Assessing the performance of an ensemble approach to rainfall—runoff modelling for prediction of the impact of climate change on streamflow

Richard Silberstein[a] and **Santosh Aryal[b]**

[a] *CSIRO Land and Water, Underwood Ave, Floreat, WA, 6014*
[b] *CSIRO Land and Water, Acton, ACT, 2601*
Email: Richard.Silberstein@csiro.au

Abstract: Climate change over the last four decades has resulted in a significant decline in streamflow in south-west Western Australia, and there is concern for the state of these streams under a future climate. In order to estimate future streamflow, a major exercise was undertaken to simulate the flows under a range of future climate projections. A suite of five rainfall—runoff models was calibrated against observed flow in 106 gauged catchments across the region. The results from weighted means of 31 ensembles, made up of all of the linear additive combinations of the five models, were examined for their ability to reproduce the measured streamflows. The mean of the daily runoff from the Sacramento and IHACRES models consistently resulted in a better model fit than any other combination of the individual models and this 'adopted model' was used to model the runoff under the climate projections. This combination produced the best overall objective function, as determined by the model efficiency of Nash and Sutcliffe (1970) and average bias of less than 2% limit.

Keywords: *Rainfall-runoff modelling, south west Western Australia, ensemble modelling, Sacramento, IHACRES*

1.   INTRODUCTION

Global warming resulting from increased atmospheric greenhouse enhancing gases has been attributed as the cause for many environmental changes, such as bleached coral reefs, insect susceptibility and death of pine forest in North America and increased severity of drought and frequency of forest fires. With increased temperatures global rainfall is predicted to increase but in some areas already short of water there has been a reduced rainfall which is projected to worsen in future (e.g., Abbaspour et al., 2009; Arnell, 1992; Arnell and Reynard, 1996). Countries with Mediterranean type climate appear to have had the greatest reductions in rainfall and consequent runoff generation and streamflow (Ragab and Prudhomme, 2002; D'Agostino et al., 2010). The reduction in rainfall and streamflow has been especially severe in the south-west of Western Australia where a 16% decline in rainfall since the mid-1970s has resulted in a 55% decline in streamflow to water supply reservoirs (McFarlane et al, 2011; Silberstein et al, 2011; Charles et al, 2011). Thus, projections by GCMs of further reductions in rainfall in the region have major implications for water resources planning.

To investigate the impact of future trends in rainfall on streamflow, we applied a selection of rainfall–runoff models, forced with the downscaled projections of GCMs. Applying the assessment over the entire south-west region of Western Australia (204 catchments with 106 gauged catchment areas in 34,000 km^2) required that the methodology for setup, operation and calibration over the domain was relatively simple, could be automated, and was robust in catering for the variations in catchment characteristics and climate. Use of a suite of models with different formulations improved the likelihood that the full range of simulated responses would be covered. The calibration results for each catchment were assessed to decide which models and calibration methods were to be used in the scenario simulations. This paper describes the process undertaken to calibrate the models and select the best for use in the future runoff projections.

2.   METHODOLOGY

2.1.   Conceptual rainfall-runoff models

A set of simple conceptual rainfall-runoff models was selected based on the ability to automate the inputs of rainfall and potential evaporation (PE), distributed across the catchments, and to automate the optimisation of model parameters. To ensure as far as possible that results were not significantly influenced by the model structure and calibration method, models tested had a range of conceptual structures and the same approach for optimisation and testing was used for all models. Therefore the results projected for future scenarios were deemed more a result of the forcing conditions, namely changes due to climate change, than to vagaries of model structure (Silberstein et al., 2012; Silberstein et al., 2013).

