Hydrol. Earth Syst. Sci. Discuss., 6, C2800–C2815, 2009
www.hydrol-earth-syst-sci-discuss.net/6/C2800/2009/


Hydrology and Earth System Sciences Discussions



# *Interactive comment on* "Selection of an appropriately simple storm runoff model" *by* A. I. J. M. van Dijk

**A. van Dijk**

albert.vandijk@csiro.au

Received and published: 3 December 2009

Author response to comments by referees #1,2 and 3

First of all, I would like to thank all three referees for their comments, which have provided valuable suggestions to improve this manuscript. Below I offer a response to the main comments (where I did not respond to specific more minor technical or textual comments I accepted the suggestions in their entirety will simply revise accordingly).

AUTHOR RESPONSE TO ANONYMOUS REFEREE #1

COMMENT) The major weakness relates to the set of candidate model structures; as indicated by the author in the conclusions (and in section 4.7), the most complex model



Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

considered in this paper is not complex enough to accurately describe the data. This is clearly evident in Figure 4, which reveals relative average errors of 50% or more in all but one of the basins. Inclusion of the true model, or a sufficiently accurate model, in the set of candidate models is a requirement for many model selection criteria (including the final prediction error criterion FPEC used here), and in general is a prerequisite for good model inference. See for example the book by Burnham and Anderson (Model Selection and Multi-Model Inference) for general discussion and background on this. Therefore, the author should consider including more complex models in the analysis; complex enough so that accurate predictions are obtained and that the model is over-parameterized (with parameter correlations above 0.9, instead of a maximum of 0.4 reported here). That model would then provide a good starting point for introducing model simplifications to reduce parameter equifinality without significantly worsening predictions. Responding to this criticism will likely require major revisions.

RESPONSE) While I understand this comment, it is not valid. There is evidence to conclude that the reason for the moderate model performance is not mainly due to the model structure, but due to (1) errors in the forcing; and (2) lack of observations on a key driver, i.e. rainfall intensity. It is worthwhile noting that Australia's environment and the observational data set available to us has some specific characteristics that affect the predictability of storm runoff response, as outlined below.

Firstly, the rainfall observation network that is generally very sparse, particularly in comparison to networks in Europe. To demonstrate this, I have added two figures below. The first from the GPCC; the other based on the complete Australian rain gauge network available to the Bureau of Meteorology and at the basis of the interpolated rainfall product used in this study.

<SUPPLEMENTARY FIGURES>

<CAPTION> Indication of density of rainfall gauging network (left) world-wide, expressed as number per 1 degree grid cell (note that national networks may be denser



HESSD

6, C2800–C2815, 2009

Interactive Comment

in some countries); (right) the density of the gauging network used to produce daily interpolated rainfall surfaces for Australia, as number per 0.25 deg grid cell.

Because isolated convective events contribute a good part of the precipitation in many parts of Australia (and most other semi-arid regions), the spatial variability in rainfall is considerable and therefore rainfall events often not captured well. The network is denser in parts of the Great Dividing Range, where rainfall is also higher, but here there spatial variability is enhanced due to orographic effects. A third perhaps relevant issue is that a time-varying part of the gauge data set is run as a volunteer network, which may mean measurements do not have the same accuracy and representativeness as automated and agency-maintained networks. In summary, estimates of daily, catchment average rainfall derived from the interpolated data are likely to have considerable error in them. Yet it is the best data we have available and does not limit the analysis. There is not an obvious a priori reason why the conclusions should be affected in any other way than degrading the degree to which the best model can explain the observations. I included comments to this effect on page 5771 (line 8-21) but can expand by introducing some of the above comments into the manuscript.

Secondly, many (but not all) of the Australian catchments investigated here either contain considerable relief, experience a relatively dry climate, or both. This means that saturated areas do not develop as extensively as in lower relief, humid catchments and hence that infiltration excess (Horton) overland flow plays a more important role than saturation related runoff processes. This in turn means that intra-storm rainfall intensity (both spatially and temporally) is likely to be an important factor determining runoff response, as argued in the manuscript. As stated on page 5771 (line 17-18), the catchments with the poorest model performance were found to be dry catchments. Unfortunately, since no accurate observations with good spatial coverage are available in Australia, there is currently no way in which this process can be included in a more complex model.

