Hydrol. Earth Syst. Sci. Discuss., 8, C244–C255, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C244/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Comparison of predictions of rainfall-runoff models for changes in rainfall in the Murray-Darling Basin" *by* J. M. Whyte et al.

J. M. Whyte et al.

jason.whyte@adelaide.edu.au

Received and published: 2 March 2011

We thank the reviewer for the prompt response, and we reply to the substantive objection before attending to specific criticisms. We note that for many of these we are able to provide a brief correction or clarification.

Response to Reviewer #1's substantive objection:
The main conclusion of our paper is that a comparison of model rainfall-runoff elasticity with the corresponding elasticity observed in the time series is a valuable adjunct to considereation of the Nash-Sutcliffe coefficients, typically calculated for model calibration and validation subseries. It is of particular relevance when investigating changes

in runoff due to projected changes in rainfall in climate change scenarios. We have not seen this published elsewhere.

The reviewer questions the relevance of a comparison of conceptual and time series models. To support our interest in this matter, we refer to quotes given in the introduction of our paper. On Page 919 we quote A. Sankarasubramanian *et al.* explaining how different rainfall-runoff models could give very different results on the same catchment. On Pages 919-920 of our paper we presented a quote from E. Todini which we paraphrase here as making the point that model comparison is not commonly seen in the literature despite the need for such comparisons. To further demonstrate the justification for comparing classes of models, we add Todini's comment, which preceded the quote presented, "From the wide range of models available, the choice of the one most appropriate for any specific task is difficult, particularly as each modeller tends to promote the merits of his/her approach." (E. Todini, "Hydrological catchment modelling: past, present and future", Hydrology & Earth System Sciences, 11(1), pp. 468-482, (2007).)

With such statements in mind, we sought to put a selection of models to the test, not merely by comparing their Nash–Sutcliffe coefficients, but also by considering their behaviour under changed rainfall conditions to obtain a broader view of the characteristics of the models on the four catchments considered. To the best of our knowledge, we are not aware of comparisons of these characteristics of time-series models and conceptual models in rainfall-runoff problems. We are aware of some extensive comparisons of conceptual models, (e.g. "Modelling Runoff and Climate Change Impact on Runoff in 178 Catchments in the Murray-Darling Basin Using Sacramento and SIMHYD Rainfall-Runoff Models", by F.H.S Chiew, J. Vaze, N.R. Viney, J.-M. Perraud, J. Teng and D.A. Post, in Proceedings of Water Down Under 2008, April 2008, Adelaide, Engineers Australia: pp. 87-97). We broaden the range of models compared by considering time series models in addition to conceptual models.

If the relative merits of models are not demonstrated and documented then surely individuals must repeat work to ascertain the relative abilities of competing models. Our comparisons of classes of models provides information which may be used to judge whether perceptions of the abilities of rainfall-runoff models are accurate under various rainfall conditions. When Reviewer #1 writes “The results for SIMHYD are much as expected..." in their first comment paragraph, we do not dispute the judgement, but wonder how it was reached.

The reviewer seems to favour conceptual models for rainfall-runoff modelling, of which we considered SIMHYD, over time series models. Time series models may be relatively simple but it seems that further complexity does not guarantee better performance. For example, consider “Regression analysis of rainfall-runoff data from an arid catchment in Oman" by N. Mcintyre, A. Al-Qurashi and H. Wheater, Hydrological Sciences Journal v52, no. 6, pp.1103-1118 Dec. 2007. The abstract states “There is a pressing need to improve capability to predict the hydrological responses of arid and semi-arid catchments. The literature indicates that physically-based rainfall-runoff models are not yet able to meet this challenge. Simple empirical or semi-empirical models may perform equally well or better, and provide basic but important information into catchment functioning."

