Hydrol. Earth Syst. Sci. Discuss., 11, C4313–C4364, 2014
www.hydrol-earth-syst-sci-discuss.net/11/C4313/2014/

Hydrology and Earth System Sciences
Open Access
Discussions

HESSD
11, C4313–C4364, 2014

Interactive comment on "Flow pathways and nutrient transport mechanisms drive hydrochemical sensitivity to climate change across catchments with different geology and topography" *by* J. Crossman et al.

J. Crossman et al.

jillcrossman@trentu.ca

Received and published: 7 October 2014

We address all issues in the order provided by reviewer #1, followed by those provided by reviewer #2. All reviewer comments are given the suffix "R#", and all author responses the suffix "A#".

We would like to thanks reviewer #1 for the time taken to provide such a detailed response. There are a number of points with which we agree, and have amended the manuscript in accordance with this. We feel, however, that some of the major concerns

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

expressed by reviewer #1 (flaws in methodology, and lack of novelty in research) are unsubstantiated, perhaps stemming from some basic misinterpretations of text and tables, a misunderstanding of the study scale, and a general misconstruing of process-based modelling.

R1: "The result of the first objective, that the catchments differ in their sensitivity to climate change due to differences in soil type and nutrient transport mechanisms, is not well backed up by the results. The results section does not include any clear comparison of sub-catchment characteristics and catchment sensitivity to climate change, and neither the results nor the discussion mention other important differences between the catchments (e.g. density of tile drains). There is no acknowledgement either that with a sample of four, differences could be down to chance. Most importantly, this is also a result that has been known for decades, and this is not acknowledged. Many studies have looked at factors influencing total phosphorus (TP) export from catchments, and factors mentioned in this paper, such as soil permeability, are already taken into account in more general risk assessment tools, such as the P Index. It's then obvious that areas with higher P risk are going to be more sensitive to changes in runoff under climate change. There needs to be much more acknowledgement of this both in the intro and the discussion and conclusion, and ore/better justification for carrying out the work in the first place."

A1: We are puzzled by the first comment, whereby the reviewer begins by saying results (that sensitivity can be related back to differences in soil type) "are not well backed up", and "could be due to chance" but continues by questioning the novelty of the research, and making unsupported claims that this is information that has been "known for decades". Whilst a quick search of the literature reveals that processes involved in P export have been widely studied, little has been defined about how these processes are linked at a watershed scale (McDowell et al., 2001) and there is much still to be explored about potential catchment-wide P responses to a changing climate (Jennings et al., 2009; Kaushal et al., 2014). Multi-million

HESSD
11, C4313–C4364, 2014

Interactive Comment

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

pound research programmes are currently underway to develop greater understandings of these interactions (including the Macronutrient Cycles Research Programme (http://macronutrient-cycles.ouce.ox.ac.uk/), and Changing Water Cycles - specifically NUTCAT 2050 (http://nutcat2050.org.uk/)).

Where the reviewer goes on to suggest that catchment sensitivity to climate change has already been demonstrated under the P index, there seems to be some misunderstanding as to either a) the role of process-based models, or b) the appropriate scale for application of the P index. The P index was originally developed by Lemunyon and Gilbert (1993) as a tool for farmers to analyse the risk of potential P loss (due to fertiliser application) to streams from individual fields. This is not an index that is suitable for catchment scale applications. For a catchment scale application of the P index, knowledge of field-scale soil P concentrations throughout each catchment would be required. This sort of high resolution data is rarely available, and makes the P index a poor tool for large scale risk assessment applications. Furthermore, as the P index was designed specifically for easy application purposes, model inputs do not include factors that can estimate responses to specific meteorological conditions (a key feature of the manuscript); there is neither a hydrological nor a meteorological component (Reid., 2011). The “risk” output from the P index relates to that of potential P to be lost directly from soils into a watercourse; there is no consideration of in-stream processing (such as dilution, settling, desorption from the stream bed), direct inputs from sewage treatment works, or importantly the effects of aggregated risk (i.e where P contributions from a field up-stream are added to P exported from soils at downstream sites).

The P index is, of course, a very useful tool for farmers analysing the current risk of P loss from their own fields. The model cannot however be used to predict future conditions, as it relies on knowledge of current soil P conditions. To assess the risk of P loss from a catchment under a future climate, the P index would require an input of future soil conditions– a gap in knowledge this empirical model cannot hope to bridge.

Process based modelling, however, is designed to fill these gaps. INCA-P, which is

HESSD
11, C4313–C4364, 2014

a distributed, dynamic model, focuses on an understanding of system understandings driving phosphorus concentrations and water flow in both the terrestrial and aquatic environment. The use of this model enables us to analyse the risk to entire catchments (not just single fields) under changing conditions. By incorporating changes in climate, with those in terrestrial (soil) and aquatic (stream) conditions, the sensitivity analysis presented in this manuscript takes into account not just the current risk, but that faced by the catchment should conditions (including soil P concentrations) change. This is a key difference. The conclusion, that clay content might be used as an indicator of sensitivity is important, because geology – unlike soil P concentrations – is unlikely to change over time. Geology (or "quaternary geology" if the reviewer would prefer) is therefore a very useful risk indicator.

The shortcomings of the P index approach for projecting future catchment-scale surface water phosphorus concentrations was not originally described in the manuscript, because the field-scale P index is widely acknowledged as a field based approached, and not applicable to analyses of future change. Additional information has been included within the introduction, however, to explain the reasoning behind the use of a process based (as opposed to an empirical) model.

In response to the data being "not well backed up", tables 1 and 2 provided a detailed, quantitative cross-catchment comparison of key characteristics, with % clay composition provided on page 8073. This information is referred to in the discussion when making links between sensitivity analyses and catchment characteristics. We chose not to repeat tables of catchment characteristics in the results section, given that the information was already included here in the catchment description. For the reviewer's benefit, additional columns have been added to tables 7 and 9 (provided as tables 1 and 2 in the attached document), to directly compare sensitivity with catchment characteristics.

With regards to the possibility of results being "due to chance" – we are unsure of the reviewer's point. The comparison of hydrochemical responses over four different catchments gives a much broader understanding of catchment interactions with climate change than can be achieved from looking at a single, or pair of catchments. We are perplexed by the apparent lack of consistency from the reviewer here, given that the reviewer cites (below) three single-catchment studies and suggests these are sufficiently comprehensive as to preclude the need for any further climate change research. It is precisely because the results derived from a single catchment study can be criticized as being performed in isolation, or "due to chance", that we compared results between additional sites.

R2: "The second objective looks at climate change impacts on hydrology and water quality. The novelty here is in looking at the uncertainty within a GCM, but this is not enough in itself to justify publication, as other studies have already considered this elsewhere (e.g. Dunn et al., 2012; Fung et al., 2013). A previous paper already describes likely climate change impacts in this region wrt phosphorus (Crossman et al., 2013), and another application of INCA-P with additional climate scenarios is not, to my mind, novel enough to merit publication in itself. If the authors think this can be justified and is novel enough to be published, then the introduction needs additional detail to justify the study and put it into further context. For example, have other studies looked at the differences between uncertainty in projected future flows and TP concentrations/loads compared to the size of the projected changes?"

A2: The reviewer states concerns over the novelty of this manuscript, and writes that our aim is in "identifying uncertainty within a GCM", citing two papers that "have considered this elsewhere". This is not the focus of the manuscript. We look to identify whether in the face of an uncertain future, hydrological and chemical processes might dampen or amplify that uncertainty. As far as we are aware, we are the first study to apply GCM uncertainty from a perturbed physics ensemble (a tool for investigating much wider ranges of plausible future climatic uncertainty which only became available within the past 6 years) to a hydrochemical impact model. Whilst it is true that the studies cited by the reviewer did look at the potential impacts of multiple GCM scenarios upon a single catchment, further novelty in our work lies in the cross-catchment comparison of responses to a range of climate scenarios, whereby adjacent catchments, with similar land use, demonstrate markedly different hydrochemical relationships with climate change. These important insights, both for process understanding and catchment management, could not have been determined through the assessment of uncertainty within a single catchment.

We use our results to investigate possible causes of the differences in catchment response, and whether generalised catchment characteristics might be used to predict the resilience of catchments to potential change. Having determined that %clay content might be used as an indicator, this finding could be further developed for broader scale risk applications (where data is limited), or as a tool for prioritising areas for future study (where data collection might be required). This novel research is therefore also fundamentally useful. Again, this determination would not have been made through a single catchment study. Ultimately, we must disagree with the reviewer's contention that just three single catchment studies (Dunn et al., 2012; Crossman et al., 2013; and Fung et al., 2013) are sufficient to put to rest the paradigm of climate change impacts on water quality.

R3: "There is no discussion of why process-based modelling is needed or used in the study. What is the added benefit? Would it not have been better to just compare the characteristics of the catchments and from those alone determine which was more sensitive to climate change?"

A3: For the reviewer's benefit, an additional paragraph comparing empirical and process-based models has been added to the manuscript. The reviewer's questions are confusing here and perhaps relate to a general misunderstanding of the role of process based models in the hydrological sciences. Perusal of the contents of recent issues of HESS confirms that the value of process-based modelling is widely appreciated by the hydrological community. In the specific case of this manuscript, process-based models are required in order to calculate catchment sensitivity to climate change, and thus it is difficult to envisage how catchment characteristics alone could have been used to determine "which was more sensitive to change". The manuscript uses sensitivity derived from process-based models to demonstrate that general catchment characteristics (quaternary geology) can be used as an indicator of potential sensitivity to climate change. As this insight is a novel finding resulting from this study, the a priori use of geology (or other catchment characteristics) to determine sensitivity would not have been possible.

Perhaps the reviewer misunderstands the term "sensitivity" in the context of impact modelling. It refers here to the response of water quality and hydrology to prescribed changes in future climate (Bates et al., 2008). This was defined on page 8070. An analysis of catchment sensitivity to future change therefore requires a quantification of changes in hydrology and water quality. Differences between the hydrologic and biogeochemical response of the individual catchments are accounted for by the catchment-specific parameterization of INCA-P. Thus, variations in catchment properties (or "characteristics") were reflected in differences in INCA-P parameter values (presented in tables 1 and 2). The sensitivity was established through process-based modelling, and an assessment carried out as to dominant causative processes, finally associated back to generic catchment characteristics. The aims and objectives have been re-phrased to help clarify this.

