

An approach for developing and comparing empirical methods to model unaccounted losses in river system models

S.S.H. Kim^a, J. Lerat^a, J. Chen^a and J. Hughes^a

^a CSIRO Land and Water
Email: shaun.kim@csiro.au

Abstract: The use of loss functions in river system modelling assumes that river losses can be derived as a function of river flow. An approach has been developed that uses multiple objectives and cross-evaluation techniques to test different loss functions for improvement to river system simulation performances (Figure 1). This approach provides rapid analysis of the loss versus flow relationship and a framework for determining the most appropriate loss function. This method is particularly useful when incorporated with parsimonious modelling paradigms to increase parameter transferability and reduce over-predicting problems that are usually associated with models with large numbers of parameters (Box and Jenkins, 1970). The approach was trialled during the CSIRO Flinders and Gilbert Agricultural Resource Assessment (FGARA) project, where a river-reach model was developed that uses streamflow routing, rainfall-runoff and the Monod loss model for river-groundwater interactions. Twenty-one reaches in Queensland's Flinders regions were jointly calibrated and then re-simulated with the Monod loss model removed. These simulations were modified to include each of five alternate loss function models: (1) a 2-parameter isotonic piecewise-linear (IPWL) loss function; (2) a 5-parameter IPWL loss function; (3) a 10-parameter IPWL loss function; (4) a 10-parameter loss table derived with a quantile-based approach; and (5) a 100-parameter loss table derived with a quantile-based approach. Slight variations of loss function approaches were also tested in attempts to explore more options. The IPWL loss function is designed to have enough flexibility to adapt to the loss characteristics of specific river reaches while still being constrained to realistic behaviour. The results showed that the IPWL loss functions could perform well in both calibration and evaluation periods if there were initial over-estimations of flow and, hence, could potentially be used to pursue alternative parsimonious loss functions.

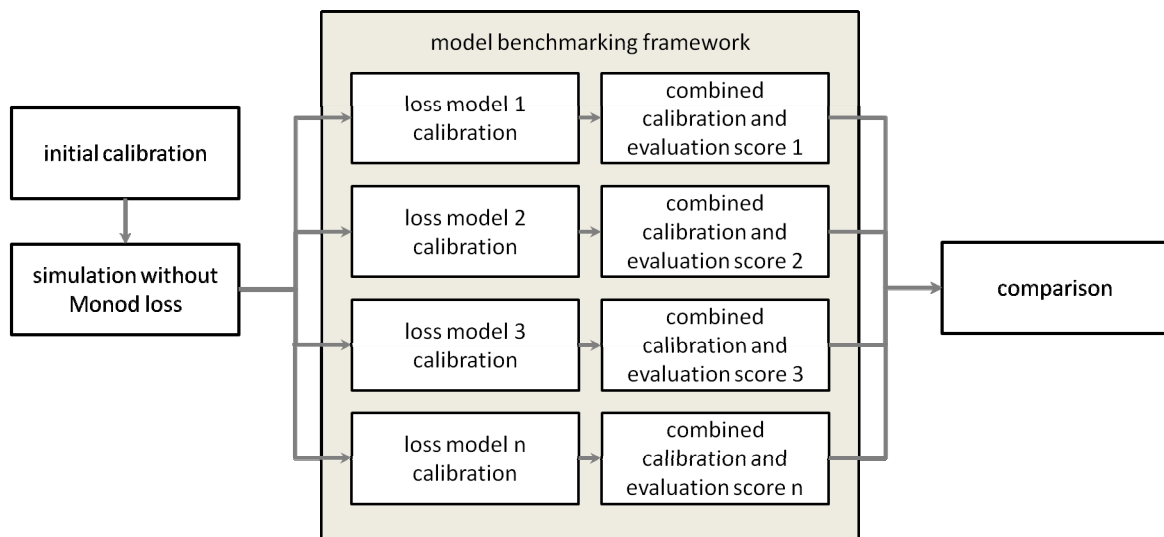


Figure 1. Conceptual flow schematic of the method used to model the different loss functions, including the model benchmarking framework.