In summary, I do not see a reason to expect that a more complex runoff model would produce a better result. I note that the model formulations tested here cover the formulations used in most common runoff models (e.g. HBV, Simhyd, SCS-CN) as well as reproducing or approximating behaviour of formulations used in several land surface models (page 5756 line 16 to 5757 line 10). Therefore, I am not sure what insight could be expected of trialling other, more complex models, or indeed, what such a model should look like. I note that a six parameter model was trialled (very complex when compared to commonly used formulations) which did not produce a better result but was indeed associated with much parameter equivalence as predicted by the referee (page 5763, line 17-20). Approaches not trialled here are highly distributed models or a multi-layer Richard's equation based model. Given the low gauge density and limited availability and quality of spatial terrain and soil data I see no reason to expect that a highly distributed model would perform better. The scaling (e.g. due to preferential flow paths and spatial variability) and consequent parameterisation issues associated with applying multi-layer models over large areas are well documented, and to derive good benefit from describing vertical water movement sub-daily rainfall intensity data would be required, which is not available In either case, this study focused on lumped catchment models, which are a mainstay in catchment rainfall-runoff modelling precisely for their parameter parsimony (also meaning they are easily automatically calibrated) and modest input requirements (which fits limited data availability in many cases). If it is deemed useful I would be happy to express this rationale better in the m/s.

As an aside, I note that the error statistic used here is should not be compared with the commonly used Nash-Sutcliffe Model Efficiencies (NSME) calculated for daily total streamflow. Such NSME values would be expected to be higher due to the contribution of baseflow (which can be estimated more accurately) and the squaring of errors in the formulation of NSME puts more emphasis on the ability to reproduce the largest few events. If the referee believes it is helpful I would be happy to illustrate this in the discussion with some numerical examples.

COMMENT) A scatter plot of predicted vs observed event runoff for the best model



HESSD

6, C2800–C2815, 2009

would be useful to evaluate how well this model performs.

RESPONSE) I appreciate that this may illustrate some of the points above and would be happy to include that.

COMMENT) Eq. (15): what are the advantages of using absolute errors instead of squared errors? Eqs. (14) and (15) are unconventional implementations of FPEC; usually epsilon is equal to the mean square error; please justify use of Eqs.(14)-(15), especially with regard to the statistical assumptions underlying FPEC.

RESPONSE) This was done for reasons outlined in the references on page 5762, line 13 (I could cite some salient points in the m/s if considered useful). An additional reason not stated by these authors (and also not the main reason in this analysis, but relevant nonetheless) would be that the streamflow rating curve is often poor for the highest flows, and therefore using a squared error statistic will put most emphasis on what are effectively the poorest observations. Happy to also include and discuss squared errors if it was deemed useful or insightful, but in writing the manuscript I felt it would not improve readability.

COMMENT) When applying FPEC it is crucial to adequately calibrate each candidate model, i.e. for each model one needs to find the parameter set that minimizes epsilon. On p5762, line 16, it is stated that calibration is done with a monte carlo approach using latin hypercube sampling. The author needs to provide convincing evidence that his method leads to identification of optimal parameter values: how was optimality evaluated? How many random samples were generated? What was the stopping criterion?

RESPONSE) LHS was chosen to ensure the full parameter space was explored, and a local downhill search was used to ensure the minimum was found (this got truncated in the final m/s, apologies). The model parameters were optimised for each station by random sampling the feasible parameter space N times (with N=10ˆd and d the number of fitted parameters), followed by a local downhill search (using the function



HESSD

6, C2800–C2815, 2009

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

fminsearch in Maltab with default parameters). I fully appreciate there can be debate about the optimal optimisation method, but it seems very unlikely the conclusions would be different if another method was used.