The models tested were SIMHYD, with a Muskingum routing method (Chiew et al., 2002; Tan et al., 2005), Sacramento (Burnash et al., 1973), SMAR with a groundwater component added (SMARG) (Kachroo, 1992; Goswami et al., 2002), IHACRES (Jakeman et al., 1990; Littlewood et al., 1997), and AWBM (Boughton, 1996). The models and calibration optimisers were part of the Catchment Water Yield Estimation Tool (CWYET), expressly built for the purpose (Vaze et al., 2011). The application of each model used distributed climate inputs with runoff generation occurring in each cell across the catchment and summed at the catchment outlet. The catchments were small enough that explicit flow routing was not required. There is no regulation on flow in these rivers. Each cell within the model catchment uses the same parameter set. All models were implemented to accept spatially distributed rainfall and PE input sequences, applied on a 0.05 x 0.05 (longitude x latitude) grid (approximately 4.5 km x 5.5 km grid cells at the latitude of the region, 31 - 32 S). The climate data were extracted from the SILO Data Drill (www.longpaddock.qld.gov.au/silo, Jeffrey et al., 2001) and the future climate scenarios generated from these (Charles et al., 2010; Silberstein et al., 2012), although these future scenarios are not discussed further in this paper. The potential evaporation formulation used was the areal wet environment of Morton (1983).

2.2.   Catchment selection

The models were calibrated on all flow data available from 1975 to 2007 from 106 gauged catchments distributed throughout south-west Western Australia (WA) (Figure 1). This period was chosen because of the rainfall reduction and change in flow regime in south-western Australian rivers mentioned above (McFarlane et al., 2012; Silberstein et al., 2012). The catchments are almost all in water reservoir reserves, almost entirely covered in native dry sclerophyll eucalyptus forest and mainly managed for water supply, timber harvesting and natural conservation, although there is some mining followed by rehabilitation with natural species. The stream gauging stations are managed by the Department of Water, WA, (DoW) who provided all data and gave the quality assurance on the data set. Not all stations had continuous record through this

period, but all had at least 10 years of good quality record after 1974 (average period 26 years). They also had an area of at least 10 km^2 (average area 380 km^2) and mean annual runoff across the catchments ranged from 5 to nearly 350 mm with a median value of 112 mm and the runoff coefficient ranged from 1 to 30 percent (Figure 2). The runoff coefficient tends to be higher in the high rainfall areas in the middle and southern part of the study region, but there is no systematic geographic trend, and no trend with catchment size (Figure 2c).
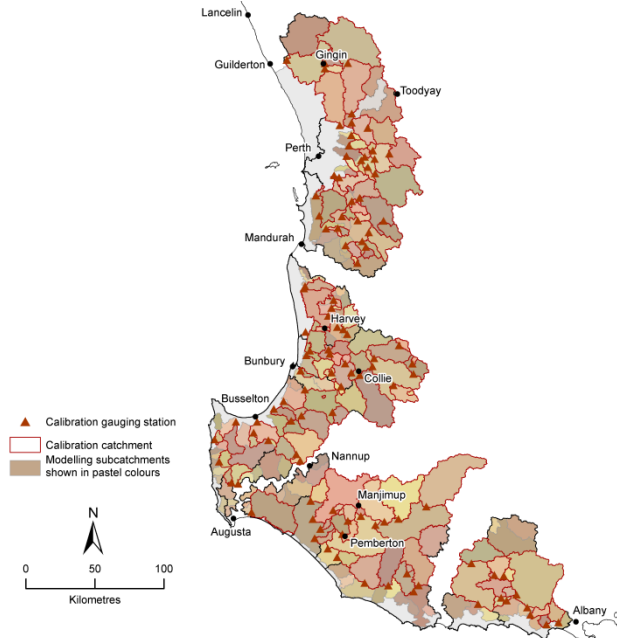

