Hydrol. Earth Syst. Sci. Discuss., 8, C1983–C1987, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C1983/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Top-down analysis of collated streamflow data from heterogeneous catchments leads to underestimation of land cover influence" *by* A. I. J. M. van Dijk et al.

N. McIntyre (Referee)

n.mcintyre@imperial.ac.uk

Received and published: 6 June 2011

This is an interesting topic and the paper includes some interesting ideas for examining it. I like the idea of using synthetic data, produced by a complex model, and adding noise to the data, to examine the implications of using more simple models (this idea has been used before several times in the hydrological literature, but little in this context). This aspect of the paper could be improved by running the AWRA model with multiple possible parameter sets and ideally another model structure too, so that results are not conditional on the assumed process-based model and parameter values.

I don't, however, like the idea that the presence of land cover signals can be deduced from the AWRA model just because it gives results which are somewhat consistent with small-scale experimental data and somewhat (arguably not) consistent with large-scale observations. This is too open to mis-conclusions due to the effect of structural error and data uncertainty. Generally, my view is that the focus of the paper should change, and the authors need to be more critical of results.

Detailed comments: 4123, Line 22. "(out of a total number of 221 and 1508 reported in the various studies)". Not clear what these two numbers refer to, or what studies are referred to.

4123, Line 24. "wi a dimensionless model parameter that characterises the hydrological behaviour of land cover class i ." Please be more specific.

4125, 17-20. It's not clear why this would influence the estimation of w.

4127. We need more information about the data in order to interpret the results. For example: What was the basis for the 'quality codes'? What was the precipitation station density, and how was the spatial interpretation done? How was the precipitation data quality judged? What is the temporal resolution?

4128. Excuse my ignorance, but I'm not familiar with this Budkyo model. It seems inappropriate for catchments where PE«P. Intuitively for such catchments, Q is almost equal to P whereas this model may predict the opposite, depending on the value of w. I know this reflects my ignorance, and the authors have referred to another paper, but it only needs two or three sentences here to explain why model that might appear unintuitive should be considered appropriate.What time resolution were the models run on?

4130 "First, we fitted the two parameter Zhang model (Eq. 3) by minimising the standard error of estimate (SEE) against Qobs from the 278 catchments" At what time resolution? Why is the SEE considered suitable? What minimisation algorithm was

used?

"We considered performance to be acceptable if the predictions were as good as that of the calibrated two-parameter Zhang model or better". It's difficult to see the relevance of this. It's reasonable to say that the first test of 'fitness for purpose' should be that the model should be consistent with observed responses, within the tolerances of the data uncertainty. But to test the model with respect to the tolerances of error (probably large, biased and hence the data are not representing reality) introduced by the Zhang model is not reasonable at all. Also, the fit measures used are not informative about the nature of the errors and hence the reader has no means to judge the significance of the model.

The second tests are not much different from the first: the benchmark consists of results from other models. The only conclusion that can be drawn is that there is no evidence that the summary outputs from the AWRA-L model are inconsistent with previous simple models. But the reason for using the AWRA-L model is that it should add information beyond these simple models. The accuracy and relevance of that additional information remains completely untested.

It seems the only way to rigorously do this, with available data, is to constrain the uncertainty in the prior AWRA model against observations from the 278 catchments and see if the model (with its posterior uncertainty) can distinguish, with confidence, the responses from different arrays of land uses. Even this would be open to the bias introduced into the upscaling by the model structural error (see below).

4130, Line 21. By using the calibrated parameters?

4131, line 5. To me, this is flawed logic. It's not clear in the paper where the empirical evidence comes from - presumably from relatively small-scale studies given the underlying arguments of the authors. If the AWRA model results are consistent with a relatively small-scale evidence base, but there is conflicting evidence from empirical analysis of large-scale data, this points to the result that the AWRA results don't reflect

the reality of upscaling – which is not surprising given AWRA's lumped treatment of routing processes.

4131, line 11. If the errors introduced are independent of the land use then it's difficult to see how this would make a major difference to results, when using 3000 samples. In reality, errors may be much larger than std=0.1, not independent from land use, and there are only 278 samples. The choice of 0.1, uniform over all variables and catchments is very artificial, and it's unclear from the paper how this is applied (e.g. randomly to each time-step and catchment, or maintaining some autocorrelation?). I have some sympathy with the authors here because inevitably such an experiment will be simplistic, but I feel a better effort could be made.

4132. This seems a better test: although still conditional on the AWRA model, at least it investigates what information can be lost or twisted by using a simplified model. It would have been better to run multiple realisations of AWRA so that results are not conditional on one set of parameters/input variables.

4133. Again, this seems a reasonable idea: for information on land use effects to be useful, the confounding effect of climate needs to be studied. However PE is a function of both climate and of land use, so the results here may be difficult to interpret. It would be better to sample fundamental climate variables which are not directly influenced by land use, and calculate PE using Penman-Monteith. Showing all the measures in Table 1 is not useful, as they are so well correlated: they (almost) all give the same information about performance. Can the authors use more informative range of performance measures?

4134, 13. I don't think "confirm" is the right word. The analysis here is very limited in the sense that the performance measures used may not reflect changes in the signals in the responses; and the overall uncertainty obscures changes. "support" would be okay.

4134, 23. How can the authors conclude that the AWRA model can accurately predict

the flow on the basis of these results? This does not provide much confidence in their critical interpretation of results. “reproduce” is too strong given results in Fig 2: the authors should be more critical about their methods and results.

4135. As the synthetic experiments (fitting simple models to the idealised flows) is, in my opinion, the most interesting and useful part of this analysis, it’s disappointing that the experiments and results were not more extensive and detailed.

4125, 20. This looks like a significant range of r2 values. At what confidence level is r2=0.1 , 0.2 significant? (after all, the main reason for doing a log-normal transform is that assumptions behind standard parametric significance tests are more valid).

4137, 11 “Our results demonstrate that a dynamic hydrological process model can reconcile this paradox” But this seems to be flawed: the process model is: 1) built on prior perceptions, which includes the perception that land use should affect flows, and hence it is inevitable that the process model will illustrate differences in responses: 2) a probable physical reason for the loss of signal at larger scales is the integration and smoothing of signals due to routing processes, which the ‘process-based’ model does not attempt to represent with any degree of realism.

4138 “Arguably, it is sufficient to demonstrate that the observations can be reproduced by a (more complex) theory and therefore can be reconciled with experimental knowledge.” The problem is that many alternative complex theories will reproduce the observations and give different interpretations of what is causing difference in response. This well-known aspect of the problem has not been addressed in the paper

The conclusions are fair. The smoothing of signals due to routing should be included in (4).

The English is generally good, but there are some grammatical errors.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 4121, 2011.