

Reducing propagation of uncertainty in river system modelling by optimal use of streamflow data

J. Lerat^a, D. Dutta^a, S. Kim^a, J. Hughes^a, J. Vaze^a

^a CSIRO Land and Water, Canberra
Email: julien.lerat@csiro.au

Abstract: Modelling of regulated river systems requires a combination of hydrological components, e.g. rainfall-runoff models, and components dealing with diversion and storage of water. The resulting models often include a large number of calibrated parameters and multiple prediction points. To reduce the dimension of the calibration problem, the modelling domain can be split into river reaches delimited by inflow and outflow gauging stations. When the model is used in predictive mode, e.g. climate change studies, uncertainty generated within the upstream reaches will propagate to downstream reaches and alter the predictions on the lower parts of the model. This study investigates the use of streamflow data to better constrain the calibration of river models. The study is based on the calibration of a river model for a set of six Australian catchments, as part of the Australian Water Resources Assessment (AWRA) framework developed jointly by CSIRO and the Bureau of Meteorology. The river model was calibrated sequentially from upstream to downstream reaches using three different scenarios. In the first scenario, the reaches are calibrated independently using gauged data as inflows to the reach. This option isolates the reaches and minimizes the propagation of uncertainty. However, it leads to sub-optimal parameters when the model is used in predictive mode with no flow data available. In the second scenario, the reach inflows were set to the simulations from upstream reaches. In the third scenario, the inflows are set to a weighted average between gauged inflow and upstream simulations. The model performance was finally assessed in predictive mode. The results from this study suggest that the use of observed inflows leads to a systematic improvement of the performance statistics compared to the use of other inflow data.

Keywords: *River modelling, Streamflow data, model calibration, multisite calibration, GR4J*

1. INTRODUCTION

A river model combines rainfall and routing components that simulate natural hydrological processes with components representing the management of water diversion infrastructures (e.g. dams, canal intakes, pumps, etc.). River models form a special class of spatially explicit hydrological models developed for water resources planning and operational management e.g., to assess the impact of water sharing rules amongst multiple users including the environment or to conduct water balance estimates of the fluxes and stores along the river system. The state and federal water agencies across Australia responsible for management of water resources within their jurisdictions use different models that are appropriate for their modeling needs (IQQM, see Simons *et al.*, 1996; Vaze *et al.*, 2011, REALM, see Diment, 1991, MSM and BigMod, see MDBC, 2002). The model presented in this paper is the river component of the Australian Water Resources Assessment system (Vaze *et al.*, 2013) developed jointly by CSIRO and the Bureau of Meteorology as part of the Water Information Research and Development Alliance (WIRADA). The AWRA system is used by the Bureau for the production of the National Water Accounts (NWA) and the Australian Water Resources Assessment (AWRA) reports, which provides an overview of water fluxes and storages at the national scale. Both products implement the Water Balance Framework introduced by Barratt (2008).

Figure 1 shows a typical configuration for the river models investigated in this paper. The configuration includes multiple prediction points located at gauged sites along a river system. When the model is run from the headwaters (referred as run order 1 in Figure 1) to downstream residual reaches (referred as run order 2 and 3 in Figure 1), there are principally two ways of dealing with streamflow data:

- The most straightforward approach is to run the model without using streamflow data (e.g. the predicted flow P1 is equal to the upstream boundary conditions for the residual reach 2 in Figure 1). This approach is referred as “simulated inflows” in this paper.
- A less common approach is to reset the flow at each gauging station with the observed flow data (e.g. the observed flow O1 is passed to the residual reach 2). This approach is referred as “observed inflows” in this paper.