R4: "The results and conclusions rely on using INCA-P to predict future stream flow and water quality, but no model validation was carried out, so we can have no faith in the model's predictive capacity. Model performance outside the calibration period is often significantly poorer, and the credibility of the model set-up for a given catchment must therefore be evaluated against independent data (Refsgaard and Henriksen, 2004). This test data set should test how well the model can perform the task it's intended for (different climate, in this study), problematic when looking far into the future. Refsgaard et al. (2014) provide a useful framework for this kind of modelling study."

A4: We feel that the reviewer has used the term "validation" here in a rather misleading way. The reviewer seems to suggest that by using additional independent data, the reliability of the model (in predicting future conditions) might be established. The notion of model validation in climate change impact studies was refuted over 20 years ago, in “Verification, validation and confirmation of numerical models in the Earth Sciences” (Oreskes, 1994). A match between predicted and observed data cannot prove any kind of accuracy in future predictions, because we are dealing with natural, dynamic, open systems, and no amount of independent data can account for the manner in which the system can change in unanticipated ways (Oreskes, 1994). When studying future change in natural systems we will always be left with the possibility that future observations will bring into question both our existing theory and our model calibrations. As a result of this then, as Oreskes points out, although models can be “confirmed” through a demonstration of agreement between observed and predicted data (during calibration), any sort of validation of their ability to predict future data is logically precluded, via an incomplete access to (future) natural phenomena. We thank the reviewer for drawing our attention to Refsgaard’s work, who also concludes that “we cannot, in a strict sense, perform validation tests on the ability of our models to project the climate change effects since we have no data truly reflecting the future conditions” (Refsgaard et al 2014).

Thus, as we can never completely verify or validate our future projections, we are left only with the ability to calibrate our model to the best of our abilities. In summary, we disagree with the reviewer’s statement that testing the model against independent data would give “more faith in the models predictive capacity”; as the model performance statistics during an observed validation period are no better an indicator of performance under a future climate than the model performance statistics during the chosen calibration period. Furthermore, recent research has shown (Larssen et al., 2007) that model performance is much greater where built on longer periods of observed data, than where shorter observed periods are used, and that it is important to use the longest time series of calibration data available (rather than to reserve additional observed data for validation). This is because where longer time periods are used, a greater range of natural phenomena are incorporated, and models can be built to predict long term trends. By calibrating the model on the largest possible range of current climate inputs, we maximise the likelihood that the model will be capable of responding to future conditions.

Ultimately, however, we stress that models are investigative tools: they can be used as a guide for management, and for prioritisation of research – but in terms of future projections, model results should be interpreted only as ranges of plausible and less plausible outcomes, given the best information currently available; results cannot be verified. Additional information on this subject has been added within the text.

R5: "Section 2.2 is lacking lots of detail on model calibration, including (i) calibration period used for each study catchment, (ii) data used for calibration, (iii) method for calibration: Graphical analysis plus manual tweaking of parameters? If so, what was the procedure followed, what performance statistics were used,. . .? (iv) Were parameters varied by sub-catchment and reach within a study catchment? (v) How many parameters therefore needed to be estimated and calibrated per study catchment? (vi) How many of these were based on some form of measured data (e.g. GIS-derived or based on literature values), how many were calibrated, but within a range derived from the literature range, and how many were purely calibrated?"

A5: We are unsure as to why the reviewer wishes to know the proportion of the total number of parameters that were calibrated using measured/observed values. This is unnecessary data, as it is most important to accurately calibrate the key parameters. Whilst obtaining measured data for 99/107 parameters within a model, a poor fit to data could still be obtained where highly sensitive parameters were not monitored. Lepisto et al (2013) and Crossman et al (2012) have determined that within INCA-P it is most important to obtain observed/accurate data for Fredulich coeffecients, soils data, and flow a parameters. This parameter information is presented in the manuscript, and INCA parameter sensitivity has now been clarified. Observed or calculated data were obtained for 30 parameters (including those identified as "key" to operation of

HESSD
11, C4313–C4364, 2014

the INCA model) and a further 6 based on literature data. The calibration procedure followed is explained in detail in section 2.2, and an overview has been added to the revised manuscript, including an outline of how the plausibility of parameter values were assessed by a) calculating P inputs [as described in the methodology and tables 1 and 2]; b) GIS and digital elevation based assessment of surface hydrological flow pathways, subcatchment size, and landuse areas; c) comparison between modelled and observed time series of flows and phosphorus concentrations [using Nash-Sutcliffe coefficients, R2 statistics, and MAE]; d) use of literature values for soil processes and e) expert judgement. The varying of all parameters by subcatchment is now included in the text for clarity. Model calibration data sources are described in detail in section 2.2, and in tables 1 and 2. Although presented in figure SI3, the period of calibration is now also repeated in the text for clarity.

R6: "To encapsulate full parametric uncertainty in the GCM, the members selected from the PPE should represent this uncertainty. However, from the description in the paper it seems that the most sensitive members were selected that still provided reasonable estimates of baseline climate. Surely the selected range should have included the least sensitive as well as the most sensitive?"

A6: We absolutely agree that the members selected from the PPE should represent the uncertainty in the GCM; this is what was done and to clarify the text has been revised accordingly. We selected 5 ensemble members from the PPE, and whilst we did indeed ensure that the "most sensitive members [. . .] still provided reasonable estimates of baseline climate" – we also included within our selection the least sensitive ensemble members available (see table 3). Our aim was not to assess the full range of parametric uncertainty under a single GCM, but to assess a wider range of plausible futures than might have otherwise been considered under a multi-model ensemble (see Collins et al., 2006 for a full explanation on the benefits of PPE vs use of traditional multi-model ensembles). We include in the supplementary material (figure 1) a comparison between the range of plausible futures provided by the PPE, and those available from a more traditional multi-model approach. In short, the PPE is a more focused and deliberate approach to addressing uncertainty in climate change. Whilst the multi-model ensemble includes a range of different parameters by chance, the PPE has varied those parameters in a methodical manner. It is clear that, although some GCMs project slightly cooler futures, the selected PPE ensemble members give a much wider range of plausible future precipitation projections than would be investigated though a multi-model ensemble approach. Therefore, the members selected for use in this study enable managers to consider a broader range of conceivable scenarios. The next step would be to derive a PPE from each GCM (Collins et al., 2006); though PPEs are not yet commonly available.

R7: "For each study catchment, only one 25km2 grid cell was used to provide the climate change data (P.8077, l.13-17). It is good practice to average at least two or more RCM grid cell projections when using climate change data. In addition, later in the paper much attention is given to looking at differences between climate projections between grid cells, which is fairly nonsensical given the errors involved. I'd recommend averaging the grid cells across the whole study area and applying a single climate change scenario to the whole catchment. This would also make it easier to compare different catchment responses, as the driving climate would be the same. If you feel strongly that this is not a good way forward, then good justification needs to be given, and the authors need to show an awareness of the lack of significance of any differences in projected climate between squares when it comes to reporting results (e.g. p.8083, l17 onwards), and the discussion."

A7: It is important to consider that this is not a top-down, scenario led approach looking to formally predict the future water quality of these catchments, but a bottom-up, "stress test" of the hydrology and water quality responses of different study to different climate projections. Suggested methods for each are very different (Brown and Wilby., 2012; Nazemi and Wheat., 2014).

Given the highly localised nature of rainfall patterns in the Simcoe region (Smith., 2010), both HBV and INCA were calibrated on point data taken from catchment-specific observed weather stations (described in the text, and figure 1 of the manuscript). The reviewer suggests that we should average all 25km2 regional climate outputs across the study area (which in total is over 2914km2) and apply a single climate scenario to all INCA models; but for the calibrated models to most accurately respond to projected climate change, they require climate data from the area closest to/covering the area under which they was calibrated.

The original GCM (at a grid resolution of 500km2) was downscaled to 25km2 grid cells using the PRECIS regional climate model, precisely to take account of the local climate variability (due to elevation, lake effects etc.). Upscaling this detailed information back to such a large extent would introduce significantly greater errors into the modelling study. If the study were to be conducted at this scale, no regional downscaling would have been required, and would have resulted in misrepresentation of projected change within every catchment.

R8: “P.8078, l14-16: The text needs to be clearer about the very important (and quite likely invalid) assumptions involved in using delta change for bias correction, namely the assumption that the relative difference between the simulated baseline and the simulated future is realistic, despite any bias. The authors also need to make clear that this method only corrects for bias in the mean, not in the variance. Very importantly, both in the methods and in the discussion, there needs to be discussion of the fact that any potential increase in the intensity of rainfall is likely subdued using this method. This is a big source of uncertainty, particularly when looking at phosphorus, which is so affected by storm events.”

A8: Delta change is the dominant method used in the US National Assessment (Hay et al., 2000) and that recommended by the IPCC (Carter., 2007). Three pages of the manuscript (8077-8079) have been devoted to the description of, and justification for use of delta change in this study. We feel that its application to the dataset has been vindicated. The claim of the reviewer that delta change is an “invalid” method is

HESSD
11, C4313–C4364, 2014

again unsubstantiated. We agree with Teutschbein and Seibert (2012) who note that this form of bias correction does not account for changes in distribution of wet and dry days, however the same can be said of linear scaling. Alternative methods such as local intensity scaling, variance scaling and power transformations each have their own drawbacks, and there is ample literature to support the use of delta change (for example Teutschbein and Seibert., 2012).

We have included additional information as to the shortcomings of the delta change methods; however it should be noted that bias correction to observed station data is absolutely a necessary process within climate change applications to hydrochemcial impact models. Whilst all methods have their limitations, this does not preclude their use (Teutschbein and Seibert., 2012). On the contrary, provided the readers are made aware of the specific method used, and the corrections applied to the data (as on pages 8077 – 8079), there is no right or wrong way to bias correct the data.