Time series (or transfer function) models of rainfall-runoff processes appear in recent publications demonstrating their relevance to hydrology. For example, consider a very recent research monograph, “Stochastic and Statistical Methods in Hydrology and Environmental Engineering: Time Series Analysis in Hydrology and Environmental Engineering" (Water Science and Technology Library) edited by Keith W. Hipel and published by Springer on December 1st, 2010. Time series models are taught in courses in hydrology for professionals and postgraduate students and are covered in well-known handbooks for engineers, consultants and students, such as “Time series modelling of water resources and environmental systems", by K.W. Hipel and A.I. McLeod, Volume 45 of the Developments in Water Science series, published by Elsevier Science in 1994, and the text for students, “Rainfall-Runoff Modelling: The Primer" by K. J. Beven,

published by John Wiley & Sons, 2001. We propose that the continuing usage of time series models justifies comparison of their performance with conceptual models. We chose the conceptual model SIMHYD for our paper as it has been widely used and is available for use in the Rainfall Runoff Library, details of which were given in our paper. We used version 1.0.5.

Furthermore, the single regime time series models considered in our paper were used for stochastic simulations. If the deterministic conceptual models cited are to be used for stochastic, rather than deterministic simulations, then a time series model for the errors must be combined with the deterministic output.

We would like to emphasize that our paper does not merely compare a range of standard models. We believe that our paper is a novel application of threshold autoregressive models with exogenous input (TARX) to a hydrological system. We have chosen to consider TARX models as they are capable of exhibiting a richer set of behaviours than what we have called “single regime" time series models, that is, those governed by the same prediction equation and parameter values for all values of the catchment variables. We have presented a version of a TARX model in the framework of our comparison paper so that it may be evaluated against simpler time series models and SIMHYD. We found that the TARX models considered produce behaviour which is much closer to that obtained from empirically derived rainfall elasticity of streamflow (see reference to Whyte’s 2011 paper in our bibliography), and SIMHYD, than the single regime time series models and hold some promise for further investigations. Moreover, the ability of the TARX model structure to model catchment runoff differently under different catchment conditions (e.g. rainfall amounts) shows that the model structure has the potential to be more flexible than those of various conceptual models where there is one set of governing equations under all conditions.

In summary, a wide variety of models are used for rainfall-runoff modelling; conceptual models, time series models, artificial neural networks and so on. It is vital that scientists are aware of their potential limitations, particularly when investigating climate change scenarios. The response of a calibrated runoff model to changed rainfall inputs is crucial in this work and we have proposed that it be used to inform model choice.

Responses to individual points raised by Reviewer #1:
In the following we provide amendments or clarifications.

P919 Line 20 The first author accepts the advice offered by the reviewer that most studies in the Murray–Darling Basin (MDB) have run models on a daily time step. We would like to stress however, that when we wrote "When considering the effect on runoff of changes in rainfall, this type of study has tended to focus on models for annual or monthly runoff...." (Page 919, Lines 19-20) we did not make particular reference to the MDB. Our comment that models run on a daily time step were not as common as studies on monthly or annual time steps in studies of the effect of change of rainfall on runoff was informed by a sampling of studies in various geographical regions. We concede that our sample may have not been large enough to ascertain which is the most popular time step. We note that a recent Compendex search by the authors for the terms "daily", "rainfall-runoff" and "Murray" in all fields produced only four hits and did not produce either of the two publications cited by the reviewer. These results give the impression that there is still scope for daily time step rainfall-modelling modelling in the MDB. Further, none of these hits appeared to be a comparison of models. We would like to note that the main feature of our paper is not the geographical location of the catchments considered nor the time step of the models used. These are choices that must be made to allow us to proceed to compare different model classes and also consider the novel TARX model application.

The first author has made a quick reading of the second publication cited by the reviewer, "Estimating climate change impact on runoff across southeast Australia: Method, results, and implications of the modeling method" by F.H.S. Chiew *et al.* This paper uses results from Global Circulation Models (GCMs) to obtain scaling factors applied to historical rainfall data to produce "future" rainfall. We draw attention to the text on Page 4, "The GCM data are obtained from the Program for Climate Model Di-

agnosis and Intercomparison (PCMDI) Web site...". This suggests that the community of developers of GCMs sees value in maintaining a program of comparing the output of rival models, and a similar practice could be adopted by the rainfall-runoff modelling community.