COMMENT) Section 1.3, lines 6-8: for sure there are numerous studies that have compared alternative model formulations for predicting runoff, more literature is needed

RESPONSE) I should have been more explicit here. Indeed, I am aware of various studies that compare alternative catchment rainfall-runoff models against total catchment streamflow (including both storm- and baseflow). However I have not found any published assessment of a variety of model formulations to predict catchment-scale event storm flow generation as a function of daily rainfall totals. I would be very grateful for any pointers in this regard.

AUTHOR RESPONSE TO ANONYMOUS REFEREE #2

COMMENT): The author takes the USDA Soil Conservation Service curve number method and explores several simplification options by using an extensive dataset from 260 catchments. While the database is impressive and should yield some useful insight, as the author states himself, even the optimal, most complex model structure studied shows only moderate performance, and implications or recommendations based on this study remain unclear. The study demonstrates the difficulties associated with relating empirically obtained relationships and parameters to processes and physically meaningful parameters. I would recommend to include different/and or more complex model structures that describe the data more satisfactorily and from there stepwise simplify and discuss the tradeoffs between model simplicity and model performance.

RESPONSE) I should note that the SCS-CN method was only one of a wide range of formulations trialled; I refer to my response to referee #1 for details.

The referee suggests that the moderate performance of the best model is less than can feasibly be obtained. This may be based on the referee's experience or assessment of previous publications. However, in my response to referee #1 is outlined why this is probably an unreasonable expectation to apply to the results of this study (that is, the sparse Australian rainfall gauging network; the importance of (unmeasured) rainfall intensity; the difference between daily streamflow and event storm flow; and the difference between the error statistic used here and the Nash-Sutcliffe criterion commonly used in other catchment model assessments). However as also explained in the response to referee #1, there is no obvious reason why these differences should affect the main conclusions of this study.

It is unlikely that more complex formulations would have produced better results, for reasons also explained in my response to referee #1.

COMMENT) Section 1.2: There have also been studies documenting and discussing threshold behaviour at different scales. Adding some of them to the discussion of threshold behaviour would be helpful.

RESPONSE) Would like to, but I would be grateful for specific suggestions in this regard...

COMMENT) The dataset of 260 catchments seems to encompass quite a range in climatic conditions. More information on climates, topography, geology and anticipated dominant runoff mechanisms would be very helpful. It should also be stressed more that all catchments were apparently taken from very humid regions – certainly a constraint for applying results to other regions.

RESPONSE) Good point, will include a more elaborate description, particularly regards runoff mechanisms. While compared to average Australian conditions the catchments data set was biased towards more humid catchments, it did include some very dry catchments, and compared to the rest of the world I do not think the catchments were particularly humid. Even so, happy add some more words.



HESSD

6, C2800–C2815, 2009

Interactive Comment

COMMENT): Section 3.2: this subset of 20 stations – how was this subset selected? RESPONSE) Manually, to cover the climate, geographic and land cover range in the data set; will add this.

COMMENT) Also, I think some of the results could be presented better in plots – e.g. the distributions of parameters. P. 5765, lines 12-13: considering the annual rainfalls, none of the catchments seem to be "dry". Please explain how you define dry and wet. Again, it would be useful to have more information on climatic conditions, rainfall distribution etc.. Also, isn't that counterintuitive that saturated area changes faster in a dry catchment than in a wet catchment?

RESPONSE) Will do; wetter and drier are meant in a comparative sense here, will revise. Regards the last comment: my intuition would suggest otherwise, but in any case I note that greater rainfall is correlated with greater relief in this data set, making it more difficult to draw any solid conclusions on this (although relief was tested as an explanatory variable in its own right).

COMMENT) This study is averaging results over a wide range of catchment sizes and climates and, I assume, topographies/geologies. Wouldn't it be helpful to analyze results for subgroups – e.g. small vs. large catchments, steep vs. shallow, depending on rainfall distribution over the year?

RESPONSE) Indeed this was the intent. However, because the potential ways of grouping are numerous, the approach taken here was to do this by correlating statistics and parameter values to catchment attributes, as done in various places.

COMMENT) Section 4.4, last sentence: Generally specifying initial losses as 12 mm is a big simplification, across scales, climates, topographies and geologies. Please provide a rationale why this "would seem realistic".

