18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

The usefulness of bias constraints in model calibration for regionalisation to ungauged catchments

Viney, N.R. [1], **J. Perraud** [1], **J. Vaze** [1], **F.H.S. Chiew** [1], **D.A. Post** [1] **and A. Yang** [1]

[1] CSIRO Water for a Healthy Country National Research Flagship,CSIRO Land and Water, Canberra, ACT, Australia
Email: neil.viney@csiro.au

Abstract: Some modellers routinely calibrate their models using an objective function that that consists solely of some measure of error minimisation at the daily or monthly time step. Such a function might be based on maximising the Nash-Sutcliffe efficiency or the correlation coefficient or by minimising the root mean square error. Other modellers routinely incorporate a bias constraint into their objective function. There are a number of ways of incorporating a bias constraint. One way is to accept a calibration only if the difference in total streamflow between observation and prediction is less than some prescribed limit. Typically, this limit might be expressed as a percentage of the total observed streamflow. In this paper we examine whether the choice of calibration method—using a bias constraint or not—has an effect on the subsequent use of the calibrated parameter sets to regionalise predictions to ungauged catchments. We calibrate five lumped rainfall-runoff models—AWBM, IHACRES, Sacramento, Simhyd and SMAR-G—to 89 gauged catchments in Tasmania ranging in size from 10 to 3500 km^2. Three separate calibration runs are done for each model. In one, the objective function is based on the daily Nash-Sutcliffe efficiency alone; in the second, we augment the Nash-Sutcliffe efficiency with a severe bias constraint that attempts to ensure that the total predicted streamflow is within 5 % of the total observed streamflow (the bucket constraint); and in the third, we combine a smooth, less severe bias constraint (the log-bias constraint) along with efficiency into the objective function. As expected, overall efficiencies tend to be larger for the unconstrained calibrations, while overall absolute biases tend to be smaller for the constrained calibrations.

We then simulate each of the 89 catchments using parameters calibrated for the nearest neighbouring gauged catchment (cross-verification) and again assess prediction efficiency and bias. The results show that the regionalised predictions using parameters derived from constrained calibrations tend to have lower absolute biases than those using parameters derived from unconstrained calibrations. Regionalised model efficiencies tend to be greater for the unconstrained case than for the constrained cases, but the differences are slight. The smooth log-bias constraint is shown to provide cross-verification predictions with efficiencies almost indistinguishable from those for the unconstrained case and with biases at least as good as those of the more severe constraint. Furthermore, this constraint does not suffer from the numerical issues that can affect predictions using the non-continuously differentiable bucket constraint. These results provide support for incorporating bias constraints into calibration routines when model parameters are subsequently used for prediction in ungauged basins. In particular, the use of a smoother constraint, like the log-bias constraint, is recommended.

Keywords: *Rainfall-runoff modelling, calibration, bias constraint, regionalisation, ungauged basins*

1. INTRODUCTION

All conceptual rainfall-runoff models have unknown parameters which require calibration using an observed sequence of streamflow. Calibration inevitably involves some form of compromise. Often this compromise might revolve around choosing whether to emphasise particular parts of the hydrograph above others (e.g., fitting peak flows at the expense of low flow prediction accuracy). In a multi-response calibration, the compromise may involve trading prediction accuracy in one response in order to improve accuracy in another. Usually, the modeller will have available a variety of levers to facilitate weighting of the different aspects of the predictions. The use of these levers will largely be dictated by the nature of the modelling application.

Some modellers routinely calibrate their models using an objective function that that consists solely of some measure of error variance minimisation at the daily or monthly time step. Such a function might be based on maximising the Nash-Sutcliffe efficiency or the correlation coefficient or by minimising the root mean square error. Other modellers (e.g., Madsen et al., 2002, Chiew et al., 2009) incorporate a bias constraint into their objective function. There are a number of ways of incorporating a bias constraint. One way is to accept a calibration only if the difference in total streamflow between observation and prediction is less than some prescribed limit. Typically, this limit might be expressed as a percentage of the total observed streamflow.

In this application, we consider the need to obtain predictions with good efficiency and bias. Although these two measures are not entirely independent—it is difficult, for example, to obtain a calibration with extremely high efficiency and extremely poor bias—calibrating to both measures does include a degree of compromise. Furthermore, here we are not so much interested in the calibration quality as in the quality of the regionalised predictions that could be obtained as a result of the calibration.