We also note that even though the abstract of the paper cited above lists only SIMHYD, Sacramento is also used and results are compared. Again, models considered are confined to the field of conceptual models. Also, we note on Page 12 "To ensure that the modeling results can reasonably reproduce the daily observed funoff series, only calibrations with NSE [Nash–Sutcliffe efficiency] of daily runoff greater than 0.4 are used...". We note that in our paper, except for Model 6, (a model linear in rainfall and quickly rejected) all time series models considered showed NSE performance above this cutoff. In view of this, we consider that there is scope for discussion of how acceptable model performance is defined.

P923 This point was addressed by comments in our "Response to Reviewer #1's substantive objection" above on the use of time series models and the apparent lack of comparisons with other model classes.

P926 Line 10 Of the conceptual models listed in the reviewers' comment, SIMHYD, Sacramento and AWBM are part of the Rainfall Runoff Library. Our paper used SIMHYD for comparison with the time series models. The conceptual models, like time series models, calculate quantities at a point in time and use these values in calculations at the next time point. The Rainfall Runoff Library documentation does not provide any information on how the package deals with missing data and what possible effects on the model calibration and validation this may have. As the intent of this study was to consider the behaviour of a selection of rainfall-runoff models on a selection of catchments, we decided to confine interest to the models and data, rather than introduce some effect of how missing data is handled. We note that it is relatively easy to allow for missing data with time series models.

Page 926 Line 20 Minimizing the residual sum of squares (RSS) and maximizing the Nash–Sutcliffe coefficient in a model calibration are equivalent. Minimizing the RSS was chosen as the method for estimating parameters in calibrations as this is the criterion used in R by the lm function when fitting the time series models considered to data and also by the tvar function when fitting the threshold autoregressive exogenous input model to data. Given this, it was decided for consistency to choose minimization of the residual sum of squares as the criterion for goodness of fit when applying SIMHYD from the Rainfall Runoff Library to data.

Page 926 Line 24 We did not intend to suggest that the current day's rainfall was not used in the prediction of the current day's runoff. This issue is resolved by replacing the text on Lines 24 and 25 "... but uses rainfall up to the day of prediction." with "... but uses rainfall up to and including the day of prediction." With this change, the text is consistent with the explanation of runoff calculation given on Page 930, Line 5.

P927 Line 21 The reviewer asks for further details on model calibration within the programming language R. This could be added as an appendix.

The R function "lm" is used to fit to data a single regime time series model that is linear in the parameters. This is achieved by minimizing the residual sum of squares where each residual is the difference between the value calculated by the model and the datum. For such an objective function, fitting a model with lm is a special case of model calibration. To illustrate this, given an observation $y_i,\ i = 1, \ldots, n$, a single regime time series model that is linear in the parameters gives a corresponding prediction $\hat{y}_i = \mathbf{x}_i^\top \beta$, where the column vector $\mathbf{x}_i$ represents the covariates present in the model (a 1 for the intercept term, lagged runoff or rainfall observations and the current day's rainfall), $\top$ denotes the transpose of a matrix or vector, and β represents the vector of parameter values. A matrix equation for the vector of predictions $\hat{\mathbf{y}}$ in terms of a matrix of covariates $\mathbf{X}$ (which has as its ith row the vector $\mathbf{x}_i^\top$) is $\hat{\mathbf{y}} = \mathbf{X}\beta$. When fitting the model to data with the lm function, if lm determines that the columns of $\mathbf{X}$ are linearly independent then it returns the least squares estimate of β, here $\hat{\beta}$, defined by the

matrix equation

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y},$$

which has a unique solution. This process for estimating the parameter vector differs from the calibration of models which are nonlinear in the parameters, such as SIMHYD, where optimization of the objective function requires iterative exploration of the parameter space.