Keywords: Flinders Catchment, loss function, model benchmarking, calibration, river system modelling

1. INTRODUCTION

Scarcity and increased pressure on water resources call for robust and accurate modelling tools that allow the stakeholders to explore and compare alternative management scenarios. In Australia, a large number of models (referred to as “river system models”) have been developed to address these issues, especially in the Murray-Darling Basin (Van Dijk *et al.*, 2008). These models describe the river network based on a semi-spatially distributed approach and include components dealing with hydrological processes (e.g. routing, rainfall-runoff) and components dealing with anthropogenic infrastructure (e.g. dams, water supply and irrigation diversions).

When the study area is large (e.g. $> 100\,000\text{ km}^2$), river models become complex structures that require several hundreds of nodes, multiple instances of rainfall-runoff and routing models, and a large number of diversion points. This paper aims to compare several river loss modelling approaches to observe the impact to the river system simulations.

In the semi-arid conditions that prevail in most parts of Australia, river systems exhibit large transmission losses due to floodplain interactions, evaporation and surface water/groundwater interactions. In this context, the rainfall-runoff (e.g. Sacramento, Burnash *et al.*, 1973; GR4J, Perrin *et al.*, 2003) and routing (e.g. Muskingum, Koussis, 1980; Laurenson, Linsley *et al.*, 1949) components are not sufficient to model the system accurately and modellers usually introduce a loss component (also referred as “unaccounted difference”). This loss component can be based on physical principles (Rassam, 2011), which may be preferred by river managers due to their perceived robustness (Lees, 2000). In addition, it is widely accepted that from a river management perspective, parsimonious modelling approaches will ensure maximised practical use of the models and the parameters (Grayson and Blöschl, 2001).

The parsimonious modelling paradigm involves reducing the number of parameters to simplify the system as much as possible while still maximising the system’s predictive performance. This strategy is fundamental to many modellers and is supposed to help avoid over-fitting problems (Grayson and Blöschl, 2001; Black *et al.*, 2011; Whittaker *et al.*, 2010). Simplification may also involve ‘lumping’ or reducing the spatial resolution of models which may have some benefit particularly in data-poor environments.

Further to the decision of the modelling option to take, there also needs to be a decision on which calibration method to use. Automatic calibration routines vary from simple Monte Carlo sampling (Metropolis, 1987) to the more sophisticated Shuffled Complex Evolution (Duan *et al.*, 1993). The objective function that is used for calibration should largely be driven by the intended purpose of the modelling exercise (Black *et al.*, 2011), and can be composed of multiple and opposing objectives. For example, mass balance bias would be the key consideration for water accounting purposes and/or Nash-Sutcliffe Efficiency for analysis of high river flows (Legates and McCabe, 1999).

The model that is used should be based on how fit-for-purpose the model is and what data is available (Black *et al.*, 2011). After a clear assessment is made on the objectives of the modelling exercise and data availability, the modelling approach, calibration procedure and objective function that will be suitable should become more apparent. However, even with clear plans of the modelling strategy there will always be other options that could be justifiably used. These could be subtle variations of essentially the same modelling procedure or could potentially be completely different conceptual approaches.

The CSIRO Flinders and Gilbert Agricultural Resource Assessment (FGARA) project has been undertaken to investigate the feasibility of increased agricultural development in the Gilbert and Flinders catchments. River system models were developed of the Flinders and the Gilbert regions to make assessments on water reliability for intensified irrigation, the environmental impacts on flows as a result of proposed reservoir constructions and reliability of reservoir delivery.

The river system models consist of reaches which are composed of streamflow routing, rainfall-runoff and loss model components. The loss component is thought to be a crucial element of the river system models with its ability to model unaccounted differences, reduce bias and fit flow duration curves (Hughes *et al.*, 2012; Panofsky & Brier, 1958). An automatic procedure is proposed that rapidly and systematically investigates similar river system loss models in an aim to improve river simulations without the need of multiple entire system calibration runs.