2. STUDY AREA AND DATA

This study uses observed streamflow data from 89 gauged catchments in Tasmania, Australia (Figure 1). All the catchments have areas of greater than 10 km^2 and have streamflow records that are at least 30 % complete during the period 1975 to 2007. All available gauged catchments in the study area that meet these criteria have been chosen. Throughout Tasmania, mean annual precipitation varies from less than 550 mm in the southeast to more than 3400 mm in the west, and is winter-dominated. For the 89 study catchments, mean annual streamflow varies from 20 mm to more than 2400 mm, and runoff coefficients range from less than 4 % to more than 90 %.

Daily rainfall input data is obtained from the Silo Data Drill (Jeffrey et al., 2001), a data set gridded at a 0.05° (~5 km) spacing. The Data Drill rainfall data is interpolated from point observations of daily rainfall. Areal potential evaporation data is also derived from the Data Drill.

About half of the catchments are natural forested catchments while the remainder are agricultural (grazing and cropping). Some streamflows are affected to fairly minor degrees by impoundment or by irrigation withdrawal. Where possible, the streamflow data for these catchments has been augmented by engineers at Hydro Tasmania Consulting to reflect pre-extraction flows. Most are stand-alone catchments, but for others there are up to three levels of nesting.

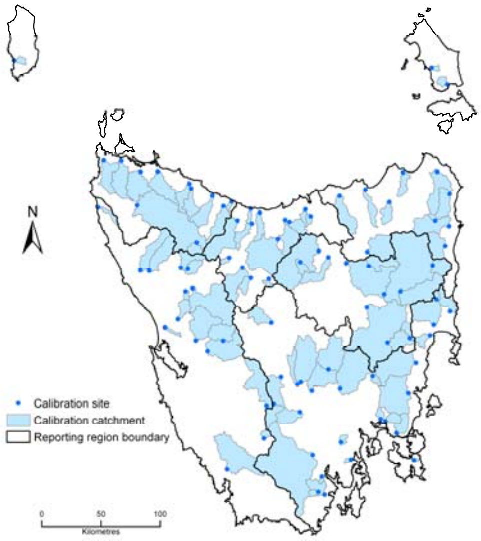

Figure 1. Location of study catchments.

3. METHODOLOGY

3.1. The rainfall-runoff models

Five lumped, conceptual rainfall-runoff models are calibrated separately on each of the 89 catchments: AWBM (Boughton, 2004), IHACRES (Croke et al., 2006), Sacramento (Burnash *et al.*, 1973), Simhyd (Chiew *et al.*, 2002) and SMAR-G (Goswami et al., 2002). All models have previously been applied widely in runoff modelling. In this study, six model parameters are optimised for Simhyd, including one parameter in a Muskingum routing algorithm (Tan et al., 2005). For the implementation of the remaining models, we optimise six parameters for AWBM, seven for IHACRES, 13 for Sacramento and eight for SMAR-G.

Each model is operated using the gridded rainfall and potential evaporation data in 0.05° x 0.05° grid cells across each catchment. For calibration, the observed runoff at the catchment outlet is compared with a spatial average of the modelled runoff in each grid cell within the catchment.

3.2. Calibration

Calibration is achieved through a sequential combination of the shuffled complex evolution algorithm and Rosenbrock methods. Tests of this procedure (not reported here) have shown it to provide reproducible results for the five models. Three objective functions are used. In one case—here termed the unconstrained case—the objective function is the Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970) of the daily runoff predictions. In the second approach, the objective function is also based on daily model efficiency. However, in this case—which we call the bucket case—a penalty is applied for any prediction whose overall bias (total model error divided by total observed streamflow) is greater than 5 %, with the penalty being proportional to the deviation beyond 5 % (Figure 2a). The objective function is given by subtracting the penalty from the efficiency. The third approach—which we call the log-bias case—uses an objective function that is a weighted combination of efficiency and a logarithmic function of bias given by

$$F = E_{NS} - 5 \,|\, \ln(1 + B) \,|^{2.5}$$

where E_{NS} is the Nash-Sutcliffe efficiency and B is the bias. The form of the log-bias constraint is shown in Figure 2b. The coefficients of this equation control the severity and shape of the resulting constraint penalty.