RESPONSE) Of course I agree that values will vary in reality and that any model is by definition is a simplification. However the statistical analysis presented here allowed the conclusion that it is not a significant over-simplification in light of the available observations. It has the important advantage of avoiding parameter equivalence. (For example the parameter equivalence between initial loss and maximum storage has been shown a big issue for the SCS-CN technique, hence the similar parameter reduction commonly applied in using that model (page 5767)). The interpretation that 12 mm seems realistic is based on the arguments provided in Section 4.4, but I could revise this to say “does not seem unrealistic”.

COMMENT) Section 4.7, lines 12/13: Doesn’t that strongly depend on climate (e.g. convective events with high intensities vs. frontal events with typically moderate/lower intensities although storm totals can be very similar)?

RESPONSE) This sounds feasible. I struggled to find studies addressing this, but will incorporate the comment somehow.

COMMENT) Many conclusions remain speculative, e.g. conclusion 4, without actual field data/observations backing those assumed mechanisms.

RESPONSE) While I agree that this study does not provide definite proof for conclusion 4, I do not agree it is speculative; it is an inference made by combining several lines of evidence. I will try to rephrase this.

COMMENT) I am not clear on what the implications of this study are and how results can be used further. Here is where I think that the inclusion of a more complex model could improve he usefulness of the study.

RESPONSE) By way of example, a very direct and practical use of this study has been for to help CSIRO decide on the storm runoff model formulation to include a large, more complex landscape hydrological model we are developing within one of our large projects. This study was intended to allow that decision to be made based on benchmarking against observations, rather than only on theoretical arguments (which is invalidated by known scaling issues) or conceptual preference (which is not scien-



HESSD

6, C2800–C2815, 2009

tific). However I would welcome suggestions as to how I might make this point more eloquently.

COMMENT) I think a figure plotting measured vs. modeled storm runoff at least for one of the models (the optimal model structure) would be very informative.

RESPONSE)Good point, will add this in.

AUTHOR RESPONSE TO ANONYMOUS REFEREE #3

COMMENT): In the conclusions' section the success of the study is minimized. There and in section 4.7 the weak performance of even the best model is mentioned. This means that the applied models are not able to reproduce the observations. The reasons for this failure might be that the model does not include (not enough or) not the adequate parameters or the model structure does not imply the relevant processes. Therefore, the author should include more complex model structures that are able to reproduce the observations better. As far as the model evaluation is not improved in this sense, there is no benefit for the community of hydrologists.

RESPONSE) Unfortunately I have to disagree with the first comment entirely. While it is true that the best model is only moderately successful in reproducing observations, there is no reason to believe that any other model would produce better results. Yet, in reality, choices about model structure still need to be made to support water management and therefore I consider this study still relevant

Perhaps the referee makes the interpretation by comparing against own experience or knowledge from reading previous studies in other regions. In my response to referee #1 and #2 I outline why this benchmark may not be applicable to interpret this study (the sparse Australian rainfall gauging network; the importance of (unmeasured) rainfall intensity; and the different nature of performance statistics commonly used catchment model assessment). However these differences do not need to affect the conclusions of this study, merely the amount of variation in observations that can be explained.



HESSD

6, C2800–C2815, 2009

Interactive Comment

Although it would of course have been great if more of the variation in the observations had been explained, there is still a practical need to estimate streamflow for many applications, and more often than not, the availability of rainfall and other observations is suboptimal. Literature suggests that better simulations might have been obtained in (often experimental) catchments where large amounts of accurate and detailed rainfall intensity measurements are available. However such findings do not translate directly to the reality of water management in many regions, where such dense observations are simply not available.

In my opinion the findings in this study have several benefits for the hydrologic community. Firstly, it establishes what an appropriately simple storm flow model formulation is for environmental and observational conditions similar to those for the catchments investigated. I refer to my response to referee #2 for a very concrete example of how the results of this study have already been used in Australia.