2. METHODOLOGY

For this study, a generic model benchmarking framework was developed that executes and compares different calibration routines and/or different simulation engines. The framework provides a platform to

automatically compare similar models and calibration approaches, with the ultimate aim of optimising the predictive performance of the modelling application. In this study, the framework was used to run pre-calibrated reach models to investigate the alternate loss functions.

Twenty-two Flinders reach models were initially jointly calibrated over the available observed data period between 1969 to 2011. Daily reach models were originally conceived with non-linear Laurenson steamflow routing (Laurenson, 1964), Sacramento rainfall-runoff (Burnash *et al.*, 1973) contributions and the Monod loss model (Monod, 1949; adopted by Kim *et al.*, 2012) described by the following equation:

$$L = \frac{M \cdot Q}{Q_{half} + Q} \quad (1)$$

where L is the Monod loss flux ($\text{m}^3 \cdot \text{s}^{-1}$), which is applied to the river flow, Q . M is the maximum river loss to groundwater and Q_{half} is the river flow value when the river loss is half the maximum value. M and Q_{half} were calibrated in this initial step along with the parameters for routing and rainfall-runoff. The next step, the reaches were simulated with routing and rainfall-runoff parameters all set to initially calibrated values and Monod loss model parameters to zero. This effectively removed the Monod loss component from the simulations. Each of the simulations were then coupled with alternate loss function approaches (Table 1) and run through the model benchmarking framework (Figure 1). The benchmarking framework was used to determine parameters of the alternate loss functions, with Monod loss parameters set to zero and all other parameters set to initially calibrated values.

Table 1. Summary of the alternate loss model methods.

Loss model	Parameters	Description
Model 1	2	Monod loss function
Model 2	2	Positive gradient linear function
Model 3	5	Isotonic piecewise linear function with evenly-spaced quantile indexes based on observed flow
Model 4	10	Isotonic piecewise linear function with evenly-spaced quantile indexes based on observed flow
Model 5	10	Loss table derived with a quantile-based approach
Model 6	100	Loss table derived with a quantile-based approach
Model 3.1	5	Isotonic piecewise linear function with evenly-spaced flow indexes between lowest and highest observed flow values
Model 4.1	10	Isotonic piecewise linear function with evenly-spaced flow indexes between lowest and highest observed flow values
Control	0	A control with no loss function used

Each reach loss model used corresponding downstream observed gauging station flow data to determine the parameters. For Model 1, 2, 3, 4, 3.1 and 4.1, a Sobol sampling method (Bratley and Fox, 1988) was adopted for sampling the loss function parameters during calibration. Parameters for Models 5 and 6 were determined using a quantile-based approach, which uses the difference between matching flow quantiles of observed and simulated flow duration curves to derive the loss versus flow relationship (Panofsky and Brier, 1958; adopted by Kim *et al.*, 2012). Extrapolation is based on the average slope of the last three points.

The framework runs a split calibration/evaluation scheme but requires input of wrapped calibration routines, models, objective functions and input data. The available observed flow data was split in half, where the first half was used for calibration and second half for evaluation. For calibrations, an objective was used that combines mass balance bias, low-flow and flow duration agreement. The objective function used during calibrations is calculated by:

$$\Omega = \left(0.1 \sum_{i=1}^{n_Q} \left[\sqrt{Q_{obs,i}} - \sqrt{Q_{sim,i}} \right]^2 + 0.9 \sum_{k=1}^{n_k} \left[\sqrt{Q_{obs,k}} - \sqrt{Q_{sim,k}} \right]^2 \right) (1 + |\beta|) \quad (2)$$

where n_Q is the total number of observed flow data points, Q_{sim} and Q_{obs} are the simulated and observed flow, respectively. $Q_{obs,k}$ and $Q_{sim,k}$ denote the k th point of the observed and simulated flow duration curves, respectively. Mass balance bias, β , is calculated by:

$$\beta = \frac{\sum_{i=1}^{n_Q} \mathcal{Q}_{sim,i} - \sum_{i=1}^{n_Q} \mathcal{Q}_{obs,i}}{\sum_{i=1}^{n_Q} \mathcal{Q}_{obs,i}} \quad (3)$$