Whereas the bucket constraint is additively symmetrical (a 50 % underprediction is penalised the same as a 50 % overprediction), the log-bias constraint is multiplicatively symmetrical (a prediction volume that is twice the observation volume is penalised the same as a prediction volume that is half the observation volume). Note that the bucket constraint, as used by Chiew et al. (2009), is much more severe than the log-bias constraint. In Figure 2 it can be seen that a bias of –0.5 leads to a bucket penalty that is 450 times that of the log-bias constraint, while a bias of +1.0 leads to a bucket penalty that is 950 times that of the log-bias constraint.

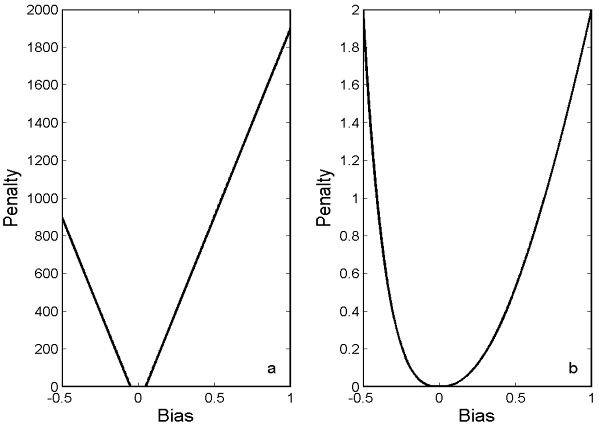

Figure 2. Graphical representation of the calibration penalties for a) the bucket constraint and b) the log-bias constraint.

Thus, these three calibration procedures can yield three separate and possibly different sets of optimised model parameters for each model.

3.3. Cross-verification

Each of the 89 catchments is simulated using donor parameters from the nearest, non-nested catchment of the remaining 88 (together with local climate input). The results of this cross-verification are assessed during the same period (1975–2007) as is used for calibration. The assessment criteria for cross-verification are the same as those used in calibration: efficiency and bias.

4. RESULTS

4.1. Calibration

The differences in calibration statistics between the three constraints are illustrated using the results for SMAR-G (Figure 3) for the 89 catchments. Not surprisingly, the best efficiencies are obtained for the objective function that maximizes efficiency only and has no bias constraint. The poorest efficiencies—especially for the poorly calibrated catchments—are generated using the most severe bias constraint: the bucket constraint. In general, the calibration efficiencies of the log-bias constraint are similar, but slightly inferior to those with no constraint.

