosphorus signal from soil hillslopes to headwater catchments. Freshwater Biology, 57(s1), 7-25, 2012.

Heathwaite, A. L., Fraser, A. I., Johnes, P. J., Hutchins, M., Lord, E., and Butterfield, D. : The phosphorus indicators tool: A simple model of diffuse P loss from agricultural land to water. Soil Use and Management, 19(1):1-11, 2003.

Heathwaite, A.L., Quinn, P.F., and Hewett, C.J.M.: Modelling and managing critical source areas of diffuse pollution from agricultural land using flow connectivity simulation. J. Hydrol., 304: 446-461, 2005.

Hracowitz, M., Savenije, H., Bogaard, T.A., Tetzlaff, D., and Soulsby, C.: What can flux tracking teach us about water age distribution patterns and their temporal dynamics? Hydrol. Earth Syst. Sci., 17, 533-564,2013. doi:10.5194/hess-17-533-2013.

Johnes, P.J.: Evaluation and management of the impact of land use change on the nitrogen and phosphorus load delivered to surface waters: the export coefficient modelling approach, J.Hydrol. 183(3):323-349, 1996.

Johnes, P.J.: Uncertainties in annual riverine phosphorus load estimation: Impact of load estimation methodology, sampling frequency, baseflow index and catchment population density. J. Hydrol., 332(1-2): 241-258, 2007.

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Field Code Changed

Formatted: Not Highlight

Jordan, P., Arnscheidt, A., McGrogan, H., and McCormick, S.: Characterising phosphorus transfers in rural catchments using a continuous bank-side analyser. Hydrol. Earth Syst. Sci., 11(2): 372-381, 2007. doi:10.5194/hess-11-372-2007

Kirchner, J.W., and Neal, C.: Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection, PNAS, 110, 12213-12218, 2013.

Krueger, T., Quinton, J. N., Freer, J., Macleod, C. J., Bilotta, G. S., Brazier, R. E., Butler, P., and Haygarth, P. M.: Uncertainties in data and models to describe event dynamics of agricultural sediment and phosphorus transfer. Journal of Environmental Quality, 38(3), 1137-1148, 2009.

Marsh, T. J. and Hannaford, J. (Eds.): UK Hydrometric Register, Hydrological data UK series, Centre for Ecology and Hydrology, Wallingford (UK), 210 pp., 2008

McIntyre, N, Jackson, B, Wade, AJ, Butterfield, D. and Wheater, H.S.: Sensitivity analysis of a catchment-scale nitrogen model. J. Hydrol. 315(1-4):71-92, 2005.

Owen, G. J., Perks, M. T., Benskin, C. M. H., Wilkinson, M. E., Jonczyk, J., and Quinn, P. F.: Monitoring agricultural diffuse pollution through a dense monitoring network in the River Eden Demonstration Test Catchment, Cumbria, UK. Area, 44(4), 443-453, 2012.

Quinn, P.F., Anthony, S. and Lord E.: Basin scale nitrate modelling using a minimum information requirement approach. In Water Quality: Processes and Policy, Trudgill S. Walling D, Webb B (eds). Wiley: Chichester, UK; 101-117, 1999.

Quinn, P.: Scale appropriate modelling: representing cause and effect relationships in nitrate pollution at the catchment scale for the purpose of catchment scale planning. J. Hydrol. 291: 197-217, 2004.

Quinn, P., Hewitt, C.J.M. and Dayawansa, N.D.K.: TOPCAT-NP: a minimum information requirement model for simulation of flow and nutrient transport from agricultural systems, Hydrol. Process. 22(14):2565-2580, 2008.

Sharpley, A.N., Kleinman, P.J.A., Heathwaite, A.L., Gburek, W.J., Folmar, G.J., Schmidt, J.P.: Phosphorus loss from an agricultural watershed as a function of storm size. J. Environ. Qual. 37, 362 – 368. http://dx.doi.org/10.2134/jeq2007.0366. 2008

Smith, J.T., Clarke, R.T. and Bowes, M.J.: Are groundwater nitrate concentrations reaching a turning point in some chalk aquifers? Sci. Tot. Env., 408(20): 4722-32, 2010.

Stephens, C., and Quinton, J.: Diffuse pollution swapping in arable agricultural systems. Critical Reviews in Environmental Science and Technology, 39(6): 478-520, 2009.

Storr, E., Adams, R. and Western, A.: How can data from headwater catchments be used to improve runoff and nutrient predictions at larger scales? In Chan, F., Marinova, D. and Anderssen, R.S.(eds) MODSIM2011, 19th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand:1652-1658, 2011.