Figure 1. Locations of the stream gauges across south-west Western Australia used for this study.
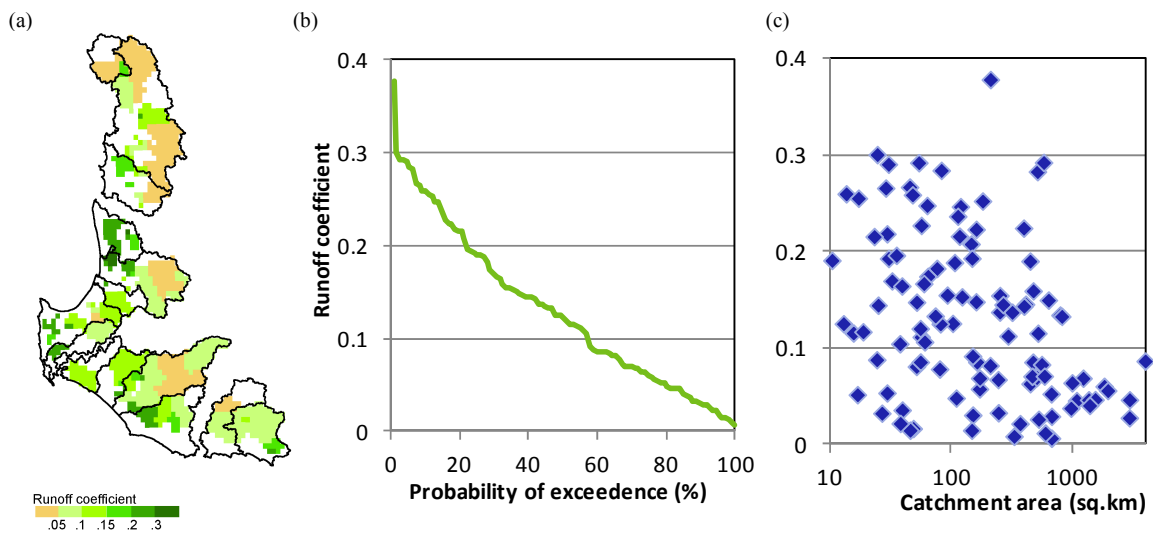

Figure 2. (a) Runoff coefficient across the study area, (b) frequency distribution of runoff coefficient, and (c) runoff coefficient plotted against catchment area

2.3. Model calibration

The automated CWYET framework used three optimisation algorithms, namely a genetic, a shuffle complex evolution and a uniform random sampling algorithm, and each was followed with the Rosenbrock local optimisation (Rosenbrock, 1960). The objective function used was the Nash and Sutcliffe (1970) modelling efficiency (NSE) with a 5% limit on model bias. The bias defined as the average percentage difference in yearly observed and modelled runoff. The NSE is influenced more strongly by high flows than low flows, but this was not considered a major issue for this project as the emphasis was on determining overall water

resources, most of which are carried by high flows. The models were all run with a 15 year 'spin-up' period prior to the calibration data commencement. The five models were calibrated separately and then the process of assessing the best ensemble was undertaken as a follow on process.

2.4. Selection of the model ensemble

There are an infinite number of options available to combine models into ensembles for catchment simulation. In this instance, arithmetic averages were made of all 31 possible combinations of the models, including the five models individually. While the calibration results indicated which models performed better and worse, no presumption of the best combination model was made prior to testing all possible combinations of the models, and unit weighting was given to all models in all combinations.

3. RESULTS

None of the models had a perfect fit to the observed data and the ensemble average approach was tested by combining simple arithmetic averages of the daily flow computed by every combination of two, three, four and all five models, and comparing these combined model averages against the observed sequences, and NSEs calculated for every combination (Table 1).

