

# Evaluation of Simhyd, Sacramento and GR4J rainfall runoff models in two contrasting Great Barrier Reef catchments

**Xike Zhang<sup>a</sup>, David Waters<sup>b</sup> and Robin Ellis<sup>a</sup>**

<sup>a</sup> *Queensland Department of Science, Information Technology, Innovation and the Arts, Australia*

<sup>b</sup> *Queensland Department of Natural Resources and Mines, Australia*

Email: [xike.zhang@science.dsitia.qld.gov.au](mailto:xike.zhang@science.dsitia.qld.gov.au)

**Abstract:** The eWater modelling framework has been used to predict streamflow and pollutant loads across the Great Barrier Reef (GBR) catchments, Queensland as part of the Reef Plan Paddock to Reef program. The Simhyd rainfall-runoff model was initially chosen to simulate streamflow across the GBR due to its extensive use for catchment hydrology studies across Australia, particularly in Queensland. The aim of this study was to determine whether an improved calibration could be achieved with alternative models, and secondly to identify the most suitable model and objective functions for high flow event simulations. Source Catchments was coupled to the Parameter Estimation Software Tool (PEST), to compare the performance of Simhyd to two widely-used rainfall-runoff models available in the Source framework, namely Sacramento and GR4J. Two contrasting catchments were selected with three nested gauging stations in each, with total drainage areas of approximately 10,000km<sup>2</sup>. The first is in the upper Burdekin Dry Tropics Region and the second in the Wet Tropics region, both are located in the northern GBR.

Statistical evaluation of the three rainfall-run models found that in Burdekin dry tropics catchment, all three models had a coefficient of determination ( $r^2$ ) greater than 0.75 and daily Nash Sutcliffe Coefficient (NSE) greater than 0.55 with the exception of a Simhyd result at one gauge. Overall, the Sacramento model achieved the highest  $r^2$  and NSE values. Similarly, all three models have simulated total flow volume ratios (bias) within 20% with the exception being Simhyd at one gauge. Sacramento outperformed the other two models when comparing high flow volume ratios (bias). When simulating middle and low range flows, the GR4J model simulated volume ratios greater than 20% for all sites. In the Wet Tropics catchment, the Sacramento model achieved the highest  $r^2$  and NSE compared to Simhyd and GR4J, although the latter two models achieved acceptable  $r^2$  and NSE values according to the criteria applied in this study. All three models simulated acceptable volume ratios of total flow and high flow in this drier Burdekin catchment.

The Sacramento model also achieved the lowest bias (volume ratio) in the simulation when comparing flow statistics for the three biggest flow events on record in both catchments. Visual assessment of the hydrographs also revealed that Sacramento model simulated the timing and magnitude of peak flow and the shape of recession curves of the large flow events better than the other two models.

The preliminary results from the calibration suggest that the Sacramento model may provide an improved calibration performance over Simhyd and GR4J. The use of Source Catchments coupled to PEST using multiple objective functions also improved calibration statistics over Source Catchment's in-built calibration tool. Future work will look at assessing other objective functions and alternative rainfall inputs to refine high flow calibrations.

**Keywords:** *Source Catchments, PEST, parameter calibration, objective function*

## 1. INTRODUCTION

The eWater ‘Source Catchments’ modelling platform has been used in the Reef Plan Paddock to Reef program for modelling the hydrology and water quality in the Great Barrier Reef catchments (Vaze *et al.*, 2011; Carroll *et al.*, 2012). For the initial calibration of all six GBR models, the Simhyd rainfall-runoff model was chosen due to its extensive use for catchment hydrology studies across Australia and in Queensland (Ellis *et al.*, 2009). However the initial calibration exercise across the GBR catchments indicates that high flow events are generally under predicted: an important consideration when the main objective of the program is focused on the estimation of pollutant loads. This paper explores whether hydrology simulations can be improved by (a) using alternative rainfall runoff models and (b) alternative high flow calibration objective functions. Post *et al.* (2007) compared the performances of three rainfall-runoff models, Simhyd, IHACRES (Identification of unit Hydrographs And Component flows from Rainfall, Evaporation and Streamflow data) and Sacramento, in calibrating and simulating six flow gauging stations in Victoria, Australia and found the Sacramento model performed best among the three models, and IHACRES better than Simhyd. Vaze *et al.* (2010) compared six models and found that the Sacramento and GR4J models outperformed others in simulating the daily runoff series in 232 catchments in southeast Australia. Based on these findings, we aim to compare the ability of three rainfall-runoff models (Simhyd, Sacramento and GR4J) in two catchments that are located in different climate zones, with an emphasis on simulating high flows more accurately.