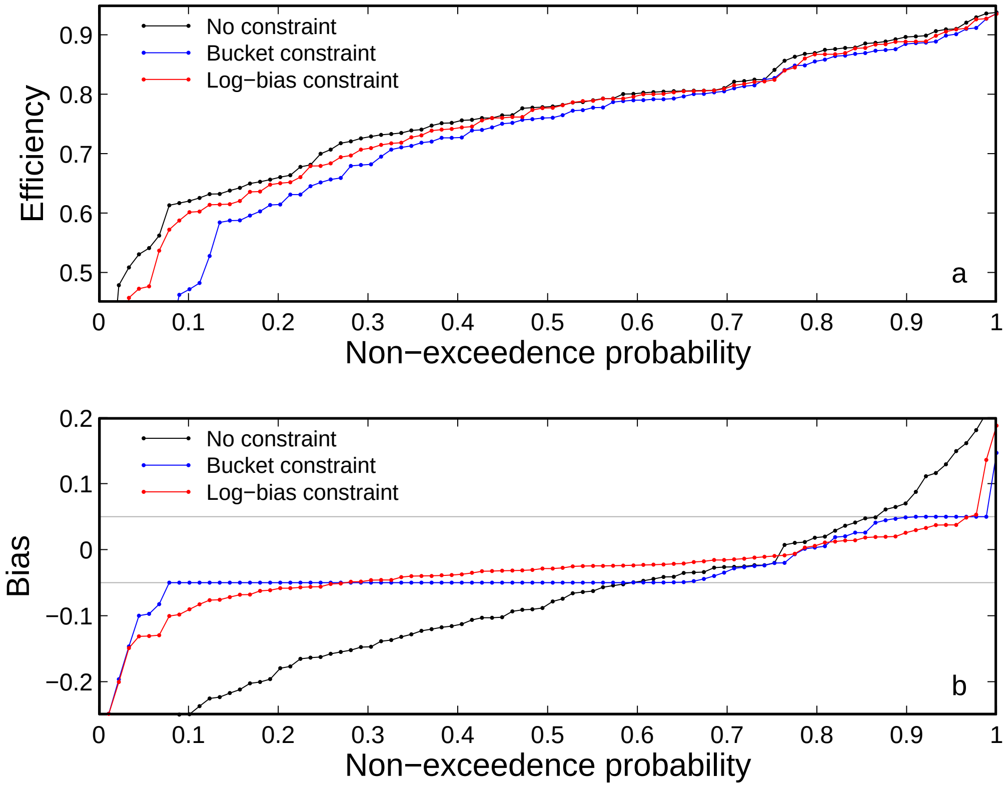

Figure 4. Cumulative probability plots of a) calibration efficiency and b) calibration bias for the SMAR-G model calibrated using three different objective functions.

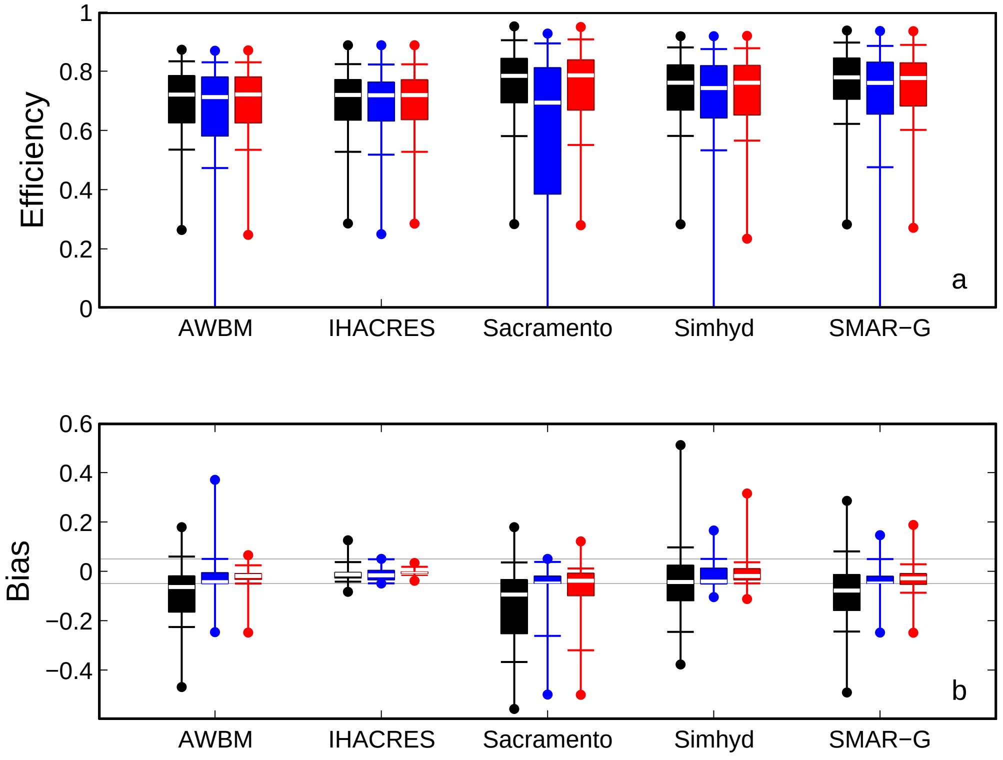

Figure 3. Box plots of a) calibration efficiency and b) bias for the five models for no constraint (black), bucket constraint (blue) and log-bias constraint (red). The boxes indicates the 25th, 50th and 75th percentiles, the whiskers indicate the 10th and 90th percentiles, and the dots indicate the extrema.

In contrast, the bucket constraint yields calibrations with the best bias characteristics. For SMAR-G, almost all catchments are calibrated with absolute biases of less than 5 %. However, despite the severity of this constraint, several catchments still have calibration biases outside these limits. In other words, the optimiser was unable to find a parameter set that yielded predictions within the bias limits. Furthermore, most of the bucket constraint biases that satisfy the limit fall right on the limit, at either –0.05 or +0.05, with relatively few occurring at intermediate biases. In comparison, the log-bias constraint yields fewer catchments with absolute biases of less than or equal to 5 %, but far more with absolute biases that are distinctly less than 5 %. The biases for the objective function with no constraint show far greater scatter and few fall within 5 %.