R9: "Section 3.1 (Results: INCA-P model calibration): This section needs re-working, including: (i) It needs to be made clearer throughout this section what is being compared with what. Are the statistics for daily, monthly or annual means? All three are mentioned, I think, but not for every catchment. For consistency, it'd be good to give performance statistics for all time periods (daily, monthly averages and annual averages) for all sub-catchments, e.g. for the catchment outflow. It's likely that the statistics for the daily data won't be great, but if for example the performance statistics for monthly or annual TP are acceptable, then that can be used to decide over which timescale it's appropriate to discuss model output for the future period. (ii) Much of the information in the results could be put into Table 4 and the text correspondingly cut down. (iii) The results should be put into the context of 'acceptable' performance statistics from the literature (e.g. Moriasi et al., 2007), taking care to make sure that like is compared with like in terms of concentrations/loads and timescales over which the data are averaged before calculating performance statistics. (iv) This section also needs validation period statistics (v) As the point of this section is to demonstrate that the model is fit for being used to predict future conditions in the study catchments, there also needs to be some discussion of whether the right processes are operating. This is particularly important given the amount of text given over to describing catchment processes in the discussion. Some of the conclusions rely on the model having correctly simulated different flow paths, for example, so it's important to establish at this stage that the model is in fact producing realistic simulations of the different flow pathways and nutrient transport mechanisms in the different sub-catchments. (vi) Finally, as dissolved and particulate phosphorus may follow very different transport pathways to the river, it would be very interesting to consider the two separately, at least in this calibration period. This would help increase confidence that the model is performing adequately for the task in hand."

A9: All model statistics reflect monthly averages, calculated over the complete calibration time period (see response to section 5 for calibration times scales). This has been clarified in the revised manuscript. Figure 5 (of the manuscript) gives an example of the INCA model output on the daily timescale, for reference only. This is actually a rather short section, given its importance (as the reviewer also notes), and we are reluctant to relegate the information to another table. We cannot put the results in the context of "acceptable performance statistics" as there is no generically acceptable level at which a model is deemed to be "correct" (see response to point 5); every application is different, and relative only to itself. Model statistics are provided so that the reader/user can decide for themselves how much confidence they have in the results. Daily statistics are not relevant here, as only monthly averages (over 30 year time periods) are used in calculating future change. The reviewer suggestion that the calibration performance accuracy be used to determine the "timescale [over which] it's appropriate to discuss model output for the future period" is nonsensical, as due to variance it is inappropriate to discuss future change projections in anything less than monthly averages derived over decadal periods (Carter., 2007).

We agree with the reviewer that the confirmation of the model's ability to represent system processes is important. Equifinality is a general issue with process based modelling. Without long-term tracer studies, and other very detailed monitoring data it is difficult to ascertain whether models are correctly representing processes involved. It is however important to consider that these are multi-branched models, which have been spatially calibrated across many tributaries and a wide ranges of sites along the main stream. As a result, we can have confidence that the upstream loads travelling into each reach are accurate (based on comparison with upstream observed data), that diffuse inputs from soils are accurate (by comparing soil coefficients with other studies) and that point source inputs are accurate (from STW input reports). Whilst this in itself is not proof of the correctness of the model representation of reality, it is strong confirmative evidence that we are getting the right answers for the right reasons.

For clarification, sections 3.1 (model calibration) and 3.2 (export coefficients) have been merged in the revised manuscript, whereby the purpose of section 3.2 was to explore the modelled nutrient responses and compare those to observed data. The consistency between INCA soil export coefficients and nutrient exports derived from previous studies of these catchments (Thomas and Sevean, 1985; Winter et al., 2007; Baulch et al., 2013) gives confidence that nutrients from both diffuse and point sources are being transferred in an accurate manner.

For further clarity on model process performance, an additional analysis into model hydrological responses is presented (see Figures 2 and 3 of the attached author comments). In figure 2, model performance statistics are given separately for the rising and recession limbs of modelled verses observed hydrographs, to determine the accuracy of model responses to precipitation events in each study catchment. The same time period is analysed in each catchment during November- December 2011, during which an intense precipitation event occurred, beginning on the 27th of December. The high R2 and low model error in both rising and recession limbs of most catchments gives high confidence in the runoff simulations performed by the HBV-INCA model chain. The higher model error in the recession limb of the Pefferlaw catchment is likely due to onset of freezing in the catchment, which is not captured by the Pefferlaw at this daily

HESSD
11, C4313–C4364, 2014

timescale. The preceding event in the Pefferlaw demonstrates a more acceptable fit, with an MAE of only -1.7% on the rising limb, and 19% on the recession.

Figure 3 analyses model performance under spring melt. Analysis is performed as monthly averages during 2010, except in the case of the Whites River, where observations are rarely collected during winter, and an average of 2010 and 2011 results were required to provide sufficient data. Again high model performance is demonstrated, and gives confidence in flow pathways. Of particular note is the high seasonal accuracy of the Pefferlaw, which despite a late onset of freezing on a daily time-scale, successfully represents seasonal frozen water stores (January and February) and the transition to spring melt runoff (March). We must stress that these analyses go above and beyond those generally presented in impact modelling studies. Further verification of flow pathways would require extensive tracer studies, for which the data are simply not available. The excellent performance of HBV in snowmelt driven catchments has been confirmed, however, through a 3-year tracer investigation in Denali National Park, Alaska (Crossman et al., 2013).

Finally, we disagree with the reviewer's suggestion that it would be "interesting to consider [dissolved and particulate phosphorus] separately" in the calibration period. Separate calibrations of dissolved and particulate P would give erroneous information. Joint calibration of total and dissolved P across all reaches increases the likelihood that we will obtain the right answer for the right reasons.

R10: "Please provide tables with all the final parameter values used, for both HBV and INCA-P, per study catchment, in the supplementary information."

A10: A table with the key parameters used has been provided. With over 107 parameters, in 4 different models (428 parameters) – these would be some large files. Additional information has been included here – at the reviewer's request (e.g. Plant uptake) (see table 3 of attached document).

R11: "Section 3.3 (Climate change): I recommend moving all of this to section 2.3, as this is the input to the modelling, not a result in itself. In addition, shorten the text as the key messages are somewhat lost at the moment, and rely more on data in Table 6. There also seems to be a bit of repetition between the text, table 6, Figs SI7-9. They do all show slightly different things, but probably don't merit the amount of space taken up."

A11: We disagree with the suggestion that the climate change results should be moved to the methodology. Whilst uncertainty in climate change is not the sole focus of the study, differences in projected uncertainty between catchments (under the perturbed physics ensemble) is most definitely a finding of the study; and not to be considered simply part of the methodology. Perhaps the captions were unclear on these tables and figures – these have been clarified in the revised manuscript - the information presented in tables 6 and SI7, and in figure SI8 is very different. Table 6 presents the average uncertainty across all probability levels within the CDF. Given the manner in which climate change uncertainty varies between probability levels (to different extents within different catchments) (demonstrated in SI7), table 6 gives a more complete measure of total uncertainty within a catchment (and is more appropriate for comparisons between catchments). SI8 provides information on the seasonal changes in temperature and precipitation (something that cannot be seen through the CDFs).

R12: "Throughout the paper, results are quoted too precisely(in terms of decimal places), given the errors and uncertainties. The authors also confuse significant figures and decimal places (e.g. tables 1 and 4 to 9). The number of decimal places should be reduced to 0 or 1 throughout. E.g: p.8074,l7-8: cm of snow falling given to nearest 0.1mm; reduce to nearest cm; percent changes throughout section 3.3, 3.4.1, 3.4.2 (and corresponding results tables)."

A12: Precision has been reduced accordingly

R13: "Section 3.4.2 (Water quality): This is hard to read at the moment, as too many numbers are quoted, breaking up the text. I'd suggest relying more on tables, and

HESSD

11, C4313–C4364, 2014

Interactive Comment

summarising only key results in the text. Splitting this section into sub-headings could help (e.g. total annual TP loads, monthly TP concentrations, seasonality). The 'cross-catchment range' is not very useful, it's clearer to just look at the differences between catchments. If the authors disagree, perhaps this could be pulled out of the text and summarised more."

A13: We agree that this section was rather heavy. For ease and consistency, each results section (climate change, hydrology and water quality) has been further broken up into 3 parts: i) likelihood of change, ii) seasonality of change iii) sensitivity to change. This structure has been implemented within the discussion, to help readers draw direct conclusions from the results. The numbers quoted within these sections, however, do highlight the points made. Within a PPE there is a vast quantity of data generated, and it is a delicate balance between over-burdening the text, and providing a "predominantly qualitative description" (a criticism the reviewer raises about the site description).

We are unsure what the reviewer means by " the cross-catchment range is not very useful, it's clearer to look at the differences between catchments"; given that the range is a measure of the difference. The focus of the manuscript is in comparing the range of climate change uncertainty (across the 4 catchments) with the range in hydrochemical responses. It would be difficult to justify removing these values. A clarification of the calculation used to determine the "range" has been given in the text (though this is a generally accepted mathematical term).

R14: "Section 3 (results): There is no attempt to link catchment characteristics with modelling results, despite this being one of the main objectives of the paper. A summary results table with the main differences between catchments in terms of modelling output, together with the main differences between catchments in terms of their topography, soils, etc. could be useful, plus some mention in the text."

A14: We agree that a summary results table would add clarification. Whilst tables 1 and 2 do quantitatively present the data that is referred to in the discussion, additional rows have been added to tables 7 and 9 to highlight the connection for readers (see tables 1 and 2 in attached document).

R15: "Discussion: This needs strengthening in a number of ways. It would be good if at some point it linked back to the original objectives. The first paragraph of the discussion could be deleted, as it belongs more in the introduction. Otherwise, I thought there were three main problems with the discussion: (a) There was a general lack of clarity of whether the text was referring to real observations, or simulations backed up by observations. Many of the processes mentioned in these paragraphs (e.g. loss of organic matter, macropore flow contributions and tile drainage, drainage of wetlands,. . .) aren't specifically included in INCA. These processes might indeed be important in reality, but did the modelling capture it? Need to link back to results showing it did or didn't, with a discussion of the model's limitations in relation to these key processes. Also need to discuss sooner how drainage of wetlands was taken into account in the model. (b) The discussion doesn't consider how the results fit into the wider work carried out on uncertainty in climate change, or sensitivity of different areas to P losses (even just for baseline climate). (c) The discussion doesn't consider any of the limitations or caveats of the study, of which there are many. It is crucial that these are acknowledged to not give a misleading impression of the confidence that can be placed in the results of this study."

A15: a) Additional clarity has been added to the discussion (See comment 13), though we would like to emphasize that the discussion does link directly back to the original aims. The two sub-headings, through which findings are discussed, are direct references to the study objectives stated in the manuscript introduction. Under each subheading, the implications of the results in relation to the relevant objectives are discussed – both in terms of importance to the specific study sites, and to the wider research community.

