Hydrol. Earth Syst. Sci. Discuss., 8, C3447–C3456, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C3447/2011/


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.

**A. I. J. M. van Dijk et al.**

albert.vandijk@csiro.au

Received and published: 14 August 2011

INTRODUCTION: We thank the reviewer for the considerable effort he has put in reviewing our manuscript and his comments, which have helped us improve the manuscript. Below we provide a detailed response to his comments, with reference to the pdf we have attached in our response to the editor, and which contains line numbering and changed text in red font.

COMMENT: "[...] 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



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

not conditional on the assumed process-based model and parameter values."

RESPONSE: While we appreciate the sentiment, we would argue any results obtained with modelling are circumstantial and that real proof can only come from observations – hence many models (or variants) would provide no more 'proof' than one. We have tried to make it clearer in the revised m/s what we do not aim to test or pretend to prove, by adding a caveats section (see response to editor).

COMMENT: "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 largescale observations."

RESPONSE: We most certainly do not claim we can deduce this from a model. Hopefully the caveats added address this in a satisfactory way too.

COMMENT: "(out of a total number of 221 and 1508 reported in the various studies)".

RESPONSE: We have clarified this: "The cited studies performed such an analysis using collated data for 221 (Donohue et al., 2010) to 1508 (Oudin et al., 2008) catchments." (line 100-101)

COMMENT: " "wi a dimensionless model parameter that characterises the hydrological behaviour of land cover class i ." Please be more specific."

RESPONSE: We have added to this: "and may be interpreted as a measure of the efficiency with which vegetation accesses and uses stored water." (line 96-97)

COMMENT: "It's not clear why this would influence the estimation of w."

RESPONSE: We added: "Depending on their covariance with land cover, these attributes may attenuate or enhance any land cover signal."

COMMENT: "We need more information about the data in order to interpret the results. For example: What was the basis for the 'quality codes'?"



HESSD

8, C3447–C3456, 2011



RESPONSE: The quality code flags whether the streamflow was interpolated or measured, and whether the rating curve was considered to be reliable for that flow rate. We did not use interpolated data and we did not use stations with poor rating curves. As this is arguably only good practice, we removed this qualification from the main text to avoid confusion (line 238).

COMMENT: "What was the precipitation station density, and how was the spatial interpretation done? How was the precipitation data quality judged?"

RESPONSE: The interpolated climate products are widely used in Australia. Rather than referring the reader to the four publications mentioned (Guerschman et al., 2008; Guerschman et al., 2009; Van Dijk, 2009; Van Dijk, 2010a) we have referred to the appropriate reference (Jeffrey et al., 2001) directly (line 242).

COMMENT: "What is the temporal resolution?"

RESPONSE: Daily; we now added this (line 240).

COMMENT: "[...] Budkyo model [...] seems inappropriate for catchments where PEÂńP. What time resolution were the models run on?"

RESPONSE: If PE«P than the Budyko model will predict that Q∼P-PET. Budyko models use long-term average variables only (line 93) and therefore do not have a time resolution as such. It may be mainly for this reason that it has been referred to as a 'top-down' (sensu Klemes, 1983) method.

COMMENT: "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?"

RESPONSE: The model uses long-term averages.

COMMENT: "Why is the SEE considered suitable?"

RESPONSE: Whether SEE is a suitable criterion may well be a valid question, however



HESSD

8, C3447–C3456, 2011



we are merely replicating the method as it has been published before, which used a least squares criterion.

COMMENT: “What minimisation algorithm was used?”

RESPONSE: We now added in the text that we used using Solver in Microsoft ExcelTM. We did test an alternative optimisation algorithm in MATLAB and also tried different starting values, but basically the error surface appeared pretty well behaved and always converged to the same parameter values. Perhaps not surprising given there are only two parameters.

COMMENT: “It’s difficult to see the relevance of this [...] 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.”

RESPONSE: Indeed this was exactly the criterion we hoped to satisfy. We note that the problem is not with the simple models per se, but with the fact that the produce a different land cover influence, expressed in different parameter values, when fitted to the data.

COMMENT: “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.”

RESPONSE: We have to disagree here, as this was not our reason. We have added some text to make this clear: “We emphasise that our objective does not require that the process model explains more variation than the Zhang models in one or both cases; equal or similar performance would be sufficient. The critical difference is that fitting the Zhang models is expected to lead to two substantially different parameter sets, producing two mutually inconsistent models in the respective applications. By contrast, the process model uses one parameter set only for both cases and therefore produces internally consistent results. That the process model parameters were estimated a



HESSD

8, C3447–C3456, 2011



priori rather than optimised is not essential but arguably preferable." (line 336-343)

COMMENT: "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)."

RESPONSE: Since 'testing the accuracy and relevance of the information provided by AWRA beyond the Budyko models' was not our goal, we do not feel this analysis would be appropriate to include. Indeed, there would seem no reason to assume the process model would not be prone to similar methodological issues as the top-down analysis tested. Essentially, our conclusion is that the land cover signal in the 278 catchment data set is swamped by other causes of variability, and therefore one simply cannot reliably determine a land cover signal from it. The choice of inference model is secondary.

COMMENT: "By using the calibrated parameters?"