Uhlenbrook, S., Roser, S., and Tilch, N.: Hydrological process representation at the meso-scale: the potential of a distributed, conceptual catchment model. Journal of Hydrology 291.3: 278-296, 2004.

Van der Velde, Y., de Rooij, G. H., Rozemeijer, J. C., van Geer, F. C., and Broers, H. P.: The nitrate response of a lowland catchment: on the relation between stream concentration and travel time distribution dynamics, Water Resour. Res., 46, W11534, 2010. doi:10.1029/2010WR009105

Formatted: Not Highlight

Formatted: Not Highlight

Wade, A.J., Whitehead, P.G. and Butterfield, D.: The Integrated Catchments model of Phosphorus dynamics (INCA-P), a new approach for multiple source assessment in heterogeneous river systems: model structure and equations. Hydrol Earth Syst Sci, 6(3): 583-606, doi:10.5194/hess-6-583-2002, 2002.

Wade, A.J., Butterfield, D., and Whitehead, P.G.: Towards an improved understanding of the nitrate dynamics in lowland, permeable river-systems: applications of INCA-N, J. Hydrol., 330,185-203, doi:10.1016/j.jhydrol.2006.04.023, 2006.

Wade, A. J., Palmer-Felgate, E. J., Halliday, S. J., Skeffington, R. A., Loewenthal, M., Jarvie, H. P., Bowes, M. J., Greenway, G. M., Haswell, S. J., Bell, I. M., Joly, E., Fallatah, A., Neal, C., Williams, R. J., Gozzard, E., and Newman, J. R.: Hydrochemical processes in lowland rivers: insights from in situ, high-resolution monitoring. Hydrol. Earth Syst. Sci.16(11): 4323-42, 2012. doi:10.5194/hess-16-4323-2012

Wheater, H. S., and Peach, D.: Developing interdisciplinary science for integrated catchment management: the UK lowland catchment research (LOCAR) programme, Int. J Wat. Res. Develop. 20(3): 369-385, 2004.

Whitehead, P. G., Crossman, J., Balana, B.B., Futter, M. N., Comber, S., Jin, L., Skuras, D., Wade, A.J., Bowes, M. J., and Read, D. S.: A cost-effectiveness analysis of water security and water quality: impacts of climate and land-use change on the River Thames system. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371, no. 2002: 20120413, doi:10.1098/rsta.2012.0413, 2013.

Williams, J.R.: The EPIC model. In Computer Models of Watershed Hydrology, Singh VP (ed.). Water Resources Publications: Highlands Ranch, CO (USA); 909–1000. 1995.

Withers, P.J.A. and Lord, E.L.: Agricultural nutrient inputs to rivers and groundwaters in the UK: Policy environmental management and research needs. Soil Use Manage. 14, 186-192. 2002.

Formatted: Not Highlight

Tables

Table 01. Attributes of Frome Water Quality monitoring datasets

Dataset	Time Period	Sampling Frequency	Average Number of Observations /Year	Measurements
Long Term Dataset (LTD) CEH/Freshwater Biological Association (Bowes et al., 2011)	1965-2009	Weekly	48	TP,TDP, Nitrate, SRP
High frequency data set (HFD) Bowes et al. (2009a)	1/2/2005 to 31/1/2006	Sub-daily	>1000 (see Table *1 for actual total)	TP,TON, SRP, TSS, instantaneous flows

Table 12. Long term nutrient concentration statistics in the LTD and HFD datasets

Dataset/Nutrient (time period)	Number of Observations	10th Percentile Concentration (mgL^{-1})	Mean Concentration (mgL^{-1})	90th Percentile Concentration (mgL^{-1})
LTD Nitrate (7/1/97-21/11/06)	384	4.6	5.6	6.9
LTD TP (7/1/97-28/2/02)	176	0.13	0.21	0.30
LTD SRP (7/1/97-28/2/02)	183	0.08	0.14	0.20
HFD TON (12/12/04-31/1/06)	1454	4.5	5.5	6.7
HFD TP (14/1/04-31/1/06)	2290	0.09	0.17	0.24

HFD SRP (1/2/05-31/1/06)	1340	0.06	0.09	0.14

Table 23. Hydrological model parameters: ~~“Expert” values;~~ bounds; and performance metrics (baseline simulation) ~~(baseline scenario)~~