We agree that on page 8089 additional clarity was required about whether we were referring to model simulations or actual processes. This has been re-worded in the

HESSD
11, C4313–C4364, 2014

revised manuscript. We were writing of simulated outputs and simulated processes, with literature support used to demonstrate that model behaviours were plausible. All processes discussed are modelled within INCA, and changes in section 2.3 (discussion of model process clarification), along with columns in tables 7 and 9, should help with clarity here. Ultimately the aim of this section was to demonstrate that the model performance matches that suggested by observations presented in the literature, and that our modelling captured these behaviours (i.e giving further confirmation that we had obtained the right results for the right reasons).

The reviewer asks for an explanation as to how drainage of wetlands was taken into account in the model. Additional information is now provided within the text, and the additional figures (2 and 3 in the author response) should also add clarity. This was explained within section 2.3 (calibration) where it was noted that the INCA-P applications presented here used 5 landuse classes (one of which is wetlands). Calibration of landuse classes includes runoff rates, soil water storage and infiltration rates - to match observed hydrological data. As part of the calibration process therefore, the hydrological behaviour of wetlands in the Holland watershed have been adjusted to match observed hydrological outflow. b) We disagree that the discussion did not consider how the results fit into the wider context of P sensitivity and climatic uncertainty. Page 8088 and 8089 explore the implications across all catchments where snowmelt hydrology is important. Page 8093 looks at studies of P sensitivity within catchments across Europe (e.g. the Rhine (van der Hurke, 2004)) and at a series of sites within Denmark (van Roosmalen, 2007). Finally, page 8093 discusses the wider applicability of the findings, with respect to a focus on internal process dynamics, rather than meteorological drivers. c) We agree that it is important to clarify study limitations. Both in the presentation of methods and of results it has been made clear the accuracy of model behaviours, and that sufficient information is given to allow informed decisions by the reader as to how much confidence to place in conclusions. We have, however, drawn this information together into a summary, in a “study limitations” section, to add clarification for the reader.

R16: “Whilst the paper is reasonably well structured, the writing is not precise enough to communicate the sometimes complex concepts in a clear and transparent way (e.g. the authors confuse variance and difference, refer to model performance statistics as model coefficients, are often not clear whether they’re referring to the climate model ensemble average or members of the ensemble,. . . .). I’ve highlighted quite a few examples below, in the minor comments. For methods that were used in the study, the past tense should also be used (e.g. p.8075, line 11; p.8079, line16). The present tense is confusing, sounding like a general statement of accepted science, rather than a description of methods used in this study.”

A16: We are sorry that the reviewer was confused by the description of differences between model ensemble average and individual ensemble members, and have added clarity to these. As per the reviewer’s preference, the methodology has been converted to past tense in a revised manuscript. Whilst the term “variance” was used on three occasions in place of “range” (corrected in the revised manuscript), there has also been some confusion on the part of the reviewer as the mathematical meaning of the term range (the difference between the lowest and highest value). Finally, we have clarified the manner in which we refer to “performance statistics” i.e. R2 , MAE and Nash Sutcliffe but must stress to the reviewer that it is equally common to refer to these terms as “model coefficients” (Weglarczyk, 1998; Cochrane, 1999; Saleh et al., 2000).

Minor Comments:

R1: “Introduction: Confusing absorb and adsorb several times”

A1: Thank you. The word absorb was used once in error, and has been converted to adsorb in a revised manuscript.

R2: “P.8071,l23: geological (i.e. bedrock) differences between catchments aren’t mentioned, only differences in drift and soils.”

HESSD

11, C4313–C4364, 2014

Interactive Comment

A2: We are speaking of quaternary geology (overlies the bedrock, beneath the organic soil layer). These geological differences are described in detail. “Quaternary” has been added for clarification.

R3:”Section 2.1 (Site description): Describe available data for model calibration and testing”.

A3: Section 2.1 is a site description of the catchments. The available data used for model calibration is described at length, both qualitatively and quantitatively under section 2.2.2 (“Model calibration”), and in tables 1 and 2. See comment 14.

R4: “P.8073, l4-28: This is an important pargraph, which currently makes for somewhat confused reading. I'd recommend summarising more, whilst keeping key information in there. Key differences between sub-catchments could be summarised, quantitatively where possible, in a table. This could then be linked to the modelling results”.

A4: It is unfortunate that the reviewer was confused here. Table 2 provides this comparison between catchments. For clarification, we have inserted a reference to this table in the suggested paragraph, and included differences in geology (or “quaternary geology”), i.e. % clay composition in tables 2, 7 and 9.

R5: “Section 2.2 (Dynamic modelling. . .): I'd suggest splitting this into (a) a description of the model; and (b) a description of the model set-up and calibration”

A5: OK

R6: “P8074 l25: the use of the word ‘parameters’ is confusing. Replace, e.g. fluxes, variables”.

A6: Agreed – changed to variables

R7. “P.8074, l25-27: confusing. I think model output timeseries are being referred to here? If so, clarify.”

A7: Clarified

R8: “P.8074, l26: soil export coefficient isn’t an output. Replace with soil erosion, this is what’s meant”.

A8: We thank the reviewer for this observation. This has been changed to “nutrient export coefficients” [from the soil]. Export coefficients represent the quantity of nutrients (or sediments) generated per unit area, per unit time (e.g. kg/ha/year). They are used for source apportionment to determine the amount of nutrients that a given landuse activity contributes to a downstream water body (McFarland and Hauck, 2001).

R9: “P.8075,l11 and l13: an individual HBV model set-up was used for each catchment, not an individual model”.

A9: Thank you. This has been altered

R10: “Model calibration: How many parameters requiring calibration does HBV have? How were these calibrated?”

A10: The method of HBV calibration has already been described (page8075, lines 6-10). A detailed model description of the parameters within HBV is provided in Crossman et al (2013) and in Saelthun (1995), and would detract here from the focus of this manuscript. Whilst it is important that the manuscript is clear about the methods used – it is important that key information is not lost by presenting details that can be obtained elsewhere.

R11: “P.8075, l20: Presumably the hydrological network was used to delineate sub-catchments, rather than flow data (i.e. discharge data)? Also, how did you decide how many sub-catchments to have? On what basis?”

A11: Flow changed to “network”, and description of the use of arc-hydro to delineate catchments has been included. ArcHydro was used to develop both the flow network and to delineate subcatchments. Catchments were derived based on 1% of maximum flow accumulation (a simple rule of thumb for stream determination thresholds). As this resulted in delineation of separate catchments within a single river reach (which

HESSD
11, C4313–C4364, 2014

were not necessary for calibration), these fine scale catchments were then grouped by stream order, i.e. so that every tributary within the catchment could be individually calibrated. The decision-matrices for determining subcatchments is not relevant to the study however, and is not included in the manuscript. As INCA is an integrated model, then provided all reaches for which observed data is available are separately delineated (so as to enable individual calibration of these areas) splitting (or merging) catchments for which data is unknown will not influence affect model accuracy.

R12: "P.8075,l20-25: refer to SI3 and tables 1 and 2."

A12: References added.

R13: "P.8075, l.26: parameters for model calibration were 'calculated'. This is a bit confusing, as calibration is the altering of model parameters by trial-and-error to optimise model performance."

A13: It seems the reviewer's understanding of the term "calibration" differs to ours. By calibration, we mean the "determination of model parameters and/or structure on basis of measurements and prior knowledge" (Janssen and Heuberger., 1995) so as to "optimise the parameter values in such a manner that the model output best fits field observation data" (Refsgaard et al. 2014). Estimation of model input parameters based on observed data is common (good) practice.

R14: "P.8076, l7-14: much of this is repeated in Table 2."

A14: We disagree. This is a description of how the quantitative parameter values (presented in table 2) were derived (described in this section). Without the description, it would be unclear how the values were obtained.

R15: "P.8076, l14: from Table 2, I see that septic inputs were classed as inputs to nonintensive agriculture. Justify this in the text."

A15: This has been clarified by including the detail of households with septic tanks being located in rural areas. By law, houses in urban areas must be connected to

HESSD
11, C4313–C4364, 2014

existing municipal sewage treatment systems (Ministry of Environment, 2006). Only those households situated in rural areas (and thus disconnected from the main sewage systems) may use septic tanks. As these houses are not built in the centre of corn or wheat fields (intensively farmed), on wetlands, or in dense forests - they are generally associated with "non intensive agriculture". We did not feel that this detailed insight into the modellers' decision making process during model set up was required.

R16: "P.8076, l.14: what about plant uptake? E.g. maximum uptake? Timing? The P budget is the key thing controlling model output, not just P inputs, so these parameters are just as important."

A16: We agree that the P budget is important, and maximum plant uptake rates have been added to the text. The result of the difference in P budgets (export coefficients) are presented in table 5 (see response to comment 8). The reviewers opinion on the importance of P budgets (i.e inputs, outputs and processes) is a little inconsistent with their comment (no.41) – where they state that budgets are nothing more than the sum of inputs and outputs (i.e. no process interactions)? Whilst the authors agree on the importance of the P budgets, there are over 107 parameters in INCA, each of which have different degrees of influence over the model output. Common practice is to focus on those of greatest importance; previous papers have demonstrated that plant uptake is not one of these parameters (Wade et al., 2001; Lepisto et al., 2013). As noted in previous comments, details of calibration procedure for the most significant parameters have been provided in the manuscript. Determination of plant uptake values was based on previous model applications to the Simcoe catchment (Jin et al., 2013), but as direct measurements were not available, it was not originally included in table 2. We feel that a detailed discussion of all 107 model parameters would be unnecessary and unusual, to say nothing of being extremely tedious for all but the most fastidious of readers.

R17: "P.8076, l16: In Lepisto et al. (2013), the equilibrium coefficient was only mentioned in terms of a PEST-calibrated coefficient, which was then compared to lab measured values (p.56 of the report). So was PEST used for calibration? Or were their lab measured values used to decide on parameter values?"

A17: There has been a slight misunderstanding here, and the text has been clarified accordingly. The lab measured values referred to in Lepisto et al (2013) are an empirical dataset of general values for a wide range of soil and landuse types (including agriculture, forestry, water and wetlands). This dataset was used to calibrate INCA soil processes.

R18: "P.8076, l20: Were the average catchment values for EPC0 determined by area-weighting values for specific soil types, based on the area of soil in the sub-catchment?"

A18: INCA is an integrated model – so there is no "catchment averaging" done by the modellers during calibration; that is done by INCA as part of the integrated output. We entered the EPCo values for the relevant landuse type, and INCA then calculates outputs [as a function of inputs and processes] through all 5 landuses, to deliver a total P export from the respective subcatchment (into the river reach). This total export is area-weighted by landuse type, and takes into account all of the different terrestrial processes operating. What needs to be known for the input is the EPCo value for each soil (or landuse) type (see SI 1), which was provided by the laboratory data – and use of which is supported by Lepisto et al (2013), where there was correspondence between expert judgement and this data.