Broadly similar patterns are seen for the other four models (Figure 4), with the no constraint case having the best efficiencies (followed closely by the log-bias case) and the worst biases. The Sacramento model's calibration efficiencies for the bucket constraint are conspicuously poorer than those for other models or objective functions. Only for IHACRES are all 89 catchments calibrated within the 5 % bias limit for the bucket constraint, although the same is also true for the IHACRES log-bias calibration. While all three objective functions tend to slightly favour underprediction, this trend is most evident in the no constraint case, for which the median calibration biases are less than –0.05 for all models except IHACRES.

4.2. Cross-verification

Prediction statistics for both efficiency and bias are significantly poorer for cross-verification than for calibration. Whereas the median calibration efficiencies for the three objective functions for SMAR-G are between 0.76 and 0.78, the median cross-verification efficiencies fall to 0.60–0.62. The SMAR-G efficiencies for the unconstrained case retain their slight advantage over those of the log-bias constraint (Figure 5). In terms of cross-verification bias, there is little difference between the bucket and log-bias constraints. The unconstrained case continues to show a strong tendency towards underprediction and has a larger median absolute bias (0.21) than the two constrained cases (0.15).

As is the case for calibration, the cross-verification statistics for the other four models show similar trends to those of SMAR-G. The bucket constraint yields slightly poorer efficiencies in the poorly calibrated catchments, while there is little to separate the efficiencies of the other two objective functions (Figure 6a). The unconstrained case tends towards underprediction and has poorer median biases (Figure 6b), and—apart from IHACRES—has greater median absolute biases than the other two constraints.

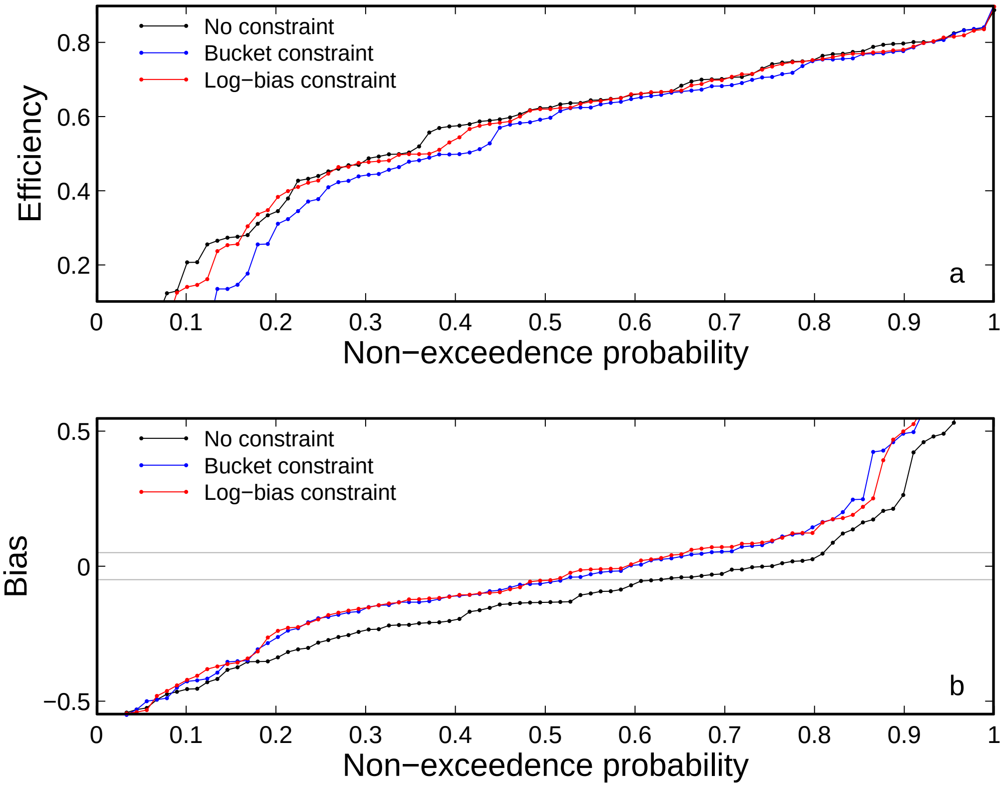

Figure 5. Cumulative probability plots of a) cross-verification efficiency and b) cross-verification bias for the SMAR-G model calibrated using three different objective functions.