Secondly, I draw the opposite conclusion from this study than does the referee: the results show that adding complexity to models does not help to improve hydrological estimation. I provide further arguments in my response to referee #1 (a complex 6-parameter model was tested; distributed or multi-layer were not considered for good reasons). Hence more complex models do not appear to offer a benefit, and better spatial and temporal measurements of rainfall intensity really are a prerequisite for improved estimation. This allows practioners to consider what accuracy and therefore observation investment is required to support water management, and dispels the unfortunate notion that observations can be substituted by models.

Thirdly, the analysis of the model parameters provides specific insights in dominant runoff processes, predictors and uncertainties, such as the predictive value of groundwater storage (inferred from baseflow) in estimating runoff response.

COMMENT): In the objectives (p. 5758) it is suggested that there has not been a comprehensive analysis of alternative model formulation. However, such model compar-



HESSD

6, C2800–C2815, 2009

Interactive Comment

isons are a wide spread research activity in hydrology and they should be summarized. A literature review in this regard is missing. An overview of similar studies should be given in order to compare the approach and to bring the results in a larger context.

RESPONSE) Referee #1 also raised this and I probably should have been more explicit here. I refer to my response to referee #1 and would be grateful for pointers to relevant publications I have missed.

COMMENT) The dataset of 260 Australian catchments is impressive and gives a good basis for the presented work. However, beside the range of size and annual rainfall rate, no useful information is given about the structure of the catchments (vegetation, soils, geology, groundwater zones, lakes, altitude, riparian zones, etc.) nor about the instrumentation of the basins (precipitation gauge network). There should be a table summarizing such information. It would make sense to choose only catchments with good data quality for this study. Likewise, the author should consider to include not every storm event but to choose "good" events due to sharp criteria. Especially, rainfall data plays (besides discharge data) a crucial role in this assessment. Therefore, the quality of the estimated daily areal precipitation is very important. In the end (p. 5771) the information is given, that the network is generally rather low. Because, even in very small catchments, the spatial variability of rainfall in space is high, it must be suggested that it has a severe influence on the results obtained in this study.

The crucial role of rainfall data quality is mentioned above. Good data could be used to show the variability of the results by using better (higher resolved) data (e.g. higher resolved data of spatial rainfall distribution in catchments where such information is available).

RESPONSE) Some very good points are raised here and I hope that the additional information I have provided in the responses to the three reviews have given some clarification. I would aim to include the key points of that in a revised m/s. Regards selecting for data quality; the referee raises a very good point that obviously deserves



HESSD

6, C2800–C2815, 2009

Interactive Comment

more attention. Given that the ultimate aim is to derive a storm runoff model that can be applied in all catchments (regardless of gauging density) I do not consider it beneficial to bias the data set by choosing only the most densely gauged catchments. It is also not obvious that the quality of rainfall forcing affects the conclusion per se.

However I do acknowledge that the importance of data quality and its influence on conclusions could have been discussed more elaborately by considering the relationship between gauge density on one hand, and the model performance and optimal model complexity on the other. I previously did do a cursory analysis of this and indeed catchments with greater gauge density are generally associated with better model performance, but without an obvious influence on optimal complexity. I would aim to elaborate on this in the m/s if it is deemed useful.

COMMENT) At least in the conclusions a proposal of ideas or approaches how the results could be improved should be presented (e.g. at p. 5769/70 good intentions to improve the assessment are mentioned).

RESPONSE) Agreed, I would be raising or emphasising the relevant points made in the responses to the three reviews.

COMMENT) In the paper some of the achieved results tried to be explained with the runoff generation theory and the involved runoff processes. These explanations are accompanied from speculations, suggestions and assumptions that are not comprehensible with the information given in this paper. On p. 5577 for example, for 72% of the catchments the calculated saturated area fsat was less than 1%. No information is given whether this might be plausible.

RESPONSE) I could include some literature references here, though they would be circumstantial rather than relate to specific catchments in this data set.

COMMENT) On p. 5769 daily rainfall intensities are brought together with Hortonian Overland flow (line 14-16). This argument is not suitable to explain higher peak flow rate.