	SDMAX (md^{-1})	SRZMAX (m)	KSURF (-)	SPLIT (-)	KGW (d^{-1})	KSSF (d^{-1})
“Expert” value	0.02	0.019	0.08[a]	0.56	0.0011	0.041
Lower Bound	1	1	0	0	0.0001	0.02
Upper Bound	100	500	5	1	0.02	1
NSE (-)	0.80					
MBE (%)	1.00					

[a] KSURF was reduced to 0.012 in the MI scenario

Formatted: Not Highlight

Formatted: Not Highlight

Table ~~3~~4. Nutrient modelling parameters; from baseline and MI scenarios (only values that were modified from baseline in the MI scenario are shown in parentheses)

Parameter	Nitrate ($mg\ L^{-1}$ N)	SRP ($mg\ L^{-1}$ P)	PP ($mg\ L^{-1}$ P)
COFMIN	0.4	0.01	0.01
CSS	8.0 (4.0)	0.03 (0.15)	
CGW	4.5	0.22 (0.08)	
KSR(*N*)[a]	0	70	700

[a] units (mg day m^{-4})x10^3

Table ~~4~~5. Nutrient modelling results; from "Expert" calibration in the baseline scenario (1997-06[a])

Dataset	C_{mod} Mean ($mg\ L^{-1}$)	Error (%)	C_{mod} 90[th] ($mg\ L^{-1}$)	Error (%)	R^2 (-)
LTD Nitrate	6.0	5.4	7.1	3.3	0.04
LTD TP[a]	0.14	-58	0.21	-50	0.02
LTD SRP[a]	0.13	-4.9	0.21	5.0	0.22

[a] Calculated up until 28/2/2002 only

Table ~~5~~6. Sensitivity Analysis Results (1997-06)

mean (min-max) C and Q	"Expert" (~~Fit~~baseline)	5th percentile Behavioural	Median Behavioural	95th percentile Behavioural
Q (mm d^{-1})	1.4 (0.46-6.4)	1.1 (0.08-4.5)	1.4 (0.20-5.6)	1.7 (0.41-8.8)
TP C [a] (mgL^{-1} P)	0.14 (0.06-1.9)	0.14 (0.07-0.22)	0.21 (0.11-1.2)	0.23 (0.19-3.9)
SRP C [a] (mgL^{-1} P)	0.13 (0.06-0.22)	0.14 (0.07-0.22)	0.20 (0.10-0.22)	0.22 (0.17-0.38)
Nitrate C (mgL^{-1} N)	6.0 (1.7-7.5)	4.5 (0.73-5.0)	4.8 (2.2-6.6)	5.9 (4.5-7.3)
TP Yield [a]	0.69	0.72	1.11	1.31

(kg P $ha^{-1}yr^{-1}$)				
SRP Yield [a]	0.62	0.72	1.10	1.28
(kg P $ha^{-1}yr^{-1}$)				
Nitrate Yield	33.2	22.8	26.1	32.1
(kg N $ha^{-1}yr^{-1}$)				

[a] Calculated up until 28/2/2002 only

Figure Captions

Figure 1 Schematic map of Frome Catchment showing monitoring points (from Bowes et al., 2009a)

Formatted: Not Highlight

Figure 2 Timeseries plots from the sub-daily HFD dataset from the Frome at East Stoke monitoring point: (2a top pane) Flow data from the catchment outlet comparing the daily mean (DMF) with sub-daily flows by showing the residual; (2b middle,) TON and (LTD) Nitrate data; (2c bottom) with the results of a two-store MIR model also shown (red line), TP, SRP and (LTD) SRP data. The numbered labels (1-5) refer to a classification of different event types described in the text

Formatted: Not Highlight

Formatted: Not Highlight

Figure 3 Pie chart showing proportion of 2005-6 Observed TP load from different event and diffuse sources calculated from HFD dataset

Formatted: Not Highlight

Figure 4 Conceptual diagram of the CRAFT model (top) and a hillslope (bottom), showing the dominant flow and nutrient transport pathways using three colours

Formatted: Not Highlight

Figure 5 Timeseries plots of modelled (from "Expert" calibration) and observed (LTD) flows and nutrient data, with the absolute error (AE) (observed-modelled) shown above: (*from top to bottom):* 5a) Flows; 5b) Nitrate; 5c) TP; 5d) SRP. Two years of data shown.

Formatted: Not Highlight

Formatted: Not Highlight

Figure 6 Timeseries plot of modelled (using Monte Carlo sampling to determine parameter values) 5th and 95th percentile and median flows, and the observed flows

Figure 7 Comparison of the nutrient yields (N and P) from the baseline (left) and MI Scenarios (right)

Formatted: Line spacing: Multiple 1.15 li, Tab stops: Not at 9.29 cm