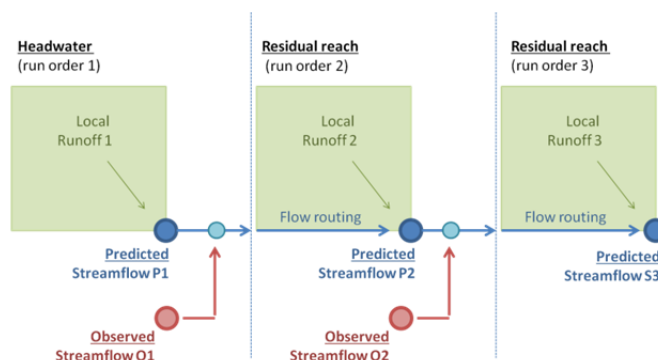


Figure 1. Combining streamflow data and prediction in a multisite river model

Most spatially explicit hydrological models presented in the scientific literature (e.g. Ajami *et al.*, 2004) are run with simulated inflows for a simple reason: the use of observed inflows requires streamflow data, which are obviously no longer available when the model is run in predictive mode (e.g. prediction in ungauged basins, streamflow forecasting or climate change scenario). As a result, using observed inflows would probably not even been considered by most hydrologists. Two exceptions can be mentioned. First, the water accounting exercise undertaken by the Bureau of Meteorology provides retrospective water balance estimates over the past year. For this specific exercise, streamflow data are available and can be used as observed inflows in the model. The development of a model in this context was the main driver for considering observed inflows as a viable alternative (Lerat *et al.*, 2011). Second, the calibration of a river model is always performed against historical streamflow records where long time series of flow data are available. In this case, even if the model is designed to *run* with simulated inflows, it can be *calibrated* with observed inflows, provided concurrent streamflow records exist at both ends of the reach.

As a result, the main question investigated in this paper is the following: given a set of inflows chosen for a river model (predictive mode using simulated inflows or retrospective mode using observed inflows), which type of inflows should be used during the calibration phase? More specifically, in the case where the model is to be used in predictive mode (i.e. with simulated inflows), two contradicting arguments could be put forward: on one hand, it seems preferable to calibrate the model with simulated inflows in order to have a similar model set-up between the calibration and the run phase. On the other hand, simulated inflows introduce errors at each prediction point that can propagate down the river system and corrupt the calibration of parameters on downstream reaches. The question is investigated here by comparing the performance of river models calibrated with observed, simulated and combined inflows. Section 2.1 presents the AWRA-R river model that is used in this study. The application of the model to six catchments in Australia is presented

in Section 0 and model calibration using different approaches is described in Section 2.3. The results are presented in Section 3 and discussed and summarized in Section 4.

2. METHOD

2.1. Model

For a given residual reach, the AWRA-R river model was run at the daily time step based on the following equation (Figure 2):

$$\widehat{Q}_{d/s} = (Q_{u/s})_{\text{rout}} + Q_r + Q_s - Q_d + Q_p - Q_e - Q_a - Q_{fp} + Q_{fpr} - Q_{gw}$$

Where, $\widehat{Q}_{d/s}$ is the estimated flow at the downstream gauge, $Q_{u/s}$ is concurrent flow at the upstream gauges (including gauged tributaries) following routing, Q_r is the local runoff contribution, Q_s is the contribution from any artificial storages including rainfall on storage area, evaporation from storage area and change in storage volume, Q_d is the loss due to irrigation diversion, Q_p and Q_e are the fluxes to and from the river due to rainfall and evaporation, respectively, Q_a is the flow diverted to anabranches, Q_{fp} is overbank flow to floodplain, Q_{fpr} is return flow from floodplain, Q_{gw} is the flux from river to groundwater.

