Hydrol. Earth Syst. Sci. Discuss., 6, C1087–C1091, 2009
www.hydrol-earth-syst-sci-discuss.net/6/C1087/2009/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data" *by* H. M. Holländer et al.

Anonymous Referee #2

Received and published: 6 June 2009

This paper addresses scientific questions regarding our abilities to predict discharge in ungauged basins, and is thus very relevant for this journal. 10 different hydrological models are used in 'ungauged' mode on the artificial Chicken Creek, which is a laudable activity and the results of such an experiment would have the potential to be very very important in catchment hydrological science. However, I was disappointed by this paper on a number of counts and think that a lot more work needs to be done to bring the exercise up to an acceptable standard. I agree with Anonymous Referee #1 that it is an interesting paper but i DO NOT agree that it is enlightening in its current form. The major criticisms I have (in no particular order) are:

1. The aim of this paper it seems is to compare underlying model assumptions in the

prediction of ungauged basins, which is the first step of a 3 step exercise. However, surely there is an essential missing first step here, where a design rainfall and design catchment are used (which could of course be based on the dimensions and characteristics of Chicken Creek). Then there would be a benchmark set of results with which to compare the next step with, where the real catchment setup is attempted. This would be an easy and essential first step to make.

2. There are effectively 2 things being tested in this paper at the same time: effects of (i) model structure (ii) modeller practice. This would be an obvious separation in terms of an introduction and discussion at the end.

3. Where is the discussion? The discussion section is just a description of results. Please think about the implications of this work. Again there is a huge potential for a hard hitting paper here.

4. Where is the sensitivity analysis!!? There are some suggestions of fiddling around with some model parameters, setup decisions to achieve more reasonable results, but surely an essential part of such an intercomparison exercise is (preferably a formal quantitative) sensitivity analysis of model parameters, initial conditions, setup decisions etc. Poor modelling practice. Some of the models (especially the physically based ones) are overparameterised, so sensitivity analysis is essential. How can you draw conclusions about the effects of model structure/modeller decisions on discharge predictions without sensitivity analysis? How can we learn from this exercise otherwise?

5. Why was only one set of results per model shown in results? E.g. why weren't monte carlo simulations run from feasible parameter space. I have first hand experience of running physically based catchment models and it is just not that difficult with current computational resources. Again, this paper displays poor modelling practice. With such a multi simulation approach some estimate of the range of feasible results could have been presented and discussed (uncertainty). Then a really useful comparison of the

different ranges provided by the different model structures could have been discussed. I think the conclusions may have been very different had this been done. I would suggest all future attempts at this PUB exercise at Chicken Creek are designed more carefully to take account of this.

6. Where is the measurement error in the catchment discussed (especially as this is a back-calculated value)! How can you compare absolute model results with absolute model predictions without taking into account the uncertainty?

7. Results are poorly presented. Please show me some graphical statistics – range of results for Q95, Q50, Q5, particular events etc.; deviation from benchmark. The results section is very dense to read and more graphical figures of any nature would be a benefit. By the very nature of this intercomparison experiment several figures are required to fully understand what's going on!

8. As already noted by Reviewer #1, English is clumsy in places. Please get manuscript checked by native speaker.

9. Terminology. Throughout the manuscript poorly defined terminology is used. E.g. Discussion of model validation (p.3203, l.5): Mostly the term 'validation' is not used now and is replaced with 'model performance'. Explain why you discuss validation. E.g. you need to discuss how you are defining 'physcially based' and 'process based' models. What is the difference (sect 3.2)?

10. Section 3.2 seems to have been written by different people- each section needs to be in the same format and written in the same way (see point above about using coherent terminology). Although several tables have been provided later I think a simple figure comparing e.g. y-axis: conceptual-physically based and x-axis dimensionality, would place these models in some kind of context.

11. Title is slightly misleading. This paper is about using the catchment in 'ungauged' mode and i think the title should focus on this, rather than using term 'sparse'?

12. The manuscript is incoherent in places. Not only is the discussion missing (and already noted above) but methodology is found in the results section and results are found in the discussion section. E.g. correction strategy for precipitation input is methodological (p. 3223). Overall the logic of the structure comes across as confusing. I was constantly turning pages to find the information that i needed to understand what was going on. I think a major restructure is required.

13. It would be very useful to know the computational time needed for this exercise by the different models (preferably using comparable computational resources, or failing this just a description of what resources were used). This is a very important consideration for modellers attempting any exercise and especially for intercomparison exercises. In addition what time estimates are there for model setup and testing time?

14. Ksat parameter p. 3225. Good that the ranges of Ksat used are documented here, however as most modellers know very well Ksat is usually a very sensitive parameter and is certainly not interpretable as a 'physical' value. Ksat is an effective parameter which is known to e.g. vary with grid cell size. Thus the ranges discussed here are dependent upon the model used as well as other variables (such as grid size). Without a benchmark sensitivity analysis it is impossible to draw the conclusions that are attempted here.

15. Why was SWAT used if it isn't normally used for small catchments (p. 3227)

16. Results. Needs to be separate sections on for example (i) what are the results – i.e. what are the range of overall discharge predictions (ii) how these compare to observations – thus model performance (iii) why this might be (links to representation of hydrological processes and model structure – which information should already have been given in setup section!)

17. The decision making of the modellers needs to be more explicitly looked at. In fact i think that this is such a valuable part that it has the potential to be a paper on its own: I would suggest using social science expertise to look at not only the obvious

decision making but also some of the tacit assumptions of the practitioners in this exercise. I would encourage the authors to look into this. This would enable not only my queries about particular points in this paper (e.g. why did the hydrus-2d modeller use a daily input time series – i can't accept that computational time was enough of an issue to warrant this decision and if it was based on a hydrological decision then it needs explaining) to be answered, but also provide a broader look at the practice of hydrological modelling which is a very valuable exercise (similar exercises are currently underway for flood forecasting and climate modelling for example).

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 6, 3199, 2009.