RESPONSE: No, the AWRA-L model was not calibrated, we used prior parameters.

COMMENT: "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."

RESPONSE: We do not fully understand this logic. We show that the process model (with one set of parameters) can reproduce both the land cover influence for extreme cases, as well as the lack of signal when analysing the mixed catchments with the Zhang curve. Crucially, a land cover influence is still inherent in the process model simulations in the latter case, but it can no longer be extracted using the top-down method. This then means that there is a methodological problem with the top-down



HESSD

8, C3447–C3456, 2011

method, and that was our hypothesis. There would not seem to be any need to look for a problem in the AWRA model (even though no doubt there are many, as with any model..) because it is sufficiently consistent with the observations.

COMMENT: “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.”

RESPONSE: We have phrased this more carefully. In each experiment, the sample is 278, just like the original data. We performed the experiment 3000 times (line 363-364).

COMMENT: “In reality, errors may be much larger than std=0.1, not independent from land use, and there are only 278 samples.”

RESPONSE: We agree; though the 278 samples are of course chosen because we need to remain consistent with the original data set.

COMMENT: “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?).”

RESPONSE: Noise was added to the long-term average values and therefore temporal structure and autocorrelation do not apply. It was added to catchments individually, otherwise it would have been an experiment probably looking more at bias rather than noise. (line 357-358)

COMMENT: “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.”

RESPONSE: We could easily make the experiment more complex, but in that case another reviewer might well argue that our choices were arbitrary and that the added complexity muddles the interpretation of the experiment.

COMMENT: “It would have been better to run multiple realisations of AWRA so that results are not conditional on one set of parameters/input variables.”



HESSD

8, C3447–C3456, 2011



RESPONSE: We appreciate the sentiment, and indeed considered doing this. However this would have added another dimension to the experiment and would have made it harder to interpret. (The various reviews suggest that our experiment is already enough to put in the right context as it is...) Moreover, as stated before (and also in the added Caveats section), even if we had used hundreds of models (whether structures or parameter values) we still would not be able to prove that this effect exist in reality. Rather, we show that it is plausible, for which one model realisation suffices. We did adjust words in places to avoid unrealistic expectations (e.g. line 489)

COMMENT: "It would be better to sample fundamental climate variables which are not directly influenced by land use, and calculate PE using Penman-Monteith."

RESPONSE: PE used in the analysis method was derived from the national gridded product and is not calculated with consideration of land cover. We appreciate that, at least in theory, the underlying station meteorological measurements may have been affected by land cover but that effect is impossible to isolate and does not appear crucial for the interpretation of this experiment. (it is noted that the process model did not use PE grids but the underlying national meteorological grids, which again in theory may equally have an implicit land cover signal)

COMMENT: "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?"

RESPONSE: We would be very happy to list more performance measures in addition to the five listed in Table 1, if the reviewer could suggest some. (btw we do not think they are necessarily well correlated; that would depend on sample size and any heteroschedasticity in the error distribution)

COMMENT: "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



HESSD

8, C3447–C3456, 2011



okay."

RESPONSE: Indeed this is the meaning we intended. We changed it to 'support'.

COMMENT: "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."

RESPONSE: 'Accurately' and 'reproduce' are obviously always going to be comparative and subjective terms in this context, but we agree that this potentially could be misinterpreted to mean that we misguidedly have started to believe our process model equals reality. We changed it to "It follows that the process model (1) can satisfactorily predict streamflow from the 278 catchments with mixed land cover, and (2) produces a land cover signal of similar magnitude as captured by the Zhang et al. (2001) models (line 442-446).

COMMENT: "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."

RESPONSE: The synthetic experiments are the core of the analysis and we have tried to describe them in as much detail as needed to interpret them. We did make a conscious decision to keep the experiments simple however, as additional complexity inevitably introduced additional assumptions. We would be keen to provide as much detail as would be helpful to the reader but would need some concrete suggestions.

COMMENT: "This looks like a significant range of r2 values. At what confidence level is r2=0.1 , 0.2 significant?"

RESPONSE: We apologise for this mistake; the numbers we gave were r not r2 values. We changed this to r2<0.04, which we trust the reviewer will agree is insignificant by any reasonable measure.



HESSD

8, C3447–C3456, 2011



COMMENT: ""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."

RESPONSE: When we state that we can reconcile the paradox, we mean that we provide a plausible explanation. We do not pretend to prove that this is the only or main mechanism. We now state this hopefully more clearly in the Caveats. We mulled over the influence of routing processes but have not been able to come up with a mechanism by which this would reduce the influence of land cover on mean streamflow. We would welcome the reviewers thoughts on this.

COMMENT: "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"

RESPONSE: We do not see this as a problem as we did not claim exclusivity and our goal was not to validate the process model; rather investigate a potential problem with the top-down analysis method. As we now made explicit, we could have used one or many other models, as long as they could reproduce the two sides of the paradox we investigate.

COMMENT: "The smoothing of signals due to routing should be included in (4)."

RESPONSE: See above, we would be happy to include this if we understood the physical mechanism.



HESSD

8, C3447–C3456, 2011

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



HESSD

8, C3447–C3456, 2011

Interactive Comment