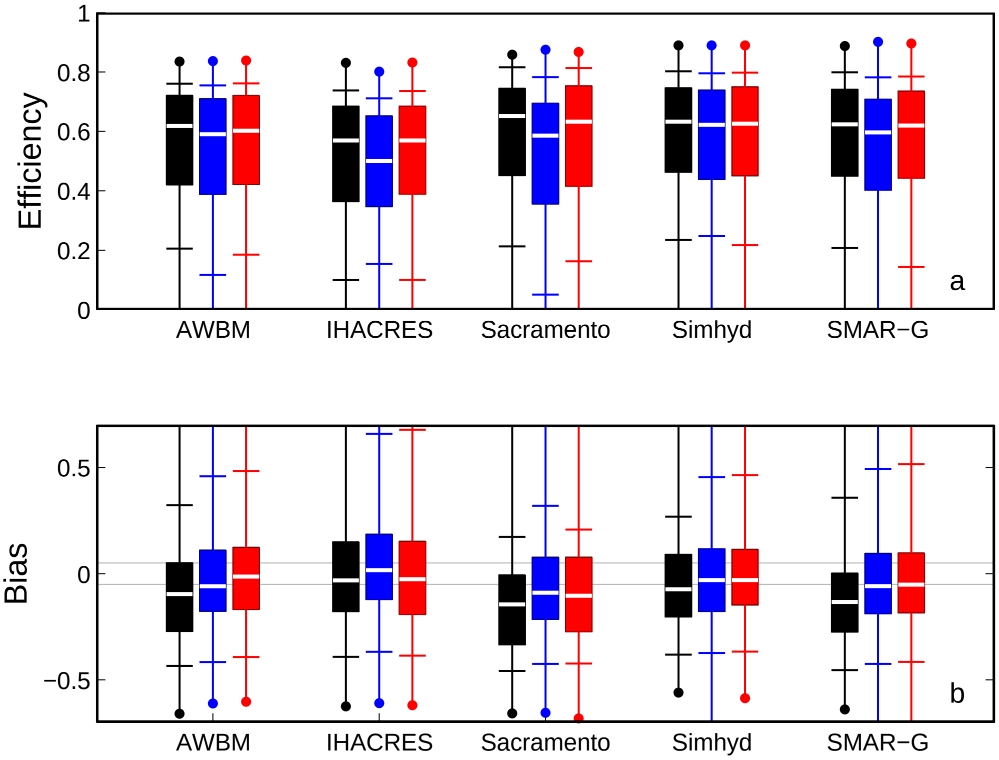

Figure 6. Box plots of a) cross-verification efficiency and b) cross-verification bias for the five models for no constraint (black), bucket constraint (blue) and log-bias constraint (red).

5. DISCUSSION

In the calibrations depicted in Figures 3 and 4, the best efficiencies are associated with the objective function that considers only efficiency, but its bias statistics are relatively poor. In contrast, the objective function with the most severe bias constraint has the best biases but the worst efficiencies. The log-bias constraint appears to provide a good compromise in that its efficiencies are almost as good as those of the unconstrained case and its biases are almost as good as those of the bucket constraint. However, more research is needed to elucidate whether this is a consequence of the relative severity of the constraints or of the functional form of the bias penalty.

One observation that points to the latter is that with the bucket constraint, most catchments have calibration biases that are exactly at the point where the bias penalty cuts in (–5 % or +5 %). It appears that during the calibration process, the optimum calibration evolves towards the constraint limit (from outside the limits). However, when it reaches the point where the penalty disappears, it becomes stuck and is unable to evolve further towards zero bias. This blockage occurs despite the fact that there are known parameter sets with better efficiencies and lower absolute biases—as evidenced by the outcomes of the log-bias constraint. It occurs in less than 20 % of catchments for IHACRES, but at least 45 % (and up to 65 %) for the other models. The problem is almost certainly associated with the non-differentiable nature of the bucket bias penalty; it does not (and cannot) arise with the continuously differentiable log-bias constraint.

In cross-verification, the efficiencies of the bucket optimiser continue to be worse than those of the other two, while the biases of the unconstrained optimiser are worse than those of the other two. The efficiencies and biases associated with the log-bias constraint are more or less commensurate with those of the best optimiser for each measure. However, with the larger scatter of results, the differences between the three objective functions are smaller in cross-verification than in calibration, especially for efficiency.

