Hydrol. Earth Syst. Sci. Discuss., 11, C2992–C3008, 2014
www.hydrol-earth-syst-sci-discuss.net/11/C2992/2014/

Hydrology and Earth System Sciences
Discussions

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.

Anonymous Referee #1

Received and published: 8 August 2014

The paper has two stated objectives: (1) to compare the sensitivity of four sub-catchments of Lake Simcoe, Canada, to climate change (in terms of their hydrological and water chemistry response); and (2) to generate estimates of future water quality under scenarios of climate change, accounting for uncertainty in General Circulation Model parameters. The approach to answering these objectives is to take five climate change scenarios, representing a range of outputs from a 17-member Perturbed Physics Ensemble (PPE) for one Global Circulation Model (GCM). Four sub-catchments of Lake Simcoe (Canada) are then described, generally qualitatively. For

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

each sub-catchment, the five scenarios are then run through a process-based hydrological model (HBV) to derive input hydrological timeseries for a process-based phosphorus model, INCA-P. Changes between baseline and future climate variables (temperature and precipitation), input timeseries for INCA-P (hydrologically effective rainfall and soil moisture deficit), and output from INCA-P (discharge and total phosphorus concentrations and loads) are then described for each sub-catchment, and contrasted between catchments. The qualitative comparison of catchment responses is used to support the conclusion that clay-rich catchments are likely to be more sensitive to climate change.

Unfortunately, despite the considerable amount of work which has clearly gone into this paper, there is not much that is novel, and much that is not considered. On top of this, a number of serious flaws in the methodology lead me to recommend at best major revisions. I have provided a rather long list of comments and suggested improvements below, in the hope of providing some guidance on how the analysis could be made more robust, and how the authors could demonstrate that the work is worthwhile.

Major comments:

1. 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

HESSD

11, C2992–C3008, 2014

climate change. There needs to be much more acknowledgement of this both in the intro and the discussion and conclusion, and more/better justification for carrying out the work in the first place.

2. 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?

3. 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?

4. 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.

5. Section 2.2 is lacking lots of detail on model calibration, including (i) calibration

HESSD

11, C2992–C3008, 2014

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?

6. 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?

7. 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.

8. 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

HESSD
11, C2992–C3008, 2014

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.

9. 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

HESSD

11, C2992–C3008, 2014

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.

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

11. 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.

12. 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).

13. 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 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.

14. 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,

HESSD

11, C2992–C3008, 2014

soils, etc. could be useful, plus some mention in the text.

15. 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.

16. 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.

Minor comments:

HESSD

11, C2992–C3008, 2014

1. Introduction: Confusing absorb and adsorb several times

2. P.8071,l23: geological (i.e. bedrock) differences between catchments aren't mentioned, only differences in drift and soils.

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

4. 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.

5. 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

6. P8074 l25: the use of the word 'parameters' is confusing. Replace, e.g. fluxes, variables,. . .

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

8. P.8074, l26: soil export coefficient isn't an output. Replace with soil erosion, this is what's meant.

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

10. Model calibration: How many parameters requiring calibration does HBV have? How were these calibrated?

11. 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?

HESSD

11, C2992–C3008, 2014

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

13. 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.

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

15. P.8076, l14: from Table 2, I see that septic inputs were classed as inputs to non-intensive agriculture. Justify this in the text.

16. 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.

17. 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?

18. 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?

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

20. 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.

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

22. 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.

HESSD

11, C2992–C3008, 2014

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

24. P.8077, l27: add ‘members’ after ‘ensemble’

25. 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.

26. 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.

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

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

29. 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.

30. 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.

31. 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.

HESSD

11, C2992–C3008, 2014

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

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

34. P.8080, l12-19: Move to methods section; not results

35. P.8080, l19 (and throughout the text from here onwards): 'model coefficients' is confusing terminology, replace with 'model performance statistics' or similar.

36. 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.

37. 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.

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

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

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

41. 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.

HESSD

11, C2992–C3008, 2014

42. 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.

43. 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?

44. 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?).

45. 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.

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

47. P.8087, l17-14: It's not quite clear what's been done here. Was the daily timeseries 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.

HESSD

11, C2992–C3008, 2014

48. 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?

49. P.8089, l6-8: expand on this

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

51. P.8091, l3: why?

52. 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.

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

54. P.8092, l3: I'm not clear what's meant by "act as a buffer to uncertainty", again line 13.

55. 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.

56. P.8094, l10: delete 'uncertainty in'.

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

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

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

Comments on the Tables and figures:

HESSD

11, C2992–C3008, 2014

Table 1: Decrease precision to just one decimal place

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.

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

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?).

Table 5: Reduce to 0 or 1 decimal places.

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.

HESSD

11, C2992–C3008, 2014

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

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

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.

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

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.

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.

Fig. 5: Give units.

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

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

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

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.

All SI Figure: resolution needs increasing.

Fig. SI2: Define acronyms within the figure caption.

HESSD

11, C2992–C3008, 2014

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.

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

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?

Table SI7: Decrease to 1 or 0 decimal places.

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?)

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.

References

Crossman, J., Futter, M. N., Oni, S. K., Whitehead, P. G., Jin, L., Butterfield, D., Baulch, H. M., and Dillon, P. J. (2013). Impacts of climate change on hydrology and water quality: Future proofing management strategies in the Lake Simcoe watershed, Canada. Journal of Great Lakes Research 39, 19-32.

Dunn, S. M., Brown, I., Sample, J., and Post, H. (2012). Relationships between climate, water resources, land use and diffuse pollution and the significance of uncertainty in climate change. Journal of Hydrology 434–435, 19-35.

HESSD

11, C2992–C3008, 2014

Fung, F., Watts, G., Lopez, A., Orr, H., New, M., and Extence, C. (2013). Using Large Climate Ensembles to Plan for the Hydrological Impact of Climate Change in the Freshwater Environment. Water Resources Management 27, 1063-1084.

Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., and Veith, T. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50, 885-900.

Refsgaard, J. C., and Henriksen, H. J. (2004). Modelling guidelines—-terminology and guiding principles. Advances in Water Resources 27, 71-82.

Refsgaard, J. C., Madsen, H., Andréassian, V., Arnbjerg-Nielsen, K., Davidson, T. A., Drews, M., Hamilton, D. P., Jeppesen, E., Kjellström, E., Olesen, J. E., Sonnenborg, T. O., Trolle, D., Willems, P., and Christensen, J. H. (2014). A framework for testing the ability of models to project climate change and its impacts. Climatic Change 122, 271-282.

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