RESPONSE) The argument was not intended to address peak flow rate, but total event runoff. Greater storm events tend to be associated with greater rainfall intensities, and that infiltration excess runoff increases in response to greater intensities. This is well published so I will add some references.

COMMENT) In line 12 perched water tables are mentioned. No information is given on the preconditions of this process.

RESPONSE) This was mentioned in the context of those two references; more information is contained in them. I would be happy to add some more detail here if deemed useful or remove these words.

COMMENT) Also on page 5766 (line 4-6) very vague and doubtful explanations are given for the behaviour of the hydrograph recession curves.

RESPONSE) the argument here is that the recession times are too long to be explained by surface flow routing.

COMMENT) The SCS-model is a rather simple and relatively weak hydrological model. Here it is used, although the adequateness of this model has been questioned in many parts of the world. The comparison should be done with a more adequate model.

RESPONSE) I am aware of the debate surrounding the SCS method and have no intention of defending the physical realism of the SCS model (indeed I identify some conceptual issues in the m/s). However, it is probably still the most widely used storm runoff model globally and contemporary scientific literature is still replete with studies on its application and development. Therefore in my opinion it would be wrong to ignore or dismiss it out of hand. Furthermore, while this study identifies conceptual issues with the SCS, it shows that the optimal model formulation among many alternatives does share some important similarities with the SCS method. In the main text I argue that the SCS method may well be 'right for the wrong reasons' (see Section 4.5).



HESSD

6, C2800–C2815, 2009

Interactive Comment

HESSD

6, C2800–C2815, 2009

Interactive Comment

COMMENT) To show some examples of good and worse results of modelled and measured hydrographs would be interesting.

RESPONSE): Good point. Since this is an event storm flow model I cannot show hydrographs, but will show some scatter plots.

COMMENT) Some catchments react strange to precipitation (e.g. annual runoff of 2 mm). Such data should not be used.

RESPONSE): These low values are accurate; there are some desert catchments in the original data set. This particular catchment and several other arid catchments were not be included in the selected data set however, because they did not have enough individual runoff events for a reliable statistical analysis. This comment reiterates that expectations of achievable model performance should not be based on experience in more humid environments. As stated on page 5771 (line 17-18), the catchments with the poorest model performance were found to be the driest catchments.

I hope my responses help to clarify aspects of this m/s and once again thank the referees for their thoughtful comments.

Albert van Dijk, 3 December 2009

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 6, 5753, 2009.

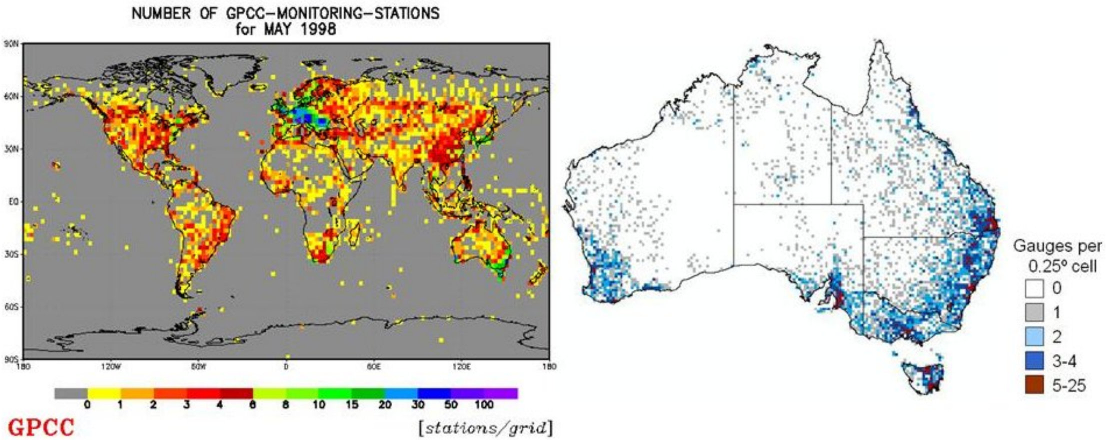


**Fig. 1.** (see main text for caption)

HESSD
6, C2800–C2815, 2009



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