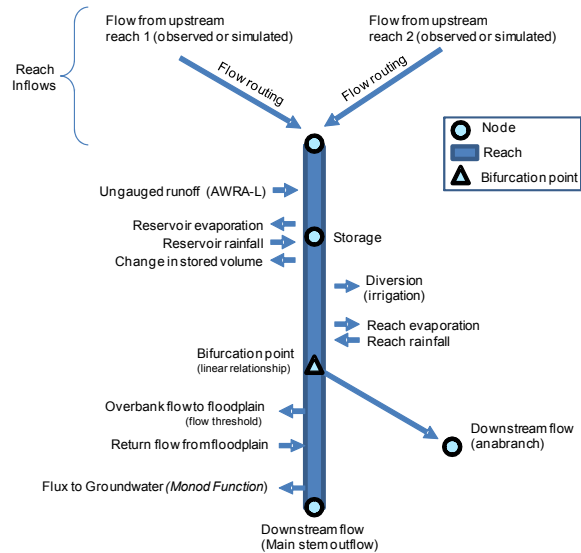


Figure 2. Conceptual representation of a river reach within AWRA-R

The routed flow $(Q_{u/s})_{\text{rout}}$ was computed by applying a delay on the reach inflows followed by a Muskingum routing scheme. The runoff Q_r was extracted from the gridded runs of the AWRA-L landscape model (version 1.5, Vaze *et al.*, 2013). A linear scaling factor was applied to the raw runoff data in order to correct for local errors in the runoff grids (Lerat *et al.*, 2012). As an alternative to AWRA-L, the GR4J rainfall-runoff model (Perrin *et al.*, 2003) was used to generate the runoff inputs. The anabranch flow Q_a was computed with a fixed linear partition between the branches. The floodplain model is detailed by Dutta *et al.* (2013). Finally, the groundwater loss was computed from the river flow with the Monod equation (Monod, 1949) $Q_{gw}=a Q/(b+Q)$ with Q the river flow and a and b the two Monod parameters.

When the model is applied to a residual reach using AWRA-L runoff, it has 25 calibrated parameters: three routing parameters, one runoff scaling factor, two floodplain parameters (see Dutta *et al.*, 2013), two Monod parameters and 17 parameters for the AWRA-L model. The model applied on a residual reach with GR4J runoff has eleven parameters (including the four free parameters of GR4J).

2.2. Data

The AWRA-R river model is currently deployed across Australia to support the production of the NWA by the Bureau of Meteorology. In this paper, five catchments presented in Table 1 were selected to cover a broad range of hydro-climatic conditions encountered across Australia. Daily timeseries of climate and streamflow data were provided by the Bureau of Meteorology.

2.3. Calibration approach

All the calibration experiments investigated in this paper were performed using a grid sampling followed by a Nelder-Mead gradient descent algorithm (Nelder-Mead, 1965). The objective function used for the calibration was a combination of the mean square error applied to the root square transform of flow with the simulation bias as per the following equation (Coron *et al.*, 2012):

$$F = \sum_i \left[\sqrt{\widehat{Q}_{d/s}(i)} - \sqrt{Q_{d/s}(i)} \right]^2 \left(1 + \left| \frac{\sum_i \widehat{Q}_{d/s}(i) - \sum_i Q_{d/s}(i)}{\sum_i Q_{d/s}(i)} \right| \right)$$

where, $\widehat{Q}_{d/s}(i)$ and $Q_{d/s}(i)$ are the observed and simulated flows on day i , respectively.

Table 1. Characteristics of test catchments

Catchment	Number of reaches	Total area (km ²)	Mean annual rainfall (mm/y)	Mean annual PE (mm/y)	Calibration period	Validation period
Namoi	28	39,800	633	610	70-92	93-2011
Ovens	17	7,800	1000	770	70-92	93-2011
Condamine-Balonne	33	136,600	510	500	70-92	93-2011
Lachlan	38	85,500	460	440	70-92	93-2011
Flinders	22	110,000	490	1860	70-92	80-90
Gilbert	23	46,400	770	1870	70-92	80-90

A split sample calibration was performed with calibration and validation periods shown in Table 1. Finally, the model calibration was repeated three times with (1) observed inflows, (2) simulated inflows and (3) an average of observed and simulated inflows. These three calibration experiments were performed with AWRAL runoff and GR4J runoff, leading to a total of six calibrated parameter sets for each reach.

3. RESULTS

Model performance was assessed by computing the daily Nash-Sutcliffe efficiency (NSE), the daily NSE on log transform flow (NSELOG), the absolute bias and the daily variance ratio between the simulated and observed flow data. The last two statistics were selected based on the decomposition advocated by Gupta and Kling (2009).

