Whole of system calibration of river models: Weighting functions and their effect on individual gauge and system performance

J. D. Hughes [a]**, S. S. H. Kim**[a]**, A. Yang**[a]**, D. Dutta**[a] **and J.Vaze**[a]

[a]*CSIRO Land and Water, Canberra, Australia*
Email: justin.hughes@csiro.au

Abstract: A recently developed method for parameter optimisation in river system models has shown potential to reduce over-fitting and improve gauge goodness of fit. This method optimises parameters across all gauges in the river system simultaneously. Optimisation requires a goodness of fit measure that aggregates performance across available gauges within the river system. To enable this, a weighting function has been used in previous studies. However the effect of the weighting function on individual gauge goodness of fit, parameter values and system states is unknown. System optimisation may have particular benefits for river models used for predictive purposes and scenario evaluations, and the effect of weighting methods needs to be assessed with these purposes in mind. The Murrumbidgee catchment in NSW was used as a test region to calibrate the AWRA-R river system model with various weighting functions. These weightings were; 1) Equal to all gauges, 2) by the length of the observed record and 3) by the sum of observed flows. In the lower reaches of the river system, weighting based on flow was lower than even weighting, since flow became progressively lower towards the downstream extremities of the study area. Goodness of fit was improved via even weighting relative to flow weighting in the lower gauges (at the expense of upper gauge fit). Model states show limited variability between weighting methods suggesting that the weighting scheme has less influence than the mode of calibration i.e., reach–by–reach or system.

Keywords: River system model, system optimisation, model state, weighting function, AWRA-R

1 INTRODUCTION

Demand for water continues to increase in line with needs from irrigation, industry, recreation and conservation. Many of these individual requirements will be in direct competition and conflict can ensue. These conditions create a strong need for management tools that can be used for water accounting and predictive purposes. River–system models can fulfill many of these management requirements, since they can account for important processes, storage, fluxes and demands in a systematic way (Welsh et al., 2012; Vaze et al., 2013).

River–system models feature a series of stream gauges linked in a river network. Flow at any gauge will, to some degree, influence subsequent gauges in the network. Hughes et al. (2014) showed that errors in the stream simulation are propagated down stream in river system models, and advocated the use of system-wide calibration procedure to reduce the deleterious effects of poor/problem gauges on the performance and parameter values for all downstream gauges in the river system model. To date, the recursive nature of the stream system does not generally influence the calibration system used since in most cases models are either calibrated manually (Hughes et al., 2006) or automatically on a reach-by-reach basis (Hughes et al., 2014). This is a practical approach to calibration but may increase the chance of over-fitting to problem gauges (Hughes et al., 2015).

Recently, Hughes et al. (2015) proposed and tested an optimisation system that calibrates parameters for all gauges in the stream network simultaneously. This approach demonstrated an ability to reduce over-fitting in individual reaches and and improve the overall goodness of fit, producing more satisfactory model states. The system calibration approach relies on a system goodness–of–fit measure that aggregates perfomance at each individual gauge location. The aggregating function or weighting vector was influenced by total observed flow at each gauge, which in turn is influenced by mean observed flow rate and number of observations (days). This weighting function was chosen rather arbitrarily, and the effect of alternative weighting functions on model states and performance was not explored. Here we aim to test the effect of different of weighting functions within a system calibration context and explore the effects in terms of model behavior.

2 METHODS

The area chosen for this study is the Murrumbidgee River system below Wagga Wagga (Figure 1). River flow is generated primarily to the east of the study area and flows to the west where the topography flattens and streams become divergent. River flow in the study area supports both the Murrumbidgee Irrigation Area (MIA) and The Coleambally Irrigation Area (CIA), two of the largest irrigation developments in Australia. Land use is predominately agriculture. The Murrumbidgee links with the Murray River system downstream of gauges 410133 and 410003. The study area features 23 stream gauges across more than 1600 km of stream channel.

The river system model used for this study is the Australian Water Resources Assessment modelling system (AWRA). The AWRA framework consists of a landscape model (AWRA-LG) and a river–system model (AWRA-R). For further information see Vaze et al. (2013). The AWRA-R model employs a semi-distributed network representing flow at each stream gauge location. Each node represents a gauged reach and will contain various hydrological procedures such as routing, irrigation diversion, loss to floodplains, ungaued inflows and losses to groundwater. For this study the routing parameters were calibrated on a reach-by-reach basis, while parameters controlling ungauged inflows, losses to groundwater and loss to floodplains were calibrated in a system context. Diversion to irrigation is estimated by an irrigation model that is calibrated outside the river system optimisation process.