The function "tvar", contained within the R package tsDyn, allows the user choice over the type of threshold autoregressive model they wish to fit to multivariate data. In our application we chose to consider a model with a single threshold and hence two regimes. Each regime has its own parameters. At the time of using tvar, the latest version of software was 0.7-40 with documentation marked as last revised 11/03/2008. This documentation did not make explicit comment on the tvar function. Within R, the tvar function has a brief help page which mentions that "... estimation can be done directly by CLS (Conditional Least Squares)." A search of "The Comprehensive R Archive Network" (http://cran.r-project.org/) has shown that there is now a newer version of the documentation with date stamp 2010-08-11 that lists tvar, but the information given does not appear to add to that given on the brief R help page referred to above.

From our observation of the output of the tvar function, the fitting process attempts to determine the combination of parameter values and the value of the threshold variable that minimizes the residual sum of squared residuals. Thus, the objective of the model fitting process is the same as that dictated by the use of R function lm for our single regime time series models and that chosen for SIMHYD.

P928 This was an oversight. We accessed the information on all four catchments from the Australian Government, Bureau of Meteorology Water Resources Station Catalogue, (http://www.bom.gov.au/hydro/wrsc). When we wrote "The catchment runoff data used to calibrate SIMHYD is that ascribed to an Australian Bureau of Meteorology

(ABoM) flow station in the catchment", we did not have in mind the owner of the flow station. A greater level of detail for all four stations follows, informed by the website given above.

The reviewer is correct on the NSW flow stations. Both stations 401013 and 401008 used in our Catchments 2 and 4 respectively are listed with owner DNR (NSW).

The flow station in our Catchment 1, Flow station #403213A, has two records. The one we used opened in 1958 has "owner" listed as DSE (VIC). The second Victorian station, #403205, present in our Catchment 3 has owner listed as BOM (VIC), [Bureau of Meteorology (Victoria)] not DSE (VIC) as claimed by the reviewer.

P934 Line 7 The first author does not have much familiarity with"parameter equifinality" raised by the reviewer, except for a brief reading of the introductory chapter of "Rainfall-Runoff Modelling: The Primer" by K. J. Beven, (2001). The comment on equifinality given on Page 22 of this text "It [equifinality] suggests that, given the limitations of both our model structures and observed data, there may be many representations of a catchment that may be equally valid in terms of their ability to produce acceptable simulations of the available data. In essence then, different model structures and parameter sets used within a model structure are competing to be considered acceptable as simulators. Some may be rejected in the process of evaluation of different model structures suggested in Section 1.7, but even if only one model is retained then the evaluation of the performance of different parameter sets against the observed data will usually result in many parameter sets that produce acceptable simulations."

The last sentence of the quote on the existence of multiple sets of parameter values giving an equally good fit of a model to data, looks to be describing the problem known in statistics as numerical or *a posteriori* parameter identifiability. The first author has some knowledge of the area as it relates to his thesis topic.

Our single regime time series models, Models 1 to 6, are linear in the parameters. As a result, our response given to the reviewer's point on **P927 Line 21** above applies here

- the lm function determined that the least squares estimate of the parameter vector is unique. We conclude that concerns of parameter equifinality do not apply for such models.

For our study with the Rainfall Runoff Library's SIMHYD, we did not see any mention of equifinality in the user guide. However, having an awareness of parameter identifiability issues led us to, as described in our section 2.2.4 "Calibration and validation of SIMHYD", Page 928, Lines 10-11, "... the optimizer used to minimze the RSS was initially pattern search multi-start with ten starts". Following this process, the Rainfall Runoff Library selects the parameter vector that gave the best fit of the model to data as the starting point for further application of an optimizer. Given that the default number of starts was four, we considered the use of ten starts to be a good attempt to explore the parameter space to find the parameter vector that actually minimized the residual sum of squares, rather than merely finding a local minimum of this objective function. We note that this approach does not guarantee that one will find the parameter vector(s) that give the actual minimum of the residual sum of squares.

As mentioned in our response to the reviewer's query on **P927 Line 21**, the documentation for the tvar package in R is quite brief and hence we are not clear on how the parameter space was explored in the fitting procedure. Given the understanding of this issue in the statistical community, we find it likely that the tvar code would use a number of random starts and will endeavor to check this.