For the assessment of the predictive performance of each loss model, a combined statistic that uses $1 - |\beta|$, Nash-Sutcliffe Efficiency (*NSE*, Nash and Sutcliffe, 1970) and *NSE* calculated on log transformed flows (*NSE_{lnQ}*, Oudin *et al.*, 2006) for both calibration and evaluation periods was used. The combined statistic for each period, *p*, weights all of these statistics equally, and is given by:

$$\omega_p = \frac{\sum_{i=1}^{n_{F,p}} F_{i,p}}{n_{F,p}} \quad (4)$$

where n_F is the number of functions used, *F* is the result of the function used. The final combined statistic is given by:

$$\kappa = \frac{\sum_{i=1}^{n_\omega} \omega_i}{n_\omega} \quad (5)$$

where n_ω is the number of periods used.

3. RESULTS AND DISCUSSION

Individual and combined statistics were calculated for each reach and compiled to compare the suitability of each of the loss models. The control (where no loss function was used) was a crucial benchmarking check since it allows demonstration of over-predicting models. During the calibration and evaluation period, Model 1, 2, 6, 3.1 and 4.1 performed marginally better than the control for combined statistics (Figure 2). The control showed comparable performance to Model 1, 2, 6, 3.1 and 4.1, where Models 3, 4, and 5 could be deemed unsuitable loss modelling approaches for this region.

The only difference between Models 3 and 4 to Models 3.1 and 4.1 are the flow indexes sampled. As suspected, the flow indexes sampled has a significant impact to the shape of the loss curve and therefore the performance of the loss function. The evenly-spaced quantile flow indexes was thought to perform better than evenly-spaced flow indexes since the quantile flows were thought to have better coverage of frequent flow events, however these over-emphasised the importance low flow events and ignored high flow events which accounted to large portions of the water balance, and contribute to the performance of *NSE*.

Some degradation of performance can be seen from the calibration to evaluation period. This could be due to uncertainty surrounding the parameters, inputs or modelling concepts. For example, perhaps the poor evaluation performances could be caused by an overly complex reach model that is over-fitting to the calibration period. Interestingly, over-prediction in Model 6 compared to the other loss models did not occur despite it having the largest number of parameters. This is thought to be the due to the inherent reduced sensitivity of each parameter as the number of parameters increases. That is, as the number of loss indexes increase, the flow range at which they apply will shorten, thus reducing the impact of each parameter.

Examination of the summary statistics (Table 2) highlights the individual statistics for each model. Bias and *NSE_{lnQ}* were the reason Models 3, 4 and 5 performed poorly overall. Mean absolute calibration biases were 16%, 26% and 14% for Models 3, 4 and 5, respectively, where the control was just 3%. Similarly, mean calibration *NSE_{lnQ}* results for Models 3 and 4 were 26% and 49% lower than the control result, respectively, indicating difficulty for these models to simulate low flows.

Plots of the loss flow curves for all tested models for the Flinders reaches were used to detect any discernible characteristics of reach losses (Figure 3). The plotted lines display each loss model's relationship between river loss and river flow. Models 5 and 6 were not constrained to be monotonically increasing, and therefore, display some high flow unaccounted gains. This is characteristic of the rainfall runoff model's inability to match observed high flows (Kim *et al.*, 2012) or possibly be due to some floodplain interactions during high flow events. Models 3 and 4 have more changes in curve shape in the low flow ranges opposed to Model 3.1

and 4.1, as expected. This implies that that the Models 3.1 and 4.1 have more accurate loss indexes in the high flows. On each plot, there are locations where there are multiple intersections of loss curves. These may be useful in refining loss ranges to further optimise the loss models.