Table 1. Proportion of the 106 test catchments for which the NSE (Nash and Sutcliffe, 1970) met range criteria for all 31 combinations of the five models

Nash-Sutcliffe	AWBM 1	SIMHYD 2	Sacramento 3	SMARG 4	IHACRES 5					
NSE>0.8	24%	34%	74%	32%	54%					
0.6<NSE≤0.8	73%	48%	24%	49%	45%					
0.4<NSE≤0.6	4%	11%	0%	11%	1%					
0<NSE≤0.4	0%	4%	2%	2%	0%					
NSE≤0	0%	3%	1%	6%	0%					
	1 & 2	1 & 3	1 & 4	1 & 5	2 & 3	2 & 4	2 & 5	3 & 4	3 & 5	4 & 5
NSE>0.8	51%	69%	49%	54%	70%	51%	55%	65%	81%	60%
0.6<NSE≤0.8	45%	30%	45%	45%	28%	37%	44%	30%	18%	36%
0.4<NSE≤0.6	3%	1%	2%	1%	1%	8%	0%	2%	1%	1%
0<NSE≤0.4	0%	0%	3%	0%	0%	3%	0%	1%	0%	1%
NSE≤0	1%	0%	1%	0%	1%	2%	1%	2%	0%	2%
	123	124	125	134	135	145	234	235	245	345
NSE>0.8	70%	52%	61%	68%	70%	62%	66%	76%	59%	75%
0.6<NSE≤0.8	29%	42%	38%	28%	30%	34%	28%	23%	37%	22%
0.4<NSE≤0.6	0%	5%	1%	3%	0%	3%	5%	0%	3%	3%
0<NSE≤0.4	1%	0%	0%	0%	0%	0%	0%	1%	0%	0%
NSE≤0	0%	1%	0%	1%	0%	1%	1%	0%	1%	1%
	1234	1235	1245	1345	2345	12345				
NSE>0.8	67%	71%	61%	76%	73%	71%				
0.6<NSE≤0.8	31%	28%	36%	23%	26%	28%				
0.4<NSE≤0.6	1%	1%	2%	0%	0%	0%				
0<NSE≤0.4	0%	0%	0%	0%	0%	0%				
NSE≤0	1%	0%	1%	1%	1%	1%				

Clearly, the Sacramento model gives the best representation of the observed runoff, with 98% of catchments having NSE greater than 0.6 and 74% greater than 0.8. Note rounding of percentages in some combinations results in the total not being exactly 100%. IHACRES was clearly the second best performing of the individual models, and it is not surprising that the best combination is the average of these two (3 & 5 in Table 1) with 81% of catchments having NSE greater than 0.8 (Table 1 and **Error! Reference source not found.**a).

However the selection of "best ensemble" or indeed "best model" is not necessarily trivial (Ajami et al., 2004), as demonstrated in **Error! Reference source not found.**. While Sacramento had the best NSE in the great majority of cases, and the selected ensemble the best NSE in all cases, its overall bias was slightly greater than the AWBM/IHACRES ensemble (**Error! Reference source not found.**b). However, since the Sacramento/IHACRES ensemble had median absolute bias of 1.2% and an average absolute bias of 1.6%, and the average and median absolute bias for the AWBM/IHACRES ensemble were 1.5%, these are all good results and small differences in bias are inconsequential.
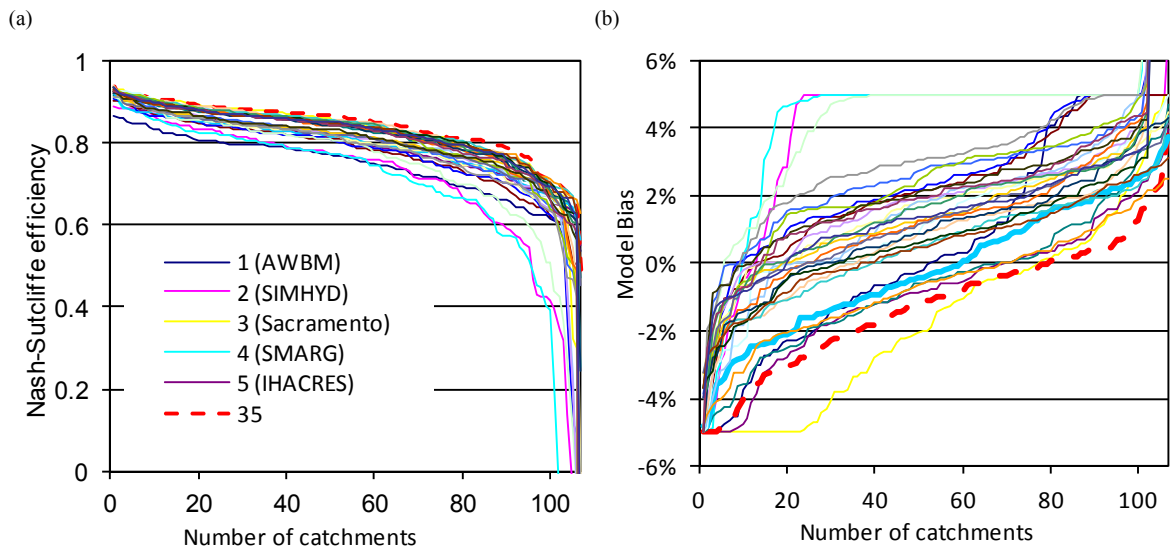
Figure 3. Frequency distribution of (a) Nash-Sutcliffe efficiency (NSE) and (b) model bias for the five models and all 26 model combinations. The heavy dashed red line shows the best model combination for NSE on both plots and the heavy cyan line shows the AWBM and IHACRES combination that had the least combined average and median bias