2.1 Calibration

Calibration proceeds in two main steps, reach by reach calibration of model parameters within each reach, followed by system-wide optimisation of a subset of parameters related to reach losses and gains. Such a process assumes that routing parameters determined during reach-by-reach calibration are acceptable and do not require fine tuning in a system context. System-wide calibration optimises parameters in all reaches simultaneously.

The objective function used for each reach is a combination of Nash–Sutcliffe efficiency (NSE, Nash and Sutcliffe, 1970) utilising root transformed values and bias;

$$OF = \left(1 + \sum_{i=1}^{m} \left(\sqrt{\widehat{q}_i} - \sqrt{q_i}\right)^2 \bigg/ \sum_{i=1}^{m} \left(\sqrt{q_i} - \overline{\sqrt{q}}\right)^2\right) *$$

$$\left(1 + \left|\left(\sum_{i=1}^{m} \widehat{q}_i - \sum_{i=1}^{m} q_i\right) \bigg/ \sum_{i=1}^{m} q_i\right|\right) \quad (1)$$

where OF is the objective function score, q_i is the observed stream flow, and $\widehat{q}_i$ is the simulated stream flow. The function is minimised in optimisation and a perfect goodness of fit will return a score of 1. The higher the score the poorer the fit.
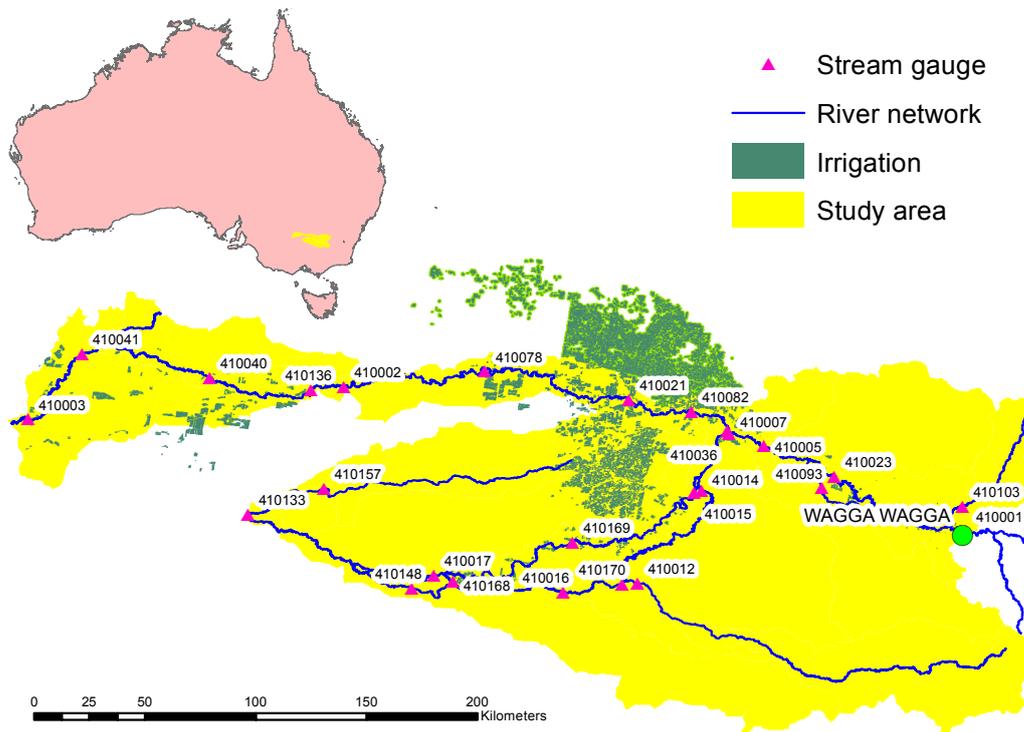
Figure 1. Location of study area showing stream gauges, river network and irrigated land.