The distribution of the performance metrics computed over the validation period for the six calibration experiments are shown in Figure 3. The figure shows the results when the model is run in retrospective mode (with observed inflows, see plots 3.a to 3.d) and in predictive mode (with simulated inflows, see plots 3.e to 3.h). The main messages from the results in the two figures can be summarized as:

- When the model is run with observed inflows, Figure 3.a to 3.d suggests that a model calibrated with observed inflows performs better than one calibrated with simulated inflows for most statistics. For example, in Figure 3.a, the median absolute bias when the model is calibrated with observed inflows is 0.18 against 0.21 when the model is calibrated with simulated inflows. Similarly, the median NSE shown in Figure 3.b is 0.72 with calibration using observed inflows against 0.64 for simulated inflows.
- The variance ratio is the only statistic that does not follow this trend as can be seen in Figure 3.d. For example, the median ratio is -0.18 for a model calibrated with observed inflows (i.e. an underestimation of the observed variance by 18%) against -0.13 with simulated inflows. The variance of a highly skewed variable like streamflow is strongly influenced by the extremes. Consequently the result shown in Figure 3.d would suggest that high flows are better represented when the model is calibrated with simulated inflows.
- When the model is run with simulated inflows, the performance statistics worsen significantly compared to the ones obtained with observed inflows as can be seen by comparing Figure 3.a to 3.d with Figure 3.e to 3.h. For example, the median absolute bias a models calibrated with observed inflows increases from 0.18 in Figure 3.a to 0.26 in Figure 3.e. This result was expected because the use of observed inflows eliminates the simulation errors that propagate from upstream to downstream reaches. It suggests that the errors introduced by observed streamflow data are of lower order of magnitude than the ones generated by the model.
- Aside from the general lower level of performance shown in Figure 3.e to 3.h by comparison to Figure 3.a to 3.d, the general trend between the calibration experiments remains similar. For example, in Figure 3.f, the models calibrated with simulated inflows exhibits a median NSE of 0.39 against 0.5 with observed inflows. In other words, when the model was run with simulated inflows, a calibration using observed inflows performed better than one using simulated inflows. This result holds for bias (Figure 3.e), high flows (Figure 3.f) and low flows (NSELOG shown in Figure 3.g). The very high flow regimes characterised by the variance ratio in Figure 3.h is again the exception with model calibration using simulated inflows performing better than ones using observed inflows.