4. DISCUSSION

The linear combination of Sacramento and IHACRES consistently performed better in terms of higher NSE, than any of the other 25 combinations or the five individual models. This was consistent across the study area even with a wide range of climate (annual rainfall ranging from 350 to 1250 mm) and geographies (Figure 4). This is particularly interesting as these two models represent the two extremes in model complexity with Sacramento having, arguably, the most physical representation of a catchment and 17 parameters, and IHACRES the simplest in concept, with no explicit process representation, although more parameters than AWBM and SIMHYD. Both are well credentialed with Sacramento used in many hydrological studies (over 500 citations in Google Scholar) and IHACRES (over 2,000 citations in Google Scholar) derived from time-series analysis origins, with its structure in this application tailored to hydrological applications and has often been found to give a good fit to data sets.

The approach tested here differed from that suggested by Viney *et al.* (2005) in which a model ensemble result was calculated by weighting each model according to the degree with which each model reproduced measured flows in the calibration data set. A weighting of zero (i.e. the model results were excluded) was used for poorly performing models, and weighting of one for perfect models. Viney et al. (2005) tested weighting according to the value of the maximised objective function, which in our case would be the NSE (Nash and Sutcliffe, 1970) with a 5% bias restriction. This is only one of many objective functions that could be used to assess model performance (Gupta et al., 2012). It gives a preferential weighting to higher flows, which in this application is appropriate as the key aim of the project was to reproduce the total water resources in streams and investigate the changes in these due to climate change. If the application had been, for example, centred on changes to ecological water provisions, then alternative measures would be better. In that case the NSE could be used on low flows or on number of no flow days (Demarty et al., 2004). We also examined the NSE on daily flow duration curves, annual total flows, and wet season and dry season flows. It was found that in almost all cases this combination remained the best performing.

There are some concerns about the combination of models into an ensemble and whether this is an appropriate strategy for hydrological simulation. In the project undertaken we assert that it is indeed appropriate as the aim was to develop a technique to give the best possible reproduction of the observed flow sequence. Although not discussed here, we also undertook "split sample" tests, in which a subset of data was used for calibration while the remainder was used to test the objective function. This process provides a greater measure of the appropriateness of the model selection by testing on a new dataset. However, it does not necessarily assist the practical aim of the project as it reduces the learning data set for the calibration.