R19: "P.8076, l23-24: mention Fig. SI3 earlier, when model spatial set-up is described"

A19: OK

R20: "P.8076, l27: confused; re-phrase to clarify that SRES-A1B is an emission scenario; HADCM3 a GCM, and the PPE reflects parametric uncertainty in the GCM."

A20: Altered

R21: "P.8076, l28: a subset of how many members of the ensemble?"

A21: Five – this has been clarified (see comment 6)

HESSD

11, C4313–C4364, 2014

R22: “P.8077, l1-6: this makes it sound like only two members from the ensemble were looked at (Q3 and Q10), not 5. It’s then stated that Q3 and Q10 were selected because they were sensitive, then that sensitive scenarios aren’t as good. This seems contradictory.”

A22: Unfortunately, the reviewer misunderstood the paragraph, and it has been rephrased accordingly (see also comment 6). Five members were chosen, and these five represented as complete a range of sensitivities as was appropriate – given the need for the ensemble members to still represent baseline conditions. Members with higher sensitivities tended to have poorer representations of the regional climate – and so the highest appropriate member was Q10. Low and intermediate sensitivity members were indeed included as part of the 5.

R23: “P.8077, l25-26: bias is as important, so report that as well”.

A23: This comment is inconsistent with reviewer comment 25. Delta change is, as the reviewer notes, a form of bias correction. This bias is presented in SI4 and SI5.

R24: “P.8077, l27: add ‘members’ after ‘ensemble’”

A24: OK

R25: “P.8078, l1: delta change is a form of bias correction (as used in this paper). Therefore this needs re-phrasing, and a bit adding to clarify what bias correction method is questionable.”

A25: Absolutely. This has been clarified in the revised manuscript. The section has been rephrased (see also comment 8)

R26: “P.8078, l28: little bias in simulated temperature is reported, so why was temperature than bias corrected? Bias correction introduces important errors of its own, so should only be done where the bias is more than a few degrees C.”

A26: We are unsure upon what scientific basis the reviewer has determined that “a

HESSD

11, C4313–C4364, 2014

few degrees C" is the limit of necessary error for the use of a bias correction. As climate change is only expected to increase by "a few degrees C" over the next 100 years, allowing such a large bias in baseline datasets is unwise. Furthermore, if bias correction were not applied to the temperature dataset, this would result in running the INCA model using a "directly forced" temperature dataset (i.e. direct from the regional climate model output), and a "delta changed" precipitation dataset (see Lenderink et al., 2007 for the contrasts). It is good practice to be consistent with data handling.

R27:"P.8079, equations: highlight in the text that an additive change factor was used for temperature; multiplicative for precipitation."

A27: This has been added, however it is standard practice for delta change.

R28. "P.8078, l17-18: there's quite a lot of repetition in these two paras; merge and make more concise"

A28: OK

R29: "P.8079, l13: The text from "these time series of temperature. . ." onwards to the bottom of the section doesn't fit in the 2.3 sub-heading; I'd recommend turning it into a new section."

A29: OK

R30: "P.8079, line18-19: "In this way, INCA-P model deficiencies were removed". This is incorrect: (a) model deficiencies are not removed by doing this, the model is just as deficient in the future as it is for the baseline; (b) this assumes that the deficiencies are the same for the future period as for the baseline, which is not necessarily true. For example, in the future different processes may become more or less important, which may affect model deficiencies."

A30: Altered to "the likelihood of model deficiencies is minimised". The reviewer's comment is confusing, however. Parts a and b make contradictory statements ("the model is just as deficient in the future as it is for the baseline"; and "the deficiencies

HESSD

11, C4313–C4364, 2014

Interactive Comment

are [not necessarily] the same for the future period as for the baseline"). As we are using the delta change method, the model is precisely as deficient in the future as it is in the baseline. If the model over-predicts flow by 7% in the baseline, and is assumed to be just as deficient in the future, then as we are calculating percentage change in flow, this 7% over-prediction is negated, and the only changes we report are those as a response to climate alterations. To argue that model deficiencies might not be the same in the future –one could then equally say reporting of any model performance statistics during the calibration stage is pointless; as process interactions during the current period might not reflect those of the future. We must assume however, that the model performance over the current period gives some confidence in future runs (see comment 4).

R31: "P.8079, lines 19-24: sorry, I don't quite follow here. On first reading, I understood from this that one cdf had been plotted per variable (flow, TDP, etc.), taking the variability model output using the different ensemble members to get the cdf. However, this isn't the case as there's one cdf plot per ensemble member. So where is the population from? Different daily values? Would be good to make a bit clearer."

A31: Text clarified. The reviewer's initial interpretation was correct. In the majority of figures, one CDF has been plotted per variable (flow, TDP etc) using all ensemble members to give the CDF. This is the "model ensemble CDF". Figure 4, however, is an example of how both the ensemble average values and the uncertainty values were derived, and how the plots should be interpreted; thus it shows both the CDF ensemble (the thick black line), and the individual outputs from each ensemble (grey lines). "A" demonstrates the ensemble average projected change in temperature at the 90% probability level (thick black line). "B" demonstrates how to derive the uncertainty in temperature projections at the 90% probability level (i.e. the range between min and max ensemble members at that probability level. All values used within the CDFs represent monthly % change over the 30 year period (daily change values are not acceptable statistics). In summary, figure 4 illustrates the PPE ensemble plot, the associated uncertainty, and how the data was derived. The methods used closely follow those of UKCP09 (http://ukclimateprojections.metoffice.gov.uk/21680).

R32: “P.8079, l26 and p.8080, l3, p.8092, l2: variance used instead of difference. Variance has a precise statistical meaning.”

A32: Changed to difference

R33: “P.8080, l5-9: Delete; belongs in introduction/conclusion, but not in methods.”

A33: Section moved to introduction

R34: “P.8080, l12-19: Move to methods section; not results”

A34: Moved to methods

R35: “P.8080, l19 (and throughout the text from here onwards): ‘model coefficients’ is confusing terminology, replace with ‘model performance statistics’ or similar.”

A35: OK

R36: “P.8080, l21-24 and Fig. SI3: That doesn’t seem justification for not including the pefferlaw to me, as there are four monitoring points in that catchment. Therefore add it to Fig. SI3 for completeness.”

A36: Figure added (see figure 4 of additional material). It should be stressed that the spatial variability of accuracy within the Pefferlaw is within the range of that achieved in other catchments, and that the reason for not providing this information is simply that, with only three sites of long-term water quality monitoring for comparison (not 4 – the fourth is a gauging station where no chemical data was collected), we considered there to be no added value to presenting model accuracy for this catchment in a spatially distributed manner (as justified in the text). Unlike the Whites and Beaver, the strength of the Pefferlaw dataset lies in its temporal, rather than spatial, extent.

R37: “P.8082: this section (section 3.2) needs an introductory phrase or two to say why these are being calculated, and how this helps achieve the objectives of the study. Just by helping increase the credibility of the model? Could also be cut down."

A37: This section has now been merged with section 2.2.2 (model calibration) (see also comment 9). An introductory statement added. This section demonstrates whether model behaviours are plausible, and quantifies how much TP is being exported from each catchment (other data presented in the paper is given as % change from a baseline). This section places our study into a broader context – which is important for readers making comparisons to other studies or catchments. We also determine where the TP is coming from i.e source apportionment; which is important later in the manuscript.

R38."P.8082, l2: re-phrase as simulated average TP export coefficients for the calibration period."

A38: Adding "simulated" to the sentence is a form of reiterative redundancy - by definition all nutrient export coefficients are aerially weighted and thus are models or simulations of some kind or another.

R39: "P.8082, l6: were the previous studies of the catchments modelling or monitoring studies? Monitoring would be better."

A39: The quoted studies use monitoring data to derive their coefficients. However, see point 38: coefficients are not observed values. It is an expression of P export per unit area (an average). Whilst observed values can be used to calculate them, TP samples are rarely taken on a daily basis over a 30 year study period, and gap filling (averaging) is usually applied. A well calibrated process model, with a daily time series output, is arguably a more accurate method than annual statistical averaging (see response to comment 40).

R40: "P.8082, l12-18: These exports from the different land uses are dependent on how the different land use classes were parameterised in INCA. To make this section relevant, it'd be good to make clear here that the point is to determine whether the

HESSD

11, C4313–C4364, 2014

simulated export fluxes are realistic, rather than presenting them as useful new results."

A40: This comment has been taken into consideration (see comment 4) and model calibration section combined with nutrient export coefficient section. However, it is important to note that modelling of TP export coefficients are another example of how process models can be used to fill gaps in knowledge that observations cannot provide. As the reviewer rightly pointed out in comment 16, this is a function of inputs, processes and outputs – and the source apportionment of P exports to different landuse types could not be achieved through monitoring alone. The detailed nature of this P export data (landuse-specific) is new. As with any model outputs, the authors agree that model parameterisation was important. Hence the detail provided in section 2.2, and tables 1 and 2.

R41. "P.8082, l26-28 and p.8083, l1-3: Is this realistic? Any data? It's just a function of the phosphorus inputs and outputs over the year (which are all very uncertain and just a function of the model parameters used), so the point of this paragraph should be to show whether the model is reasonable or not, rather than just describing something that could be unrealistic."

A41: The references used demonstrate that the INCA model behaviours are exactly what would be expected of soils responding to an input of fertiliser, and shows that the model is responding reasonably (Haygarth et al., 1998; Borling et al,. 2003).

R42: "Throughout the results sections, it would be useful if the authors, when stating results that are interesting, referred to parts of the discussion in which these interesting results were then explained and discussed in more detail (and made sure there was some discussion of them somewhere in the discussion). E.g. p.8084, l7-8. A more structured discussion with sub-headings would be needed for this to work, but I think it would make the paper tie together better."

A42: The discussion has been restructured for clarity. We do not agree, however, with referring to sections of the discussion throughout the results. This will be confusing, and is not standard practise for this journal.

R43: “P.8084, l8: what does this mean? That 50% of the time flow increases by 23%? Is this a value from the median of the ensemble members?”