Table 2. Summary results of models' performance for all reaches in the Flinders.

model		1	2	3	4	5	6	3.1	4.1	control
Calibration	Mean NSE	0.66	0.67	0.66	0.65	0.63	0.66	0.68	0.68	0.66
	Mean NSE_{inO}	0.55	0.56	0.39	0.27	0.56	0.57	0.56	0.55	0.53
	Mean $1 - \beta $	0.99	0.97	0.84	0.74	0.86	0.97	0.97	0.96	0.97
	ω_{cal}	0.74	0.73	0.63	0.56	0.69	0.73	0.73	0.73	0.72
Evaluation	Mean NSE	0.47	0.48	0.49	0.49	0.45	0.47	0.51	0.52	0.47
	Mean NSE_{inO}	0.45	0.45	0.32	0.23	0.47	0.47	0.45	0.44	0.43
	Mean $1 - \beta $	0.73	0.73	0.71	0.67	0.69	0.73	0.73	0.73	0.74
	ω_{val}	0.55	0.55	0.51	0.46	0.54	0.56	0.56	0.56	0.55
Both Periods	Mean NSE	0.57	0.57	0.57	0.57	0.54	0.57	0.59	0.6	0.57
	Mean NSE_{inO}	0.5	0.5	0.36	0.25	0.52	0.52	0.5	0.5	0.48
	Mean $1 - \beta $	0.86	0.85	0.78	0.71	0.78	0.85	0.85	0.85	0.85
	κ	0.64	0.64	0.57	0.51	0.61	0.64	0.65	0.65	0.63