Computer simulation of natural systems is always prone to uncertainty. Despite good quality data, only some of these uncertainties are measurable and can be accounted for. The 'equifinality' problem (Beven, 2006) in modelling, due to different combinations of parameter values producing the same results, is hard to quantify and, as for the study reported here, is not necessarily an operational problem if the focus is to reproduce an integrated observed behaviour such as total flow. While the reliance on NSE with 5% bias limit in this project produced a good result, the results may well be improved by examining in more detail the components of the objective function score as discussed by Gupta et al (2009).
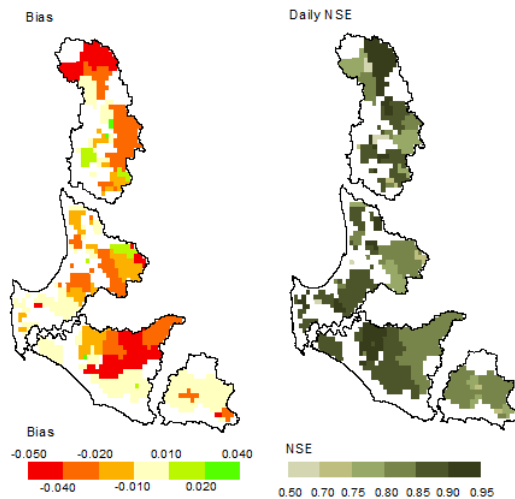

Figure 4. Geographical distribution of (a) model bias and (b) Nash-Sutcliffe modelling efficiency

5. CONCLUSIONS

In this paper we assessed the performance of a set of conceptual rainfall-runoff models of varying complexities, from the simplest, a time-series transfer function model, to a 17 parameter model that has some level of process representation. The rationale was to limit modelled process complexity to the minimum necessary, always ensuring the models were 'fit for purpose' and only as complicated as necessary while as simple as possible. While the most detailed model (Sacramento) produced the best overall objective function, the simplest (IHACRES) produced the second best, and the arithmetic average of these two resulted in the overall best ensemble combination. However, the selection of 'best model' is not necessarily straight forward, and is dependent on the application required. Appropriate combination of selection of objective function, requirement for minimum absolute or average bias and simplicity, or otherwise, of model, by which parsimony of parameter selection may be significant, all need to be considered in addressing any particular modelling problem (Vaze et al., 2012).

6. ACKNOWLEDGMENTS

REFERENCES

Ajami, N.K., Gupta, H., Wagener, T. and Sorooshian, S., 2004. Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. Journal of Hydrology, 298(1-4), 112-135. doi:10.1016/j.hydrol.2004.03.033.

Beven, K., 2006. A manifesto for the equifinality thesis. Journal of Hydrology, 320(1-2), 18-36.

Boughton, W.J., 1996. AWBM water balance model calibration and operation manual, CRC for Catchment Hydrology, Australia.

Burnash, R.J.C., Ferral, R.L. and McGuire, R.A., 1973. A Generalised Streamflow Simulation System – Conceptual Modelling for Digital Computers, US Department of Commerce, National Weather Service and State of California, Department of Water Resources.

Chiew, F.H.S., Peel, M.C. and Western, A.W., 2002. Application and testing of the simple rainfall-runoff model SIMHYD. In: Singh VP and Frevert DK (Editors), Mathematical Models of Small Watershed Hydrology and Applications. Water Resources Publication, Littleton, Colorado, pp. 335-367.

D'Agostino, D.R., Trisorio, L.G., Lamaddalena, N. and Ragab, R., 2010. Assessing the results of scenarios of climate and land use changes on the hydrology of an Italian catchment: modelling study. Hydrological Processes, 24(19), 2693-2704. doi:10.1002/hyp.7765.

Demarty, J., Ottle, C., Braud, I., Olioso, A., Frangi, J.P., Bastidas, L.A. and Gupta, H.V., 2004. Using a multiobjective approach to retrieve information on surface properties used in a SVAT model. Journal of Hydrology, 287(1-4), 214-236. doi:10.1016/j.jhydrol.2003.10.003.

Goswami, M., O'Connor, K.M. and Shamdeldin, A.Y., 2002. Structures and performances of five rainfall-runoff models for continuous river-flow simulation, 1st Biennial Meeting of International Environmental Modeling and Software Society, 1, Lugano, Switzerland, 476-481.