The biases of the IHACRES model present an interesting case study. In calibration (Figure 4), the biases are extremely small for all three objective functions, and significantly smaller than the biases for the other four models. This is most likely due to the presence in IHACRES of a parameter that effectively scales rainfall. However, in cross-verification, despite IHACRES appearing to not have as great a tendency towards underprediction as the other models, its biases have larger interquartile and interdecile ranges than the other models. It appears that the rainfall scaling parameter does not translate well to other catchments. Perhaps this is indicative of errors in the gridded rainfall data in areas of high spatial rainfall gradients.

A cross-verification study like this gives an indication of the quality of predictions that are likely to occur when parameters are regionalised to enable prediction in ungauged catchments. Where predictions are required on a single ungauged catchment, prediction efficiency is probably the most important metric, although a good bias is also desirable. In contrast, where broad-scale predictions are required in a large number of ungauged catchments across a region, the user is likely to be at least as interested in mean annual flows as in day-to-day flow fluctuations. In this context, bias becomes more important. Given these considerations, it would seem desirable to include some sort of bias constraint in model calibration. Compared to the unconstrained case, the use of either of the constraints assessed here leads to improved bias characteristics in regionalisation with little diminution in efficiency. In particular, the log-bias constraint would appear to provide the best solution.

6. CONCLUSIONS

Calibration of five rainfall runoff models in 89 catchments using three separate objective functions which give varying weights to efficiency and bias measures has shown that objective functions weighted more heavily towards efficiency give predictions with relatively poorer bias, while those weighted more heavily towards bias give predictions with slightly poorer efficiency. These trends persist when the calibrated parameters from nearest neighbour catchments are used in cross-verification tests. The log-bias constraint is shown to provide cross-verification predictions with efficiencies almost indistinguishable from those for the unconstrained case and with biases at least as good as those of the more severe bucket constraint. Furthermore, this constraint does not suffer from the numerical issues that can affect predictions using the non-continuously differentiable bucket constraint. These results provide cautious support for incorporating bias constraints into calibration routines when model parameters are subsequently used for prediction in ungauged basins. In particular, the use of the smoother log-bias constraint is recommended.

ACKNOWLEDGMENTS

This work is supported by the Tasmania Sustainable Yields project and by the water information research and development alliance between CSIRO’s Water for a Healthy Country Flagship and the Australian Bureau of Meteorology.

REFERENCES

Boughton, W.C., (2004). The Australian water balance model, *Environmental Modelling and Software*, 19, 943–956.

Burnash, R.J.C., R.L. Ferral and R.A. McGuire (1973), A generalized streamflow simulation system—conceptual modeling for digital computers. Tech. Rep., Joint Federal and State River Forecast Center, Sacramento, 204pp.

Chiew, F.H.S., M.C. Peel, and A.W. Western (2002), Application and testing of the simple rainfall-runoff model SIMHYD. In Singh, V.P. and Frevert, D.K. (eds.), Mathematical models of small watershed hydrology and applications, Water Resources Publications, Littleton, USA, pp. 335–367.

Chiew, F.H.S., J. Teng, J. Vaze, D.A. Post, J. Perraud, D. Kirono and N.R. Viney (2009). Estimating climate change impact on runoff across south-east Australia: method, results and implications of modelling method. *Water Resources Research* (in press).

Croke, B.F.W., F. Andrews, A.J. Jakeman, S.M. Cuddy and A. Luddy (2006). IHACRES Classic Plus: A redesign of the IHACRES rainfall-runoff model. *Environmental Modelling and Software*, 21, 426–427.

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

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

Madsen, H., G. Wilson and H.C. Ammentorp (2002), Comparison of different automated strategies for calibration of rainfall-runoff models. Journal of Hydrology, 261, 48–59.

Nash, J.E. and J.V. Sutcliffe (1970), River flow forecasting through conceptual models, I, A discussion of principles. *Journal of Hydrology*, 10, 282–290.

Tan, K.S., F.H.S. Chiew, R.B. Grayson, P.J. Scanlon and L. Siriwardena (2005), Calibration of a daily rainfall-runoff model to estimate high daily flows. Congress on Modelling and Simulation (MODSIM 2005), Melbourne, Australia, pp. 2960–2966.