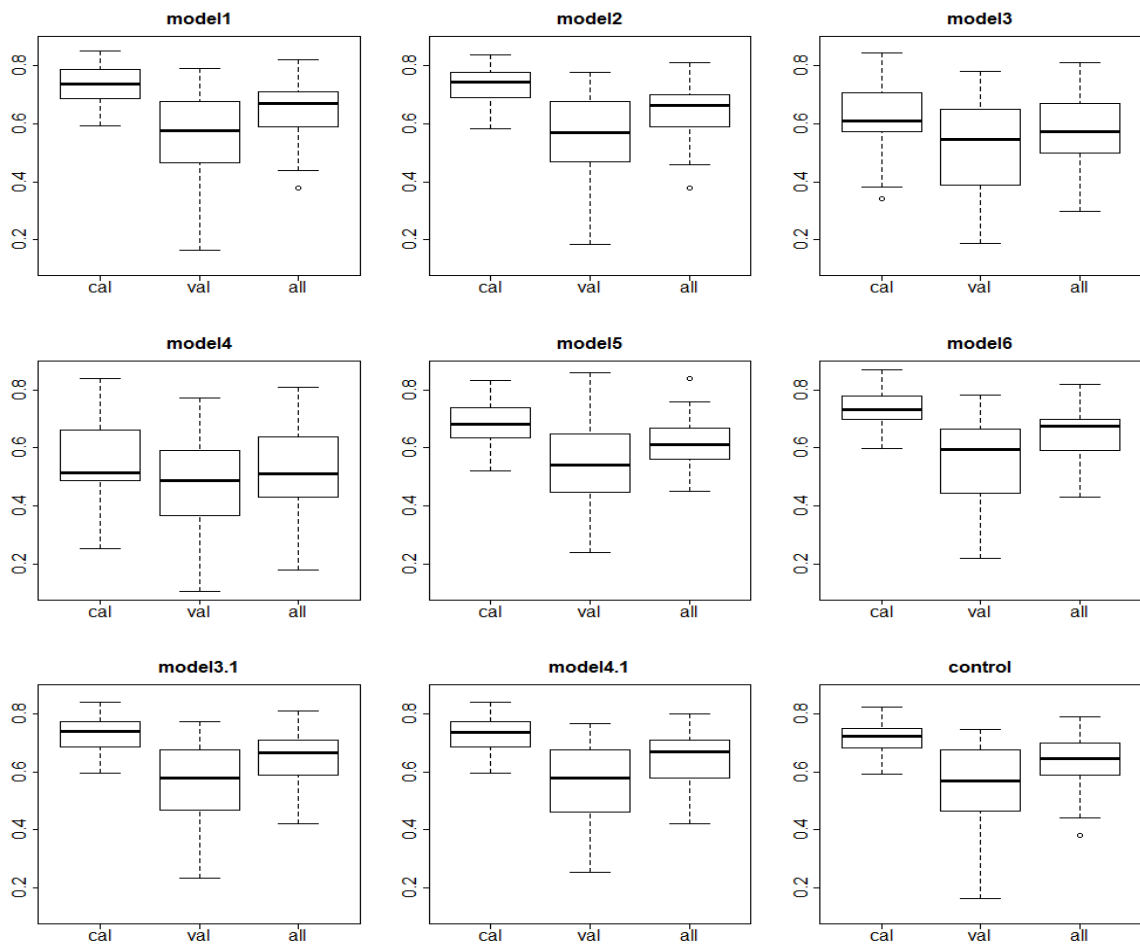


Figure 2. Combined statistic performances for all the reaches for the calibration (cal), evaluation (val) and the whole (all) period. Note: Whiskers of the boxplots represent the lowest datum still within 1.5*interquartile range (IQR) of the lower quartile, and the highest datum still within 1.5*IQR of the upper quartile.

An examination of flow duration curve plots showed characteristics that dictated the success of the method (Figure 4). If control simulations over-estimated the flows, then there was capability for the loss functions to

improve results. However, if flows were under-estimated by the control, the loss functions (with the exception of Models 5 and 6) were unable to deal with the unaccounted gains. When examining only the reaches that had over-estimating controls, average results for all models for NSE_{lnQ} over both periods gives 0.54 with a control score of 0.51. These results are much better than reaches with under-estimating controls, which had an average NSE_{lnQ} of 0.42 with a control score of 0.47.