During the first stage of calibration (reach–by reach), initial parameter values are obtained. In this mode, simulated flows for each gauge are used as the input to the the next reach. Optimisation occurs for each reach in turn, so parameters (6 per reach) are obtained to maximise the goodness of fit for each gauge in a sequential fashion. The second stage (system optimisation) optimises 4 parameters for each reach in system-wide calibration.

System-wide optimised parameters are related to reach loses and gains i.e.; ungauged inflow correction factor (k_r), flood plain return flow coefficient (k_{fpr}) and GW loss (α and β). The same ranges were used in reach–by–reach and system calibration.

A single objective function score was utilised for optimisation. This score needed to take into account the goodness of fit at all gauges simultaneously. This was achieved using a weighting function.

$$OF_{sys} = \sum_{j=1}^{n} OF_j * w_j \tag{2}$$

where OF_{sys} is the system objective score, and OF_j is the objective score for each gauge.

This study tests three different weighting functions to asses their effects on modelled system behavior. At its most simple, the weighting function could be simply the mean of all individual objective scores;

$$w_{even,j} = 1/n \tag{3}$$

where $w_{j,even}$ is the objective score weighting for each gauge j and n is the total number of gauges. This gives equal weighting to every gauge regardless of flow rates or length of observation record. Secondly, the weighting function can take into account the length of observation record for each gauge, giving more weight to those gauges with a longer records;

$$w_{days,j} = \sum_{i=1}^{m} c(q)_{i,j} \left/ \sum_{j=1}^{n} \sum_{i=1}^{m} c(q)_{i,j} \right. \tag{4}$$

where c is binary and indicates if an observation was made at gauge j on day$i$;

$$c(q) = \begin{cases} 1 & \text{if } q \text{ is observed} \\ 0 & \text{if } q \text{ is not observed} \end{cases} \tag{5}$$

However, this would also mean that gauges that with lower mean flow rates would have equal weighting to gauges with higher mean flow rates which was considered to be poorly constrained since one aim of calibration is to reduce the mass balance error of estimates. Given the potential importance of mean flow rate and length of observed record, the final weighting function used was the relative sum of all observed flows;

$$w_{flow,j} = \sum_{i=1}^{m} q_{i,j} \left/ \sum_{j=1}^{n} \sum_{i=1}^{m} q_{i,j} \right. \tag{6}$$

where w_j is the objective score weighting for each gauge j, and $q_{i,j}$ is the flow for each gauge j on each day i. The sum of all gauge weights for any weighting function will be 1.

System calibration uses the Shuffled Complex Evolution (SCE) algorithm (Duan et al., 1992), as implemented in the R "hydromad" package (Andrews et al., 2011).