Returning to the first part of Beven's text given above, the notion that different model structures may produce comparable representations of observations appears to support our interest in the comparison of models and our extended framework (beyond NSE alone) for considering their properties.

P934 Line 9 The reviewer is correct on the points raised as far as comparison of Nash–Sutcliffe coefficients are concerned. There was a late change to the Nash–Sutcliffe values for validation (E_v) in Model 4's table of results which had some flow on effects.

The Nash–Sutcliffe values for calibration and recalibration (E_c and E_r respectively) are slightly better for Model 4 compared to those of Model 3 but indeed the E_v for Model 4 are substantially lower than those for Model 3 in three of the four catchments.

The sentence on Page 934 Lines 10-12 should be replaced by “Model 4 produces slightly higher Nash–Sutcliffe values E_c and E_r for all four catchment compared to those of Model 3, however Model 3 has substantially better E_v values, particularly for three of the four catchments." In a similar manner, the sentence on Page 934 Lines 12-14 should be amended to “Model 5's E_v values are substantially larger than those seen for Model 4."

We note that Model 4 was proposed for consideration initially to obtain insight into the behaviour of a model which has quadratic rainfall and runoff terms. Table 3 shows that Model 4's E_c and E_r values are in some cases better than those determined for SIMHYD and for Catchment 3 the E_c, E_v and E_r values for Model 4 were substantially better than those obtained for SIMHYD. We thought this provided some justification for considering Model 4 at the next stage, however, the instability of the model when making multi step ahead predictions (as noted on Page 935) showed that Model 4 was not suitable leading to its rejection.

We are aware of the similarity of results from Models 1, 2 and 3, which is why Model 2 was omitted from the rainfall scaling part of the analysis. We felt it useful to analyse Models 1 and 3 to ascertain whether Model 3 with its extra parameters and squared rainfall terms would exhibit different behaviour under the rainfall scaling compared to that of Model 1. We do not agree that the behaviour of Models 1 and 3 are indistinguishable in this part of the analysis.

The behaviour of Model 1 may appear qualitatively similar to that of Model 3, but there are some differences. These are summarized by Model 1 results in (10) on Page 936 and Model 3 results in (12) on Page 937. Model 1 shows that the ranking of rainfall scaling factors determined for Catchments 1 to 4 is unchanged whether the rainfall

scaling factor is greater than or less than one. Model 3's results are not symmetric; the ordering of rainfall scalings is quite different depending on whether the rain input is increased or decreased. For Model 1 the catchments either attenuate or amplify the effect of the rainfall scaling in the runoff scaling, regardless of the value of the rainfall scaling considered. Catchment results for Model 3 do not have this property.

The differences between Model 1 and 3 are also evident for runoff scaling predictions resulting from the more extreme values of rainfall scaling. As one example, for Catchment 3 for a rainfall reduction of 20$\%$, Table 8 shows that Model 1 predicts a theoretical runoff scaling of 0.750, whereas Model 3 predicts a value of 0.791. Similarly, for a rainfall increase of 20$\%$ on Catchment 1, Model 1 predicts a theoretical runoff scaling of 1.156, whereas Model 3 predicts a value of 1.211.

To illustrate the importance of the difference consider the Brisbane River, given its prominence in recent Queensland flood events. For mean annual runoff 1,112,933 Megalitres per year, (http://www.anra.gov.au/topics/water/availability/qld/swma-brisbane-river.html, last updated May 13th 2009) we have mean daily runoff of approximately 3049 ML/day. Suppose that a rainfall-runoff model reproduces this mean annual runoff. Under a 20% increase in rainfall for this historical mean daily flow, the expected mean daily flow using Model 1 would be 3525 ML/day and using Model 3, 3692 ML/day, a difference of 167 Megalitres per day. The mean daily flow difference between that predicted by the two models by this rough calculation is around 29% of Brisbane's daily water usage of 575 Megalitres, (http://www.seqwater.com.au/public/catch-store-treat/dams/wivenhoe-dam, page updated 28th February 2011) showing that the difference of results predicted is not negligible.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 917, 2011.