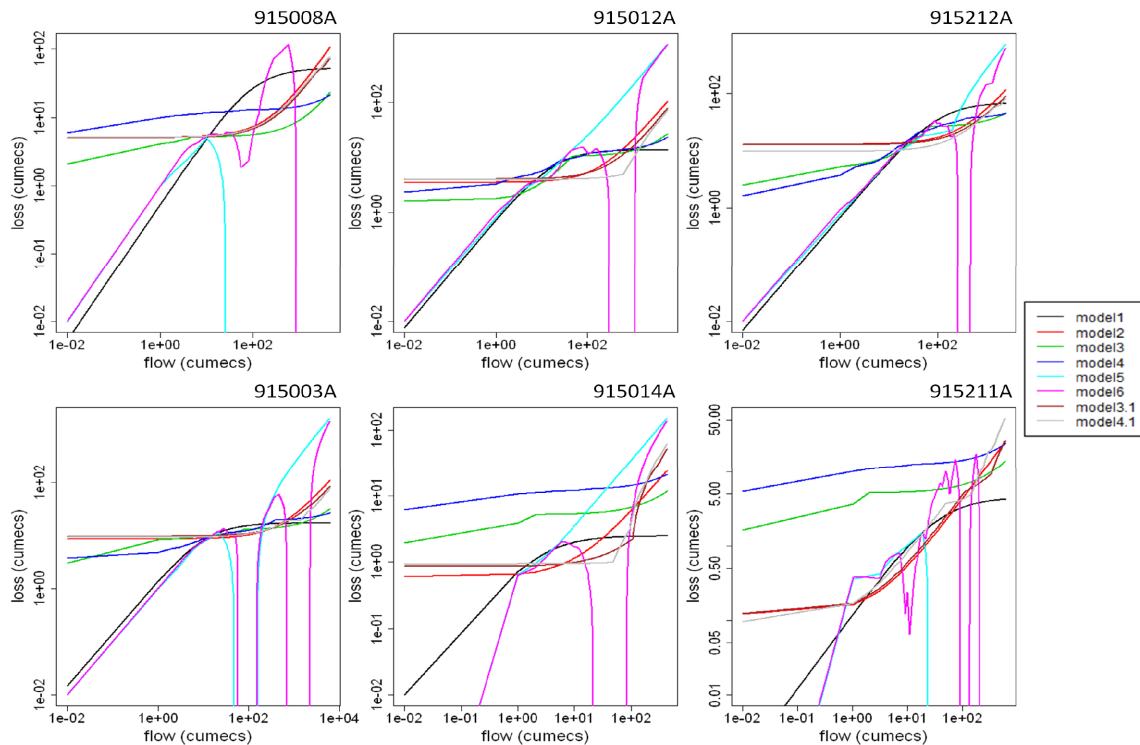


Figure 3. Loss function log-log plots for each of the models for Flinders gauge locations 915008A, 915012A, 915212A, 915003A, 915014A and 915211A.

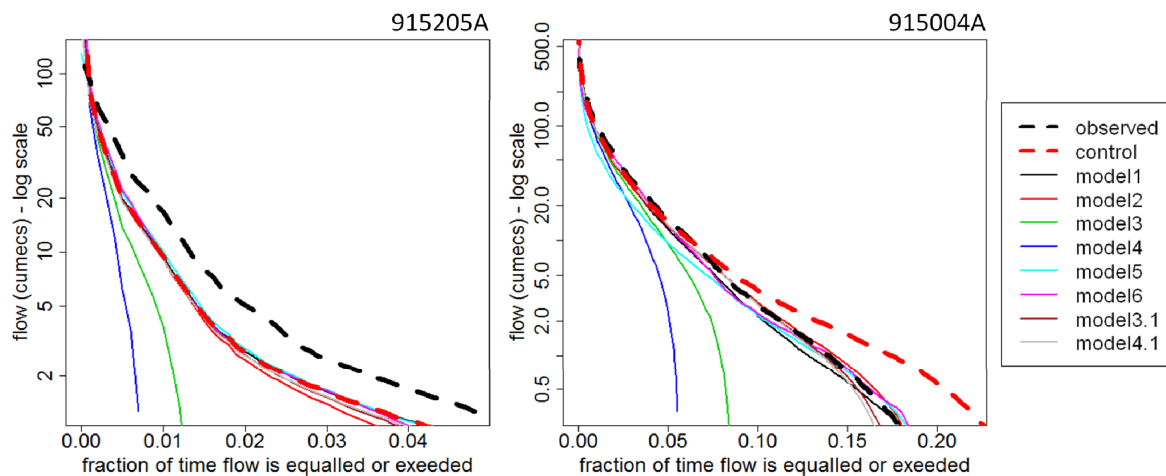


Figure 4. Flow duration curve plots demonstrating under-estimation (915205A) and over-estimation (915004A) of flow in the control.

The approach taken in this study has proven to be a simple and rapid way of investigating different loss functions. The approach's success is obviously dependent on the initial calibration and whether the Monod model is able to represent river losses. It is expected that while using this approach the Monod function would be one of the best performing loss functions, since the reaches were initially optimised using Monod, however, other loss functions have shown to perform just as well, and without following the same curve shape, indicating there is potential for more exploration. The next step may be to use the model benchmarking framework on the most successful loss models using a full reach calibration method.

4. CONCLUSION

The method used in this study has proven to be a rapid and simple way of benchmarking river system loss models and provides insights to reach-specific loss characteristics. The method's success was largely dependent on the initial calibration, where improvement could be obtained if the initial calibration over-estimated flows. The model benchmarking framework could be used to incorporate whole reach model calibrations instead of just the loss components. IPWL loss functions with evenly-spaced quantile indexes and the 10-parameter loss table derived with a quantile-based approach performed poorly compared to using no loss model, however, all other tested models (Monod function, positive gradient linear function, IPWL loss functions with evenly-spaced flow indexes, 100-parameter loss table derived with a quantile-based approach) showed marginal improvement in agreement for low-flow, high flow and bias statistics.