A43: No, please see table SI10. The value “23.37” is the comparison of flow projections between catchments. The paragraph summarises the data in table SI10 (which contains information from the CDF of an ensemble average of all QUMP members, and the uncertainty between members – see comment 31), showing that by 2070, there is a 50% chance that flow in the Holland and Pefferlaw will have decreased by up to 5.82 and 8.2%; and in the Whites and Beaver will have decreased by up to 29.19 and 20.35%. It comments that across the catchments, this is a large range of projected changes (23.37%) – which are significantly smaller in the Holland and Pefferlaw, and largest in the Beaver and Whites.

R44: “Section 3.4.1: re-structuring would be useful, starting with HER and SMD, and then looking at flow changes (which depend on HER and SMD). Sub-headings could help, and linking sentences describing (a) what the main change in climate change drivers is; (b) what the change in HER and SMD is, and whether this fits with the climate change drivers; (c) what the change in flow is, and whether this matches the changes in HER and SMD. It’s hard to extract this key information from the text as it is at present. Reducing reference to the cross-catchment variability would be useful (move to a table?).”

A44: This comment is confusing. Comparisons of the cross-catchment range (note: not variability) between different responses and climate drivers is the main focus of the manuscript. We have consistently followed a very clear structure for sections 3.3, 3.4.1, and 3.4.2. Each section starts with presentation of the CDF ensemble average and uncertainty for the relevant variable; continues on to discuss seasonal changes; and finishes with an analysis of sensitivity per unit change in driver. The current structure is concise, consistent, and easy to follow. We are sorry the reviewer did not see it.

HESSD

11, C4313–C4364, 2014

Interactive Comment

R45: “P.8085, l10: I disagree that the Pefferlaw is different to the Beaver and Whites. From Table 7, the Holland is the only odd one out. Subsequent discussion needs to be altered to reflect this.”

A45: Statistical significance of the difference between catchment responses cannot be analysed as there are insufficient data points. The authors are unsure, therefore, as to on what basis the reviewer states that the Holland is an “odd one out”? The original statement made is “the Holland and Pefferlaw yield the least HER in response to changes in precipitation, whilst the Beaver and Whites generate the most”. The Beaver and Whites both generate a very high HER output response to changes in precipitation (almost 1:1) at 0.94 mm. The Pefferlaw generates less, at 0.87, and the Holland 0.52. Whilst the Holland certainly generates the least, the authors maintain that their original statement is correct: the Beaver and the Whites generate the most. The discussion does not need to be altered. We hope that the additional columns in tables 7 and 9 (attached) will help to clarify this.

R46: “Throughout results section: likelihood is often used, when I think the authors mean probability.”

A46: This is more complicated that it at first appears. Probability is generally used by researchers in a statistical sense, involving mutually exclusive events, where all outcomes are accounted for. It has to be between 0 and 1, and all probabilities must add up to a total of 1. Statistical probability cannot be used in projecting climate change impacts, because it is impossible to account for all outcomes, due to the random and open character of natural systems. Likelihood is much weaker than probability; but essentially it gives the odds of an event given specific data. This is referred to in UKCP09 as “subjective probability”, defined as “an estimate based on the available information and strength of evidence” (http://ukclimateprojections.metoffice.gov.uk/21680). The authors are happy to change all references of likelihood to “subjective probability”, but would urge caution to the scientific community that this not be misinterpreted as a statistical probability, as is more commonly used.

R47: “P.8087, l17-14: It’s not quite clear what’s been done here. Was the daily time-series of TP concentration, averaged over ensemble members, taken as the starting point? In Table 9 the Beaver and Whites have massive increases of 0.2 to 0.5 mg TP/l with 1mm of rainfall in one season of the year. This needs highlighting and coming back to in the discussion.”

A47: Monthly averages are used, as advised by the IPCC (Carter et al., 2007). Daily changes are not appropriate in climate projections. This has been clarified in section 2.4. The method for the unit change sensitivity analysis was described in section 2.4 (originally pages 8079-8080 of the manuscript). The change in TP (mg/l) [between the baseline and future period] was divided by the change in precipitation (mm) [between the baseline and future period], giving an output of catchment response per unit change in precipitation. All values are calculated over 30 year averages, on a monthly basis. The model ensemble average was used, as this takes into account all the variations in uncertainty. Additional clarification has been provided in the methods (section 2.4).

R48: “P.8088, l21: the direction of change in projected HER and flow MUST have matched climatic drivers (precipitation, temperature), as that’s what forces them. Should this just say precipitation?”

A48: We strongly disagree with this statement. The direction of change of HER and flow does not have to match the direction of change of their climatic drivers. Climatic drivers have complex interactions with a variety of processes; for example in regions where plant growth is primarily temperature limited, warmer conditions can lead to a higher growth rates, and greater primary productivity. This can result in an increase in evapotranspiration which can offset an increase in precipitation, leading to an overall reduction in runoff. In this catchment, the interactions under discussion are during spring, and are related predominantly to ice and snow. This is a snowmelt driven catchment, and in winter and early spring, an increase in temperature results in a reduction in frozen water stores. This results in a reduction in spring melt. Therefore, despite an increase in precipitation in winter and spring, an increased soil moisture deficit (due to

HESSD
11, C4313–C4364, 2014

having no snow, and having removed the spring melt), results in an overall reduction in HER and flow. So – in spring, we have seen an increase in temperature, an increase in precipitation – but a decrease in HER and flow. The direction of change in hydrological response does not always match the direction of change in climatic drivers, when there are other significant (snowmelt) influences to consider. This is one of the reasons that detailed, process-based models are quite useful.

R49: "P.8089, l6-8: expand on this"

A49: Lines 9-24 do exactly that.

R50: "P.8089, l9: Model calibrations didn't demonstrate this, they were consistent with observations that." . .

A50: The model results do demonstrate this. As no observed time series data of soil TDP or labile P exist for these sites "consistency with observations" is not possible. It is exceptionally rare to find study areas with 30 years of observed TDP or labile P data. The literature, however, supports the model behaviours. See SI6

R51:"P.8091, l3: why?"

A51: Explained in lines 10-18.

R52: "P.8091, l8-19: again, not from my reading of the results (Pefferlaw has 0.87, which is much closer to 0.9 than it is to 0.6). All the subsequent discussion therefore needs altering."

A52: A line has been added in the discussion to clarify: "The hydrological sensitivity of the Pefferlaw is a little higher than that of the Holland, but is consistent with the difference in residence times and clay content between the two catchments (Table 2)." The hydrological sensitivity of the Pefferlaw is lower than both the Beaver and the Whites, but higher than the Holland. The difference between the Pefferlaw and the Holland is consistent with the difference in residence times and clay content (with the Pefferlaw having slightly shorter soil water residence times, and slightly higher clay content than the Holland). This is consistent with the discussion. Information has been added to tables 7 and 9 for clarity.

R53: “P.8092, l1: 0.14 mg/l, is this annual mean concentration?”

A53: It’s the average coefficient of TP per unit change in precipitation over a 30 year period. See table 9.

R54: “P.8092, l3: I’m not clear what’s meant by “act as a buffer to uncertainty”, again line 13.”

A54: To act as a “shock absorber”, a “cushion”, a “shield against”. . ..

R55: “P.8092, l24-25: results not presented to back this up. Note also that soil is only a small part of geology; bedrock differences are not discussed at all. Differences in P inputs and P saturation between catchments – I can’t find where that was mentioned in the results.”

A55: This a puzzling comment. The manuscript goes to great length to present results on climatic inputs, and sensitivity of the catchments to the changing drivers [which the reviewer has commented upon]. In terms of differences in P inputs – these are presented in table 2. Data relating to P saturation conclusions are presented in SI6. Information derived from the climate uncertainty analysis, catchment sensitivity analysis, and data derived from the model calibration, are combined in the discussion – together with existing literature - to ascertain possible reasons for specific hydrological and chemical responses. Additional data on clay content has been added to table 2 for clarity (and in tables 7 and 9). Differences between catchments in terms of runoff: and soil matrix flow were presented on page 8085, with significantly more HER flowing as surface runoff in the Whites and Beaver, than in the Holland and Pefferlaw.

The results in table 9 clearly indicate that timing of P export, combined with differences in runoff:matrix flow is associated with sensitivity to climate change (with the Holland and Pefferlaw being most sensitive during summer and autumn). The high P export during summer is a result of fertiliser inputs (shown clearly in SI6), where soil TP is highly mobile following this TP applications. There is little export in winter because the soil TP has been used up (SI6) and furthermore there is little contribution from surface runoff (page 8085). The spring/winter export in the Beaver and Whites is associated with the higher overland flow and macropore contributions, where TP is delivered directly to the streams via soils (page 8085). All of this data has now been combined and included in table 7 for clarity.

R56: "P.8094, l10: delete 'uncertainty in'."

A56: The sentence has been modified for clarity

R57: "P.8094, l17-18; 'catchment sensitivity to climate uncertainty was lower. . .'; presumably should read as catchment sensitivity to climate change?"

A57: The sentence has been modified for clarity

R58: "P.8094, l10-15: not backed up by results presented here."

A58: This again is very puzzling. These results are presented by the manuscript. The modification of table 7 should help to clarify this, and "geology" has been changed to "soil type", as requested earlier. See response to comment 55 on presentation of results

R59: "P.8095, l3-7: This doesn't make sense; how can hydrochemical model uncertainty affect catchment sensitivity to climate change?"

A59: To address this comment, we clarify the mechanisms at work here (see also comment 3). The manuscript assesses catchment sensitivity (i.e. phosphorus and hydrological responses to climate change – see comment 3) using a process based model. The response of the INCA models to climate drivers is dependent upon their calibration accuracy, and thus uncertainty in this calibration leads inherently to uncertainty in the results. The authors went to great lengths to calibrate the models using measured data; however there are always issues such as equifinality to consider. Additional information on the caveats of the study has been added, but we would ask the reviewer to carefully consider their interpretation of the manuscript.

Comments on the Tables and figures:

R60:"Table 1: Decrease precision to just one decimal place"

A60: OK

R61: "Table 2: Add groundwater TDP concentration, parameters relating to amount and timing of plant uptake. Round catchment area to the nearest km. Re-name the first column something like 'Parameter/Data type', as it is not just model parameters but also input timeseries. I don't understand the values for the first four rows of 'hydrological characteristics' – these are input timeseries, so what are the values? Means of some kind? For fertiliser inputs, make consistently to 1 decimal place (d.p.). Sewage inputs to 0 d.p. Define acronymms in table caption. The Beaverton is refered to as Beaver in the text."