3   RESULTS AND DISCUSSION

The river system model was firstly calibrated in reach–by–reach mode then calibrated in one of the system calibration modes, namely system–even (Equation 3), system–days (Equation 4) or system–flow (Equation 6). Performance at each gauge was assessed using the objective function score outlined in Equation 1. Results for each calibration system/weighting function are shown in Figure 2. The most obvious difference between the calibration methods is that, in general, goodness of fit is not radically different. The median goodness of fit appears to be improved with system calibration methods relative to reach–by–reach calibration, but there are only minor differences due to system calibration weighting function. Goodness of fit scores were analysed for differences in treatment using the Kruskal–Wallis test (Kruskal and Wallis, 1952) using data grouped by calibration method. This showed no significant difference in performance between calibration systems, although the weighted objective function score (Equation 2) will almost always be lower for system calibration relative to reach by reach since this is the object of the system optimisation process.
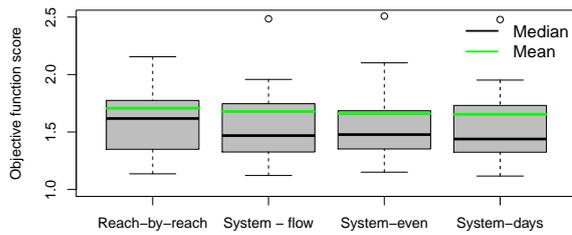
Figure 2. Goodness of fit calculated at each gauge by calibration method
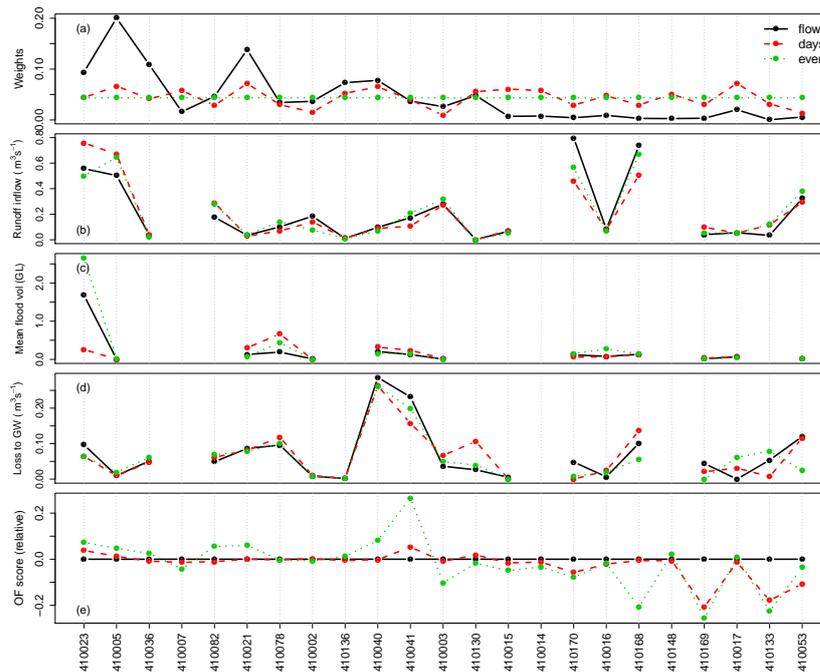
Figure 3. (a) Gauge weight, (b) estimated ungauged inflow, (c) estimated mean floodplain volume, (d) estimated loss to groundwater and (e) objective function score for each gauge in the study area. OF scores are given relative to the flow weighted OF score i.e., flow weighted OF score is subtracted

The nature of the study area is such that flows are highest in the east and become progressively smaller to the west. This partly due to losses (natural and anthropogenic), stream divergence and increasing aridity to the west. Flow weighting, therefore puts less emphasis on gauges in the west, to the advantage of the higher flow gauges to the east (Figure 3). The differences in weighting from the east (leftmost part of panel (a)), to west (right most part of (a)), are reflected in the objective function score (Figure 3, panel (e)). Goodness of fit is poorer (higher OF score) for flow weighting in the western and southern gauges. The converse is true from even weighting. It performs better in low flow gauges since it has a higher weighting.

Examination of model states and parameters shows that the system calibration method (regardless of weighting system) has far lower variance than reach by reach calibration (Figure 4). In the system calibration method any changes in system behavior at each gauge will propagate downstream and have consequences for those gauges downstream. For example, any improvement in goodness of fit for a particular gauge is likely to be rejected during system optimisation if these changes also incur commensurate degradation in downstream gauges. Hence over–fitting that may occur during reach–by–reach calibration can be avoided. This echoes

similar findings of the study of Hughes et al. (2015). Within system calibration, even weighting is slightly different to the other weighting schemes in that it produces higher floodplain volumes (and hence losses) for the reaches 410021, 410078, 410040 and 410041 (Figure 3). Without independent data to confirm the flood plain volumes for these reaches, it is difficult to say which is the more correct. Losses to groundwater are very similar in absolute terms, but also in the spatial distribution of the values. In the main east–west branch of the river system, loss to GW values are very similar for all weighting functions, however there is more variability in the values in the southern anabranch from gauges 410170 to 410133. Similar variability is apparent for ungauged runoff.
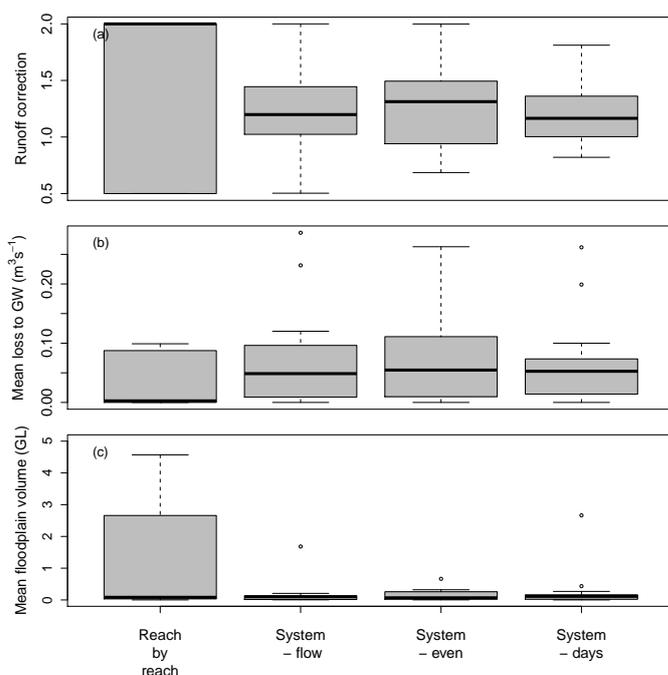
Figure 4. Model parameter/states for reaches across the model domain; (a) runoff correction factor, (b) mean loss to GW and (c) mean floodplain volume, for each calibration method

