Draft: Response to the editor (deadline Oct.18, 2013)
Edited by G. Blöschl, H. Flühler, H. Holländer, H. Bormann, G.B. Chirico, T. Blume, J.-F.Exbrayat
In italics (font 10): Points raised by the reviewers
In regular fonts (plain text font 11): Response by author and co-authors

Interactive comment on “Impact of modellers’ decisions on hydrological a priori predictions” by H. M. Holländer et al.

Anonymous Referee #2

Received and published: 20 September 2013

Below review comments under inverted commas, response in plain text.

“1) I guess it is now too late to comment on the study design, but for completeness I will make this comment anyway. First of all, I seriously question whether this problem can even be related to the "predictions in ungauged basins" problem, considering that this is a totally artificial, almost unique catchment. Almost all of the models used here derive their origins to their development and application in a particular climatic or landscape context, and when you extrapolate to some other place you assume implicitly some kind of similarity: this completely broke down here.”

We are not sure why one would want to **relate** the problem of the paper to the predictions in ungauged basins problem: it **is** a prediction in an ungauged basins problem, i.e. estimating runoff in the absence of stream flow data for the Chicken Creek. Perhaps what the reviewer is implying here is that he/she does not consider the Chicken Creek representative of natural catchments. Representativeness is an issue, of course, and one has to ask in what way would a natural catchment be different from an artificial catchment that affects model performance? Our view is that this depends on the model type. There are some models that are based on laboratory scale equations (such as Topmodel) and if the assumptions (such as exponential decay of conductivity with depth) apply they should be equally applicable to natural and artificial catchments; in fact, they should be more suitable for artificial catchments given assumptions such as soil spatial homogeneity. There are other models that do not use laboratory scale equations but imply some sort of organisation of the catchment structure as a result of the co-evolution of climate, landscape, vegetation and soils. We agree that the artificial catchment (given the limited time allowed for landscape evolution) differs from natural catchments. But then catchments in different climates would also exhibit different structures and mechanisms at the catchment scale (e.g. infiltration/saturation excess) and it is a fact that many models are used outside the location/climate for which they were developed. Perhaps there is too much of a practice of using models for climates/soils they have not been developed for (and this may be a problem), but this is certainly the current state of the practice. We believe that the conditions of the Chicken Creek are within the spectrum of catchments that can be encountered anywhere in the world, so we would not expect its representativeness to be fundamentally different from the case when extrapolating a model to completely different climate/soil conditions.

“2) Given the nature of the catchment (small, artificial etc.) I have to appreciate the courage (or bullheadedness) of the participants who volunteered to use TOPMODEL or SWAT (or even any of the other models). They had "no chance in hell" of getting the runoff responses right in this case. What thought processes were gone through in choosing to apply these models? I can only think that they were chosen, in a perverse way, to show hydrological models and modelers in a bad light.”

We object to the language the reviewer is using and have difficulties in understanding the reviewer’s frustration. What the reviewer seems to imply is that, with the right model structure, predicting runoff would have been straight forward but with the unsuitable model structure actually used, the model results will be far off the target, whatever parameters are chosen. We do not think this is true. The models used in the comparison are all quite flexible because of their high dimensionality. Virtually any hydrological response can therefore be simulated with “appropriately selected” parameters. Several studies have proved how model parameters can compensate for model structural errors, often by assuming unrealistic values (not consistent with the conceptual framework) during parameter estimation procedures (e.g. Clark and Vrugt, 2006).

We therefore think that, at the end, the main differences are due to model parameter choice (in the context of each model) and the parameters have been subjectively chosen by the modellers. It would be straightforward to fit all models to the observed hydrographs with close to perfect match but we think this is not needed as the high dimensionality implies that this is possible.

We do not claim that all the model structures are perfectly suited to the Chicken Creek and we will highlight this more clearly in the revised manuscript. The reality is that the modellers who were volunteering for participation were given the freedom to use any model of their choice, and they opted for the models as presented in the paper. It appears that the reviewer, if tasked with the same problem of estimating runoff in the Chicken Creek, would have chosen yet another model. All participants of the study have chosen different models so this would be expected and is not inconsistent with the design or the conclusions of the study. There is a story to be learned that models are often selected on the basis of familiarity with the code rather than on the basis of model assumptions, in particular if the exact runoff processes are unknown.