A61: As requested, maximum plant uptake has been included in table 2, and discussed within the manuscript. However, this is a table demonstrating sources and values for measured or calculated input values within the calibrated INCA model, and inclusion of plant uptake values here is questionable (based on literature). Groundwater TDP is not a key model parameter (see comments 5 and 10), however it has now been added to table 2. For the reviewers' benefit, groundwater TDP data from the Provincial Groundwater Monitoring Network (http://www.ontario.ca/environment-and- energy/provincial-groundwater-monitoring-network-pgmn-data) was used to guide calibration of model parameters, and chosen inputs to the models ranged between 0.001mg/l (Beaver) and 0.007mg/l (Holland and Whites).

R62: "Table 3: Need better caption. No acronyms. Are these all the members? What are the ones in bold? What's sK? What's delta?"

A62: Here, it seems, is where the reviewer's confusion RE: comments 6 and 22 stems

HESSD

11, C4313–C4364, 2014

from. The ensemble members in bold indicate those used in the study, and clearly demonstrate a wide range of sensitivities. This has been clarified in the caption. The title has been clarified.

R63: "Table 4: Needs re-doing. Just providing locations with the best model performance statistics is not ok (cherry picking). Instead, replace with something like performance statistics for the worst and the best reaches for each study site, as well as for the catchment outflow. Please provide model performance statistics for daily data, as well as monthly and/or annual averages/loads if desired. Add in the number of observations and Nash Sutcliffe efficiency for comparability with other modelling studies. In the caption, replace 'model fit coefficients' with 'model performance statistics. Explain how the model error was calculated (difference of the means, i.e. bias or root mean squared error?)."

A63: The equation for model error has been added in the revised manuscript. Daily statistics are not appropriate for this study (see comments 9 and 31). The authors should not be accused of "cherry picking", having provided performance statistics for every reach in the catchment in Figure 6. This is a summary table of best performing reaches, and performance at the catchment outflow. At the reviewer's request, average performance statistics for the whole catchment are now also presented in this table.

R64: "Table 5: Reduce to 0 or 1 decimal places."

A64: OK

R65: "Table 6: Is the average uncertainty +/- the value given, or the width of the interval? I don't understand the units in this table (degrees C given for temperature; % for the rest). It could really help if there was a sentence in the figure caption explaining how this should be interpreted. E.g. "for the Holland sub-catchment, by 2030 precipitation simulations are +/- 19% of the ensemble average" (or whatever's correct). Decrease all to just 1 d.p."

A65: These are the units for which climate change are generally presented (o C for temperature, % change for precipitation and flow). Temperature cannot be reported as a % change due to discrepancies in units (Kelvin vs Celsius). The interpretation of "average uncertainty" is explained clearly on page 8079. It is a measure of "the mean value of uncertainty across all CDF probability levels". This reason for use of this metric can be better understood by looking at table SI7. It can be seen that at different probability levels, the precipitation uncertainty in, say, the Whites, is very different (ranging from 16.4 to 24.8% in 2030). The lowest uncertainty here is at the 10% level. However, the lowest uncertainty in the Beaver projections is at the 50% level. So – it would be quite misleading to present only the ensemble average value and its corresponding uncertainty value at the 50% level. It might lead us to think that the data did not have many uncertainties within it, when in fact some catchments are very poor at conveying extreme change (10 and 90%). To better present the overall uncertainty within the different catchments, then, uncertainty was calculated at every probability level within the CDF plot, and an average taken. The results in Table 6 don't mean that the simulations are +/- x% of the ensemble average, because this is an average of uncertainty taken across the CDF. Table 6 gives an overall uncertainty response of the PPE for each catchment and study variable, providing an effective summary for the reader. The more specific data the reviewer is looking for is provided in SI7. Table 6 is a more of a mechanism for comparison between catchment responses.

R66: "Table 8: Add dates for future periods on left hand side. Decrease all to just 1 d.p."

A66: OK

R67: "Table 9: Clarify the caption – is this averaged over one year (2030), or a 30 year period?"

A67: 2030 refers to the period 2020-2049. This was explained in the methodology. An "'s" was missing from the caption in table 9 however and has been provided.

HESSD
11, C4313–C4364, 2014

R68: “Fig. 1: Define acronyms used in legend in the caption. Use of colour in catchment boundaries isn’t good as they overlap. Annotate instead? Don’t need central points of RCM squares marked. Hard to pick out sub-catchments with selected RCM squares in grey; maybe try highlighting in some other way (e.g. bold edges). Not sure what the word ‘analysis’ refers to in the figure caption.”

A68: Figure has been adjusted for clarity

R69: “Fig. 2: Suggest deleting this figure and just giving statistics (average difference, or similar).”

A69: See comment on figure 3

R70: “Fig. 3: In caption, say what the Qs are (selected members of the PPE). The use of Q is a bit confusing, as it makes me think of quantiles, so need to be clear about this throughout.”

A70: We would prefer to keep figure 2, if we are keeping figure 3. It would be inconsistent to present just the precipitation and not the temperature data. Q has been clarified in the caption

R71: “Fig. 4: Not clear what data each line is representing. Daily values? E.g. should this be interpreted as 90% of days have a temperature change less than or equal to 3.2C? An an example of how these plots should be read would be great.”

A71: Figure 4 (of the manuscript) is an example of how the plots should be read. “A” demonstrates the ensemble average projected change in temperature at the 90% probability level (thick black line). “B” demonstrates how to derive the uncertainty in temperature projections at the 90% probability level i.e. the range between min and max ensemble members at that probability level. Temperature changes are monthly averages over 30 year periods.

R72: “Fig. 5: Give units.”

HESSD

11, C4313–C4364, 2014

Interactive Comment

A72: OK

R73: "Fig. 6: What is each point? Mean over whole model run? Mean of annual means?"

A73: Clarified in caption

R74: "Fig. 7: See comment on Fig. 4, and amend fig caption."

A74: OK

R75: "Figs 8 and 9: Merge into one figure. Define acronyms in figure caption."

A75: They would not be suitable as one figure – the level of detail would be too high for publication.

R76: "Fig. 10: Replace 'QUMP' with 'ensemble', or define QUMP. Are TP concentrations daily or mean monthly or seasonal? If true, say that there is one box per ensemble member."

A76: OK

R77: "All SI Figure: resolution needs increasing."

A77: OK

R78: "Fig. SI2: Define acronyms within the figure caption."

A78: OK

R79: "Fig. SI3: Why is this schematic, rather than a simple realistic map for each study catchment with the sub-catchments and reaches marked on? Why do some of the reaches appear to not connect to the main stem? Please add a scale bar for each catchment."

A79: Scale bar added. This is a model schematic because a catchment map has already been provided. It should be noted that a) model schematics give more information as to how the model is used to represent the study catchments b) provide a more direct comparison between the availability of observed (monitoring) data, and model sub-catchment calibrations. It is more difficult to interpret this data from maps derived from satellite data. We have provided such a map for comparison (Figure 5 of the attached document) and are happy for either map to be used – but do feel the original provided greater clarification on model structure. In relation to the "disconnection" of reaches in the schematic, this is simply a function of the INCA graphical user interface - these reaches are in fact connected to the main stem of the model.

R80: "Figs. SI4 and SI5: In the caption, need to say that Q0 to Q15 are ensemble members. Delete 'applied to the observed data'."

A80:OK

R81: "Fig. SI6: Which study area? Which sub-catchment? Which time period? What do the boxes represent – variability in daily labile P pools for one sub-catchment? If so, why present as boxplots rather than as a timeseries?"

A81: This is the average for intensive agricultural areas throughout all of the Holland catchment (i.e average of all subcatchments, not a single sub catchment). In both plots, monthly averages were used over the calibration period. This has been clarified in the caption.

R82: "Table SI7: Decrease to 1 or 0 decimal places."

A82: OK

R83: "Fig. SI8 and 9, 11, 12: Define QUMP and what Q0, Q3,. . . are in the figure caption. Table SI10: This is a key table, so put in the main text, not the SI. Could be combined with Table 6. Decrease to 1 d.p. Make clear what these probability levels mean (number of days with up to this change?)"

A83: The subjective probability levels are to be interpreted as explained in section 2.3 (now section 2.4: data analysis), whereby 10%, 50% and 90% is the likelihood of change being less than a certain amount, having accounted for projections from all ensemble members. For example, in the Holland the temperature result at the 50% likelihood level is 2.55; this means that by 2030 there is a 50% likelihood that temperature change in any month over that 30 year average period is going to be less than 2.55o C. At the 90% likelihood level the temperature value is 5.03, and means that by 2070 there is a 90% likelihood that in any month over that 30 year average period, temperature will have changed by less than 5.03o C.

R84: “Table SI13: Is this the mean of the ensemble members? Is it monthly TP loads and monthly average concentrations? This is as important as table 8 in the main text; suggest moving from the SI to the main text.”

A84: Caption clarified

Reviewer 2:

We would like to thank the second reviewer for the analysis of the manuscript. The study is indeed ambitious, being the next step forward in providing practical applications of data-intensive perturbed physics ensembles for catchment management (Collins et al., 2006). We included the more intensive data in the SI, so as to enable others to replicate our methods and expand upon the work; and aiming to leave the main manuscript fully accessible to those for whom the core conclusions from the study will be generally interesting.

R1: “Is the supplementary really needed? It looks like some of the results are now presented twice, first in the main text and then in supplementary file, e.g. Figure 7 and SI7. “

A1: Information within the supplementary material is not critical to the understanding and interpretation of the manuscript. This is why it is included in the supplementary section, and not within the main text. It is true that there is some overlap between the data in figure 7, and that within SI7 – however, overall the two present quite different

HESSD

11, C4313–C4364, 2014

Interactive Comment

results. Figure 7 gives the CDF of each individual catchment, whereas the data in SI7 gives a summary of values at key percentile intervals, and the uncertainty associated with each of those percentiles. To demonstrate all of this information on one figure (different catchments and associated uncertainty) would have been too confusing – indeed, reviewer 1 was confused with just the single CDF. The table of highlighted percentiles and their uncertainties in SI7 is not critical to the paper (the numbers are quotes individually in the text), but the overall table does add context for those who wish to investigate the results more fully.

R2: "The first too pictures in supplementary are already published, and the third one is really difficult to understand. Instead would be nice to have a map which shows also location of agricultural land, wetlands and artificial areas."