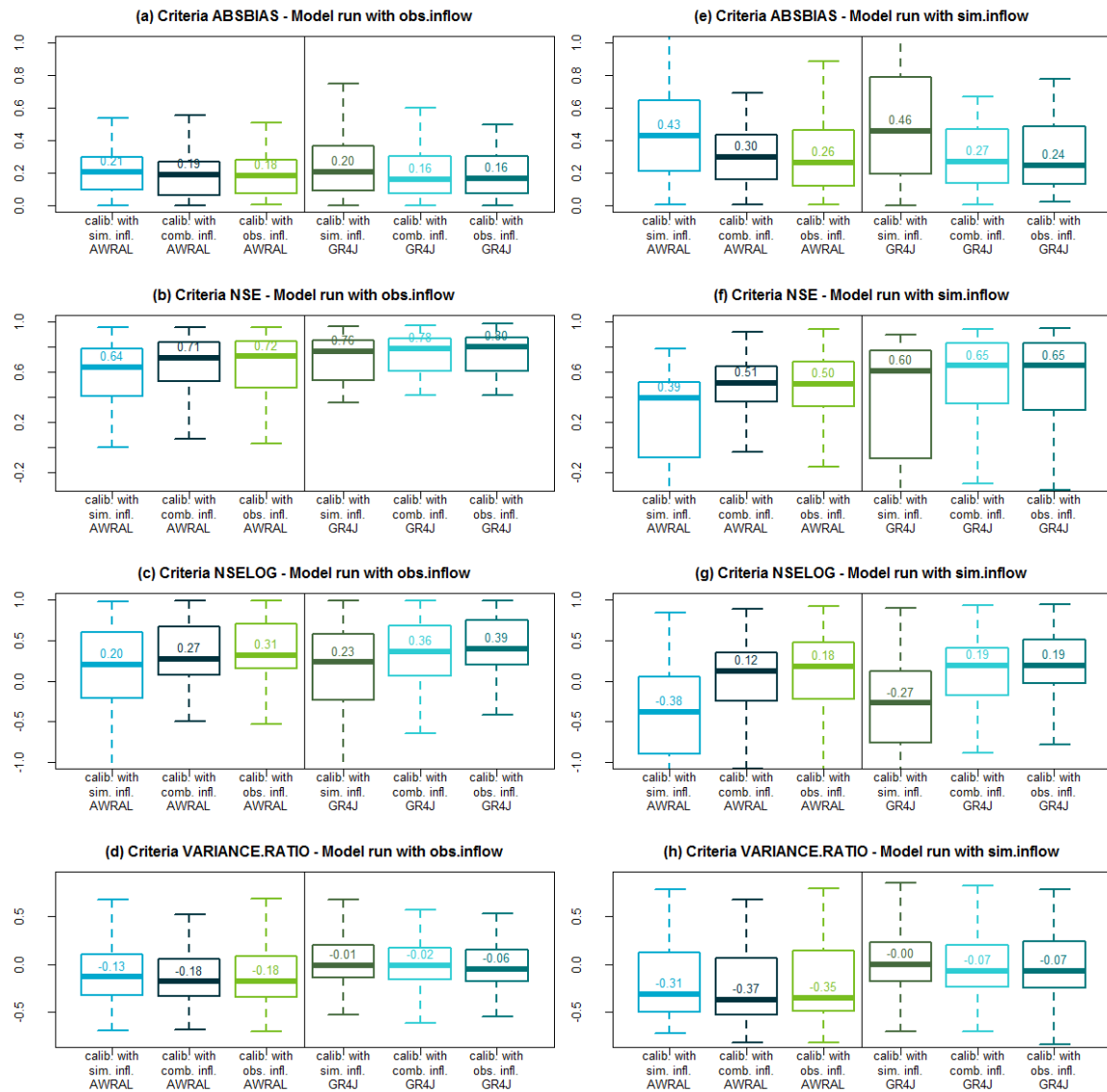


Figure 3. Performance statistics over the validation period on residual reaches in retrospective mode (observed inflows, plots (a) to (e)) and predictive mode (observed inflows, plots (e) to (h)) for the six calibration experiments. Each plot shows the distribution of model performance calibrated with simulated inflows, combined inflows and observed inflows, and two rainfall-runoff models (AWRA-L and GR4J) to compute ungauged lateral inflows to the reach.

- The use of GR4J rainfall-runoff model produces systematically better performance statistics than AWRA-L for all statistics presented in Figure 3. This result was expected because GR4J has more degrees of freedom (4 free parameters against one scaling factor for AWRA-L) to adjust to local runoff conditions. For example, the median NSE shown in Figure 3.f is 0.5 when AWRA-L runoff is used against 0.65 with GR4J runoff. It is important to indicate that the version of AWRA-L used for this work (AWRA-L 1.5) has now been superseded by recent developments which are expected to bring the performance level closer to GR4J. However, in spite of these differences, the general trends described in the previous points remains valid for a river model using GR4J: the use of observed inflows during calibration always produce better performance in validation mode, even when the model is run with simulated inflows.
- The use of combined observed and simulated inflows produced performance statistics lying between the statistics obtained with configuration using observed and simulated inflows, respectively.

Figure 3 presents an aggregated view of the results over the whole dataset. In Figure 4, the distribution of the performance statistics was split according to the run order of the reaches (e.g. headwater=1, immediately

downstream of headwater=2, etc...). The model was run over the validation period in predictive mode with the GR4J rainfall-runoff model. Figure 4 can be commented as follows:

- The model performance generally improves when moving away from headwaters up to run order 10. For example, the median NSE shown in Figure 4.b goes from 0.64 for run order 2, to 0.70 for run order 3-4 and 0.79 for run order 4-9. However, all performance suddenly drop for run orders greater than 10, which suggests that the model can be improved in lower part of the system.
- The trend observed in Figure 3 is confirmed in Figure 4: models calibrated with observed inflows performed better than models calibrated with simulated inflows. The trend is particularly clear for the low flow simulations, as can be seen by comparing Figure 4.c with Figure 4.f.