Gupta, H.V., Clark, M.P., Vrugt, J.A., Abramowitz, G. and Ye, M., 2012. Towards a comprehensive assessment of model structural adequacy. Water Resources Research, 48. doi:10.1029/2011wr011044.

Gupta, H.V., Kling, H., Yilmaz, K.K. and Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1-2), 80-91. doi:10.1016/j.jhydrol.2009.08.003.

Jakeman, A.J., Littlewood, I.G. and Whitehead, P.G., 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. Journal of Hydrology, 117, 275–300.

Jeffrey, S.J., Carter, J.O., Moodie, K.B. and Beswick, A.R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environmental Modelling and Software, 16, 309-330.

Kachroo, R.K., 1992. River flow forecasting. Part 5: Application of a conceptual model. Journal of Hydrology, 133(1/2), 141–178.

Littlewood, I.G., Down, K., Parker, J.R. and Post, D.A., 1997. IHACRES – Catchment-scale Rainfall-Streamflow Modelling (PC version) Version 1.0, The Australian National University, Institute of Hydrology and Centre for Ecology and Hydrology, Canberra, Australia.

McFarlane, D.J., Stone, R., Marten, S., Thomas, J., Silberstein, R.P. and Ali, R., 2012. Climate change impacts on water yields and demands in south-western Australia. Journal of Hydrology, 475, 488-498.

Morton, F.I., 1983. Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology. Journal of Hydrology, 66(1-4), 1-76.

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part 1 - A discussion of principles. Journal of Hydrology, 10, 282-290.

Ragab, R. and Prudhomme, C., 2002. Climate change and water resources management in arid and semi-arid regions: Prospective and challenges for the 21st century. Biosystems Engineering, 81(1), 3-34. doi:10.1006/bioe.2001.0013.

Rosenbrock, H.H., 1960. An automatic method for finding the greatest or least value of a function. The Computer Journal, 3, 175-184. doi:10.1093/comjnl/3.3.175. ISSN 0010-4620.

Silberstein, R.P., Aryal, S.K., Durrant, J. and Braccia, M., 2013. Rainfall—runoff model performance suggests a change in flow regime and possible lack of catchment resilience, MODSIM 2013. The Modelling and Simulation Society of Australia and New Zealand, Adelaide, this volume.

Silberstein, R.P., Aryal, S.K., Durrant, J., Pearcey, M., Charles, S., Braccia, M., Boniecka, L., Hodgson, G.A., Bari, M.A., Viney, N.R. and McFarlane, D.M., 2012. Climate change and runoff in south-western Australia. Journal of Hydrology, 475, 441-455. doi:doi:10.1016/j.jhydrol.2012.02.009.

Tan, K.S., Chiew, F.H.S., Grayson, R.B., Scanlon, P.J. and Siriwardena, L., 2005. Calibration of a daily rainfall-runoff model to estimate high daily flows, MODSIM 2005 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand, December 2005, 2960-2966. ISBN: 0-9758400-2-9.

Vaze, J., Perraud, J.-M., Teng, J., Chiew, F.H.S., Wang, B. and Yang, Z., 2011. Catchment Water Yield Estimation Tools (CWYET), 34th IAHR World Congress, 27th June to 1st July, Brisbane, Australia, 1554-1561.

Viney, N.R., Croke, B.F.W., Breuer, L., Bormann, H., Bronstert, A., Frede, H., Gräff, T., Hubrechts, L., Huisman, J.A., Jakeman, A.J., Kite, G.W., Lanini, J., Leavesley, G., Lettenmaier, D.P., Lindström, G., Seibert, J., Sivapalan, M. and Willems, P., 2005. Ensemble modelling of the hydrological impacts of land use change. In: A. Zerger and R.M. Argent (Editors), MODSIM 2005. International

Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, 2967-2973.