“3) Now I come to the main criticism of the paper. Really, I did not get anything useful from this paper. This way of idiosyncratic, anecdotal presentation is totally unhelpful. My problems come from two sources: (i) I do not know much about the catchment (no information is provided to guide me, and (ii) I do not know much about the idiosyncracies of the various models. How am I expected to read through page after page of details (this model, this process, this parameter etc....) and still gain some wisdom out of the paper. All this material should appear on the first day of a brainstorming session on model inter-comparisons, but by the third day something profound should come out, and that distilled essence is what should show up in the paper. The "ball by ball description" of the minutiae of detail can be avoided in a paper like this. It puts the whole exercise in a bad light.”

The comments of the reviewer refer to the presentation of the material in the paper. We appreciate that it is the authors’ responsibility to present the material in a way that is helpful to

the reader. The reviewer suggests there is too much detail in the paper and yet he/she does not know enough about the model and the catchment. While this seems like a contradiction to us we will restructure the paper. We will move some material into an appendix and focus on the main message rather than on the details. Of course, runoff modelling is a complex issue and, often, detail is important to understand the bigger picture, so some detail needs to be retained.

"4) As a way to guide the authors towards this, I have a suggestion. I assume that the authors have now come to the final stage of their experiment. I think they should use the hydrologic data they have collected so far, i.e., rainfall, runoff, soil moisture (and any other internal or surrogate information) to piece together (without the benefit of any model) what actually happens or happened on this catchment. Surely, they now have a better clue on this, at least as best as they can. They can extract all kinds of metrics and signatures on the observed variability. This summary will satisfy me on learning about the catchment and its hydrology. This was my first point earlier (getting to know the catchment)."

This paper is the sequel to a previous paper where we exactly did what the reviewer suggests. We indeed pieced together the main runoff generation processes in the light of the data and surrogate information. In the revision we will better highlight how this paper relates to the previous one to address this concern of the reviewer.

"5) Then they should evaluate performance of each model juxtaposed against the actual observed response. This will expose the fundamental strengths and weaknesses (including their inappropriateness for this application). This will bring the focus back to the model structures and process descriptions in each model, and away from the pure litany of detail that is presently there. It will provide structure to the paper, and will become an object lesson to future modelers."

We did this by calculating the root mean square error and the Nash-Sutcliffe efficiency for each model and presented the improvements in terms of modelling stages. However, we are not sure whether evaluating the performance of each model separately and in more detail against the observed response would increase readability. In fact, a long list of model structures and processes may increase redundancy rather than clarity. As mentioned above, we will restructure the paper to better bring out the main points and avoid listing details.

"6) In fact, if at all possible, the authors should develop minimalist model(s) on the basis of the rainfall-runoff data they have collected, and in this way throw light on the essential or dominant processes happening, and this is another vehicle they can use to throw light on the strengths and weaknesses of the models."

We are not sure what makes the reviewer recommend develop minimalist models and fit them to the runoff data. In what way would this help in predicting runoff for the case where no runoff data are available (which is the topic of the paper)? Neither can we see why minimalist models would help in testing the strengths and weaknesses of the models used in the paper given that their structures and parameters are different. We realise that a spectrum of different models

(including minimalist models) could be developed from the data of the catchment but this would be the subject of a completely different study.

Furthermore, our idea/strategy was totally different; instead of developing an optimal model based on all information available we aim at analysing the predictive potential of existing state of the art models in such data limited case. And we aim at learning about the impact of modeller's decision on the improvement of the performance. Doing this systematically we can learn a lot on a) the limitations of the available models and b) on reasonable parameterisation strategies.

"7) The object lesson for me (something that the authors can expand on) is the value of even minimal amount of rainfall-runoff data before one does any kind of modeling, in any kind of circumstance. Even a month (let alone year) worth of data is far in excess of what you can learn from soils data or field visits. This may be disallowed in PUB, but this modeling study has demonstrated the folly of believing all the nice process theories in the classroom for predicting the real world, even in the case of a human-made catchment (even more so I would say)."

Whether a month worth of data is far in excess of what one can learn from soils data or field visits we are not sure, as we would expect this to depend on the runoff regime and the runoff characteristics of interest, but we agree that knowing even a minimal amount of rainfall-runoff data can be useful for choosing model parameters. We will more strongly emphasise the importance of runoff data in the paper but note that this has been said before, such as in the following statement (Blöschl, 2005, p. 15): 'It may seem strange to end a review of rainfall-runoff modelling in ungauged catchments with a note on the value of runoff data, but that in my opinion is the state of the science. As highlighted throughout this article, there are methods for dealing with rainfall/runoff modelling in ungauged catchments, but significant uncertainties remain. Short-runoff data series may contain very valuable information […] so the single best recommendation on the issue of rainfall-runoff modelling of ungauged catchments may be to install a stream gauge!'