REFERENCES

- Black, D., Wallbrink, P., Jordan, P., Waters, D., Carroll, C., Blackmore, J. (2011). *Guidelines for water management modelling: Towards best-practice model application*. eWater CRC.
- Box, G.E.P., Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-D. iv, San Francisco.
- Bratley, P., Fox, B.L. (1988). Algorithm 659: Implementing Sobol's Quasirandom Sequence Generator. *ACM Transactions on Mathematical Software*, 14, 88-100.
- Burnash, R.J.C., Ferral, R.L., McGuire, R.A. (1973). *A generalised stream-flow simulation system – conceptual modelling for digital computers*. Technical report, Joint Federal and State River Forecast Center, Sacramento.
- Duan, Q.Y., Gupta, V.K., Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of optimization theory and applications*, 76(3), 501-521.
- Grayson, R., Blöschl, G. (2001). *Spatial patterns in catchment hydrology: observations and modelling*. Cambridge University Press, Cambridge.
- Hughes, J., Dutta D., Kim S., Vaze J., Podger G. (2012). *An automated calibration strategy for a river system model*. Proceedings of the National Conference on Water and Climate: Policy Implementation Challenges, Engineers Australia, 1-3 May 2012, Canberra.
- Kim, S.S.H., Vaze, J., Dutta, D., Hughes, J., Dawes, W. (2012). Investigation of different loss and gain accounting methodologies for river system modelling. Hydrology and Water Resources Symposium, Engineers Australia, 19-22 November 2012, Sydney.
- Koussis, A.D. (1980). Comparison of Muskingum method difference schemes. *Journal of Hydraulic Division*, ASCE, 106(5), 925-929.
- Laurenson, E.M. (1964). A catchment storage model for runoff routing. *Journal of Hydrology*, 2(2), 141-163.
- Lees, M.J. (2000). Data-based mechanistic modelling and forecasting of hydrological systems. *Journal of Hydroinformatics*, O2.1, 15-34.
- Legates, D.R., McCabe, G.J. (1999). Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1), 233-241.
- Linsley, R.K., Kohler, M.A., Paulhus, J.L.H. (1949). *Applied Hydrology*. McGraw-Hill, New York, 689 pp.
- Metropolis, N. (1987). The beginning of the Monte Carlo method. *Los Alamos Science*, 15(584), 125-130.
- Monod, J. (1949). The growth of bacterial cultures. *Annual Review of Microbiology*, 3, 371-394.
- Nash, J.E., Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I – A discussion of principles. *Journal of Hydrology*, 10(3), 282-290.
- Oudin, L., Andréassian, Mathevet, T., V., Perrin, C. (2006). Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. *Water Resources Research*, 42, W07410.
- Panofsky, H.A., Brier, G.W. (1958). *Some Applications of Statistics to Meteorology*, The Pennsylvania State University, 224 pp.
- Perrin, C., Michel, C. and V. Andreassian, (2003). Improvement of a parsimonious model for streamflow simulations. *Journal of Hydrology*, 279, 275-289.
- Rassam, D.W. (2011). A conceptual framework for incorporating surface-groundwater interactions into a river operation-planning model. *Environmental Modelling and Software*, 26, 1554-1567.
- Van Dijk, A.I.J.M., Kirby, M., Paydar, Z., Podger, G., Mainuddin, M., Marvanek, S. Peña Arancibia, J. (2008). *Uncertainty in river modelling across the Murray-Darling Basin*, A report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project, CSIRO Australia, 89 pp.
- Whittaker, G., Confesor, Jr., R., Di Luzio, M., Arnold, J.G. (2010). Detection of overparameterization and overfitting in an automatic calibration of SWAT. *Transactions of the ASABE*, 53(5), 1487-1499.