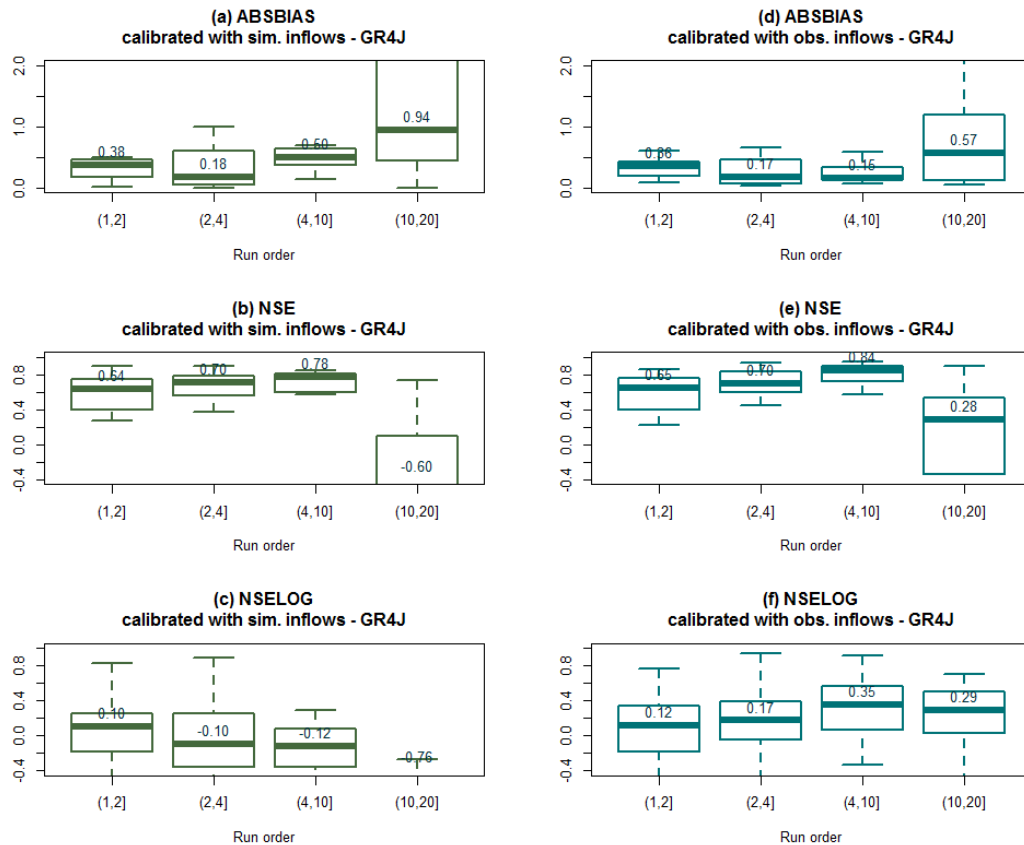


Figure 4: Distribution of performance statistics from most upstream (reach order 2) to most downstream residual reaches (reach order 10 to 20). Statistics are computed over the validation period in predictive mode (simulated inflows)

4. DISCUSSION AND CONCLUSION

This paper clearly shows that the choice between a configuration using simulated inflows (propagation of flow predictions from one reach to another) and observed inflows (reset of inflows to observed flow data at each gauge) has a significant impact on the calibration of river model. The use of observed inflows during calibration produced superior performance, including when the model is run with simulated inflows over an independent validation period. These results suggest that the propagation of the simulation errors from one reach to another deteriorates the calibration of parameters for downstream areas of the model.

The advantage provided by the use of observed inflows during calibration was confirmed when a different rainfall-runoff model was used and when performance was analysed at different locations of the river system. The performance gains were particularly significant for low flow simulations at the lower end of the river systems.