4 CONCLUSIONS

System calibration of the river system model had obvious advantages over reach by reach calibration, especially in terms of reduced over–fitting, noted by variance in model states and parameters. The weighting function used during system calibration had moderate effects on model performance for individual reaches within the model. Even weighting increased the distribution of weighting towards the lower (western and southern) parts of the catchment relative to flow weighting and therefore improved performance in those areas at the expense of higher flow area. The most notable finding is the stability of model states and performance despite differing weighting schemes. This indicates that the weighting scheme is less important than the mode of calibration i.e., reach–by–reach versus system.

REFERENCES

Andrews, F. T., B. F. Croke, and A. J. Jakeman (2011). An open software environment for hydrological assessment and development. *Environmental Modelling and Software 26(10)*, 1171–1185.

J. D. Hughes *et al.*, System calibration of river models...

Duan, Q. D., S. Sorooshian, and V. K. Gupta (1992). Effective and efficient global optimisation for conceptual rainfall-runoff models. *Water Resources Research 28*(4), 1015–1031.

Hughes, D. A., L. Andersson, J. Wilk, and H. H. Savenije (2006). Regional calibration of the Pitman model for the Okavango River. *Journal of Hydrology 331*(12), 30 – 42. Water Resources in Regional Development: The Okavango River.

Hughes, J., D. Dutta, J. Vaze, S. Kim, and G. Podger (2014). An automated multi-step calibration procedure for a river system model. *Environmental Modelling & Software 51*(0), 173 – 183.

Hughes, J. D., D. Dutta, S. S. H. Kim, J. Vaze, and G. M. Podger (2014). The effect of stream network simplification and gauge error on river model performance. *Australian Journal Of Water Resources 18*(2), 151–161.

Hughes, J. D., S. S. H. Kim, D. Dutta, and J. Vaze (2015). Optimisation of multiple gauge river-system models. a system approach. *Water Resources Research In review*.

Kruskal, W. and W. A. Wallis (1952). Use of ranks in one-criterion variance analysis. *Journol of the American Statistical Association 47*(260), 583–621.

Nash, J. E. and J. V. Sutcliffe (1970). River flow forecasting through conceptual models part I - a discussion of principles. *Journal of Hydrology 10*(3), 282–290.

Vaze, J., N. Viney, M. Stenson, A. V. Dijk, D. Dutta, R. Crosbie, J. Lerat, D. Penton, J. Vleeshouwer, J. Teng, S. Kim, J. Hughes, W. Dawes, Y. Zhang, B. Leighton, J.-M. Perraud, A. Yang, B. Wang, A. Frost, A. Elmahdi, A. Smith, and C. Daamen (2013, December). The Australian Water Resource Assessment Modelling system (AWRA). In *20th International Congress on Modelling and Simulation*, Adelaide. Modelling and Simulation Society of Australia and New Zealand.

Welsh, W. D., J. Vaze, D. Dutta, D. Rassam, J. M. Rahman, I. D. Jolly, P. Walbrink, G. M. Podger, M. Bethune, J. M. Hardy, J. Teng, and J. Lerat (2012). An integrated modelling framework for regulated river systems. *Environmental Modelling and Software*. 10.1016/j.envsoft.2012.02.022.