A2: The comment is understandable, but SI 1 and 2 were originally drawn by the lead author, and both have been adapted for this publication. Since their original publication, INCA-P has been modified, and an adapted schematic was required. SI3 has been redrawn at the reviewer's request (see figure 5 of the attached document). The absolute locations of different landuse classes are not included on the figure – the ecological land classification of Ontario is highly detailed, and does not improve clarification of the site figures. As percentage land cover has already been quantified (Table 1 of manuscript) this would not add any scientific value.

R3: "On page 8070, line 5. Eutrophication is also other harmful aspects than reduction of oxygen. In worst case it alters the whole ecosystem."

A3: Altered to "it can result in eutrophication, where death and decay of excess algal matter leads to reduction in stream oxygen concentrations (Nicholls, 1995; Jarvie et al., 2006) which affects fish spawning and survival (Evans, 2011). In addition, the composition of algal species may be altered, and blue-green algae (e.g. cyanobacteria), can become dominant. These species may produce toxic compounds, which are harmful to terrestrial and aquatic animals (Chorus and Bartram, 1999), resulting in high water

HESSD
11, C4313–C4364, 2014

quality treatment costs (Smith, 2003)."

R4: "Only part of the PP is converted to bioavailable P, e.g. Hartikaninen et al., 2010"

A4: Absolutely. We did not intend to suggest that all PP is converted to DP. This has been rephrased for clarification.

R5: "Page 8071, line 6. Are unknown physics of climate processes really only parameter un-certainty? Not a system uncertainty etc? And parameter uncertainty comes then from model description and measurements, where measured value does not completely de-scribe what it is supposed to describe."

A5: Absolutely, though we may be speaking at crossed-purposes. We say that parameter uncertainty stems from the unknown physics of climate systems, ie. that the unknown physics "introduces a source of uncertainty....known as parameter uncertainty". It is not claimed that parameter uncertainty causes system uncertainty. This can be rephrased for clarification. It should be noted that these are perturbed physics ensembles – not just perturbed measurement ensembles. Here, perturbations are made not just to individual measurements of existing model functions, but to the physical structure of the model itself – so as to explore the impact of unknown physics on our certainty of projections (Collins et al., 2006). Within GCMs, the physical functions controlling the systems are generically referred to as "parameters".

R6: "In methods section would be to have description of water quality measurement, where, how often and the analysis methods."

A6: TP samples were analysed colorimetrically following digestion. These are long term records, amalgamated from a number of sources (table 2 in the manuscript) and different providers collected samples at different frequencies; this varied from twice-weekly to monthly, and included both event-based and routine sampling. These details are now included in the manuscript.

R7: "Also, some sentences of agriculture, as it covers the main land use. What is the main crop, what is the growing season. And especially, how is the climate change assumed to affect crop and growing season”

A7: Major crops in the area are alfalfa, corn (for grain), and soybeans (Statistics Canada, 2008). Livestock raised in the area are primarily poultry and cattle. The growing season for individual farms is not recorded, but are recommended by OMAFRA (as quoted in the manuscript), and it was assumed that these recommendations were followed. INCA simulates a plant growth index, related to seasonal variation in solar radiation, which in combination with a user-specified growing season, is used to quantify nutrient uptake by plants. Should the climate become warmer, the amount of nutrients taken up could increase. Whilst a longer growing season might become possible, it is not necessarily true that it would be permitted. This would be a Ministerial decision, as Simcoe is currently an area of concern with regards to P loads, and has in place significant nutrient reduction targets. As this is a physical assessment of the primary impacts of climate upon hydrology and water quality (and the different sensitivity of individual catchments), the secondary impacts (such as OMAFRA decisions on alterations to growing season, and subsequent choices by farmers as to changes in crops) are not incorporated; such assumptions (which would include the need for socio-economic projections of population growth) would lead to modeller bias in the interpretation of results.

R8: “Page 8074, line 5. What is the annual P? “

A8: Annual average in-stream concentrations range from 0.026mg/l in the Beaver to 0.142mg/l in the Holland. It is difficult to give comparative “true” loads, as the rivers are not monitored over the same time periods, nor at the same sampling frequency. The average annual P export (kg/km2) is given in table 5 of the manuscript.

R9: “Reference to the HBV model is missing”

A9: The version used was the Nordic HBV model (Saelthun, 1995). The reference has been added to the manuscript. References

HESSD

11, C4313–C4364, 2014

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Borling, K. 2003. Effects of long-term inorganic fertilisation of cultivated soils. Swedish University of Agricultural Sciences, Uppsala. Doctoral Thesis.

Brown, C., and R.I. Wilby. 2012. An alternate approach to assessing climate risks. EOS Transactions, American Geophysical Union. 93 (41): 401-412

Carter, 2007. General guidelines on the use of scenario data for climate impact and adaptation assessment, version 2. Task Group on Data and Scenario Support for Impact and Climate Assessment (TGICA) IPCC Report

Cochrane, T.A., and Flanagan, D.C. 1999. Assessing water erosion in small watersheds using WEPP with GIS and digital elevation models. Journal of soil and water conservation 54 (4): 678 - 685

Collins, M., Booth, B.B.B., Harris, G.R., Murphy, J.M., Sexton, D.M.H. and Webb, M.J. 2006. Towards quantifying uncertainty in transient climate change, Clim Dynam. 27: 127 – 147

Chorus I, Bartram J, Eds (1999): Toxic cyanobacteria in water - A guide to their public health consequences. E and FN Spon, London, England

Crossman, J., Futter, M.N., and Whitehead, P.G. 2013. The significance of shifts in precipitation patterns: modelling the impacts of climate change and glacier retreat on extreme flood events in Denali National Park, Alaska. PLOS ONE 8 (9) e74054

Haygarth, P.M., Hepworth, L. and Jarvis, S.C. 1998. Forms of phosphorus transfer in hydrological pathways from soil under grazed grassland. Eur J Soil Sci. 49: 65 – 72

Janssen, P.H.M., and Heuberger, P.S.C. 1995. Calibration of process-oriented models. Ecological Modelling. 83 :55 – 66

Jarvie, H.P., Neal, C. and Withers, P.J.A. 2006. Sewage-effluent phosphorus: a greater risk to river eutrophication than agricultural phosphorus? Sci Total Environ. 360:246–53.

Jennings, E., Allott, N., Pierson, D.C., Schneiderman, E.M., Lenihan, D., Samuelsson, P., and Taylor, D. 2009. Impacts of climate change on phosphorus loading from a grassland catchment: implications for future management. Water Reserch 43 (17): 4316 - 4326

Jin, L., Whitehead, P.G., Baulch, H.M., Dillon, P.J., Butterfield, D., Oni, S.K., Futter, M.N., Crossman, J., and O'Connor, E.M. 2013. Modelling phosphorus in Lake Simcoe and its subcatchments: scenarios analysis to assess alternative management strategies. Inland waters - Journal of the International Society of Limnology 3: 207 – 220

Kaushal, S.S., Mayer, P.M., Vidon, P.G., Smith, R.M., Pennino, M.J., Newcomer, T.A., Duan, S., Welty, C., and Belt, K.T. 2014. Land use and climate variability amplify carbon, nutrient and contaminant pulses: a review with management implications. Journal of the American Water Resources Association 50 (3): 585-614

Larssen, T., Hogasen, T., Cosby, B.J. 2007. Impact of time series data on calibration and prediction uncertainty for a deterministic hydrogeochemical model. Ecological Modelling. 207: 22-33

Lemunyon, J.L, and Gilbert, R.G. 2011. The concept and need for a phosphorus assessment tool. American Society of Agronomy doi:10.2134/jpa1993.0483

McDowell, D., Sharpley, A., and Folmar, G. 2001. Phosphorus export from an agricultural watershed. American Society of Agronomy 30 (5): 1587 – 1595

McFarland, A.M.S. & Hauek, L.M. 2001. Determining nutrient export coefficients & source loading uncertainty using in-stream monitoring data. Journal of the Amer. Water Res.Assoc. 37: 223-236.

Ministry of Environment, 2006. Clean Water Act. S.O. C.22. Accessed online at www.ontario.ca/ministry-environment

Nazemi, A., and Wheater, H.S. 2014. Assessing the vulnerability of water supply to changing streamflow conditions. EOS. 95 (32): 288-289

Nicholls, K.H., 1995. Some Recent water quality trends in Lake Simcoe, Ontario: implications for basin planning and limnological research, Can Water Resour J. 20 (4), 213–226.

Oreskes, N., Shrader-Frechette, K., Belitz, K. 1994. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. Science 263 (5147): 641-646

Refsgaard, J. C., Madsen, H., Andréassian, V., Arnbjerg-Nielsen, K., Davidson, T. A., Drews, M., ... & Christensen, J. H. (2014). A framework for testing the ability of models to project climate change and its impacts. Climatic Change,122(1-2), 271-282.

Reid., K.D. 2011. A modified Ontario P index as a tool for on-farm phosphorus management. Ontario Ministry of Agriculture, Food and Rural Affairs, Stratfor, Ontario.

Saelthun, N. 1995. Nordic HBV Model. Norwegian Water Resources and Energy. Administration

Saleh, A., Arnold, J.G., Gassman, P.W., Hauck, L.M., Rosenthal, W.D., Williams, J.R., Mcfarland, A.M.S. 2000. Application of SWAT for the Upper North Bosque River Watershed. Transactions of the ASAE 43 (5), 1077-1087

Smith, V.H. 2003. Eutrophication of freshwater and coastal marine ecosystems. Environmental Science and Pollution Research. 10 (2): 126 – 139

Smith, G.J. 2010. Deriving spatial patterns of severe rainfall in Southern Ontario from rain gauge and radar data. A thesis presented to the University of Waterloo. Waterloo, Ontario.

Teutschbein, C., and Seibert, J. 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different models. Journal of Hydrology 456 – 457: 12 - 29

Wade, A.J., Hornberger, G.M., Whitehead, P.G., Jarvie, H.P., and Flynn, N. 2001. On modelling the mechanisms that control in-stream phosphorus, macrophyte and

HESSD

11, C4313–C4364, 2014

epiphyte dynamics: an assessment of a new model using general sensitivity analysis. Water Resources Research 37 (11): 2777 – 2792 Weglarczyk, S. 1998. The interdependence and applicability of some statistical quality measures for hydrological models. Journal of Hydrology. 206 (1-2): 98-103

Please also note the supplement to this comment:
http://www.hydrol-earth-syst-sci-discuss.net/11/C4313/2014/hessd-11-C4313-2014-supplement.pdf

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 11, 8067, 2014.