A combination between observed and simulated inflows was tested but did not bring any performance improvement compared to the use of simulated inflows. However, the combination tested was based on a simple average between the observed and simulated inflows. More advanced combinations could be

considered with, for example, time varying weighting factors between the two variables. Eventually, the weighting scheme could be based on statistical modelling of the streamflow data (e.g. inverse weighting with data variance).

In all the calibration methods presented in this paper, the river model was calibrated sequentially reach by reach from headwater to downstream reaches. Future work will explore alternative approach performing a simultaneous calibration on all prediction points (system calibration). As suggested by Micevski *et al.* (2011), this is expected to further constrain the model simulation and may help better identify the trade-off between spatial consistency, which is limited in the current approach, and model performance at a given prediction point.

ACKNOWLEDGMENTS

This work was funded by the WIRADA research alliance between the Bureau of Meteorology and the CSIRO and carried out in the CSIRO Water for a Healthy Country Flagship. The climate and streamflow data were provided by the Bureau of Meteorology.

REFERENCES

- Ajami, N. K., Gupta H., Wagener T., Sorooshian S. (2004). Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *Journal of Hydrology*, 298.1: 112-135.
- Barratt D.G. (2008) A Proposed National Water Balance Framework. *Water Division, Bureau of Meteorology*.
- Coron L., Andréassian V., Perrin P., Lerat J., Vaze J., Bourqui M. and Hendrickx F. (2012) Crash testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments. *Water Resources Research*, 48, W05552.
- Diment, G.A. (1991). Wide use of a generalised headworks and resources model: REALM. Proceedings International Hydrology and Water Resources Symposium, Perth, Australia, Institution of Engineers, Australia, National Conference Publication, 91(22), 579–583.
- Dutta, D., J. Teng, J. Vaze, J. Hughes, J. Lerat, S. Marvanek (2013). Building Flood Inundation Modelling Capability in River System Models for Water Resources Planning and Accounting, *In: Climate and land surface changes in hydrology, IAHS Red Book (ed by Boegh et al.), IAHS Publication*, 205-212.
- Gupta, H. V., Kling, H. (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377.1, 80-91.
- Lerat, J., Z. Paydar, Peña-Arancibia J., Dawes W. (2012). Development and evaluation of the AWRA-R model, *CSIRO: Water for a Healthy Country National Research Flagship*
- Micevski T., Lerat J., Kavetski D., Thyer M., Kuczera G. (2011). Exploring the utility of multi-response calibration in river system modelling. *In International Congress on Modelling and Simulation (19th: 2011: Perth, Australia) MODSIM2011*.
- Monod J (1949) The growth of bacterial cultures. *Annual Reviews in Microbiology*, 3, 371-394.
- Nelder J.A. and Mead R. (1965) A simplex method for function minimization. *Computer Journal*, 7, 308–313.
- Perrin C., Michel C., Andréassian V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279.1: 275-289.
- Simons, M., Podger, G., and Cooke, R. (1996) IQQM—A hydrologic modelling tool for water resource and salinity management. *Environmental Software*, 11, 185 – 192.
- Vaze, J., Davidson, A., Teng, J., Podger, G. (2011). Impact of climate change on water availability in the Macquarie–Castlereagh river basin in Australia. *Hydrological Processes*, Vol 25, Issue 16, 2597–2612, DOI: 10.1002/hyp.8030.
- Vaze, J., Viney, N., Stenson, M., Renzullo, L., Van Dijk, A., Dutta, D., Crosbie, R., Lerat, J., Penton, D., Vleeshouwer, J., Peeters, L., Teng, J., Kim, S., Hughes, J., Dawes, W., Zhang, Y., Leighton, B., Perraud, JM., Joehnk, K., Yang, A., Wang, B., Frost, A., Elmahdi, A., Smith, A., Daamen, C (2013). The Australian Water Resource Assessment System (AWRA). Proceedings of the 20th International Congress on Modelling and Simulation (MODSIM2013), Adelaide, Australia, 1–6 December 2013.