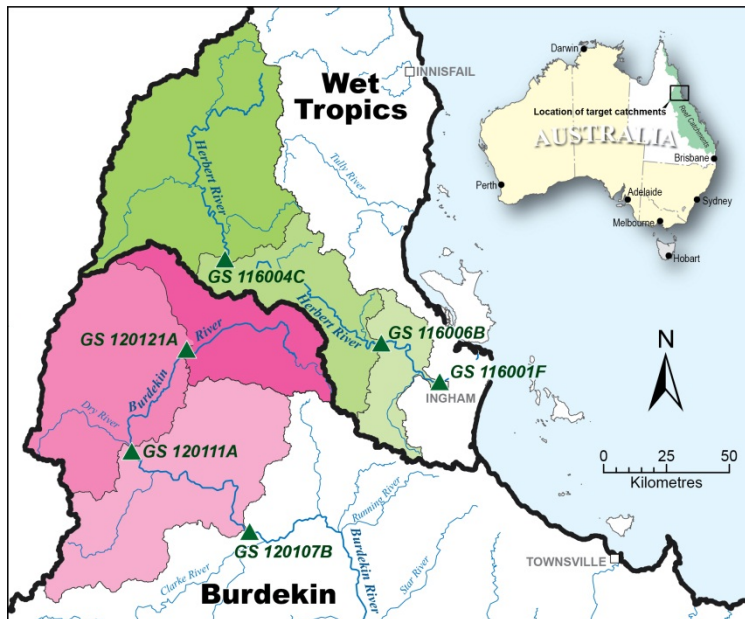
The command-line version of Source Catchments which allows running the model with many different input parameter settings was coupled with model-independent parameter estimation tool, PEST (Parameter ESTimation), to perform the calibration. A unique technique called Singular Value Decomposition Assist (SVDA) was also implemented in PEST to reduce model run time by defining a set of ‘super parameters’ from the large number of possible base parameters. For example, Ellis *et al.* (2009) used the SVDA to define 120 ‘super parameters’ from a possible 800 base parameters to calibrate water quality parameters in a previous GBR catchment modelling study. Another innovative technique called BEOPEST has been implemented in PEST. The name is derived from the original Beowulf Computer Cluster technique which ties a number of common computers together using a reasonably fast network and ‘slaves’ out calibration runs to any number of computers simultaneously. This process dramatically speeds up the calibration process and computation time.

The primary objective of this study is to assess the performance of three rainfall-runoff models when representing the hydrology of large complicated river systems typical of the GBR; and secondly, to identify appropriate objective functions for weighting calibration more towards high flow events.

## 2. STUDY CATCHMENTS

The Upper Burdekin sub-catchment is the part of the Burdekin basin upstream of where the Burdekin River flows into Lake Dalrymple and drains an area of approximately 10,528 km<sup>2</sup> (Figure 1). The sub-catchment covers less than 30% of the total Burdekin River catchment area but contributes more than 50% of the average annual stream flow of the Burdekin River. The three nested gauges chosen were Burdekin River at Lake Lucy Dam (GS 120121), Burdekin River at Lucy Downs (GS 120111), and Burdekin River at Blue Range (GS 120107). Blue Range, the most downstream gauge has a drainage area of 10,528 km<sup>2</sup> with an average annual stream flow of 1,165GL. Streamflow data recorded at the gauges ranged from 13 – 35 years.

The Herbert River Catchment is the largest of the Wet Tropics Catchments in North Queensland. In this study, we selected the Lower Herbert sub-catchment for calibration. Three nested gauging stations were chosen: Herbert River at Glen Eagle (GS116004), Herbert River at Abergowrie (GS116006) and Herbert River at Ingham (GS 116001). The Ingham gauge is the most downstream gauge having a drainage area of 8,581 km<sup>2</sup> with an average annual stream flow of 2,424GL, over double that of the Burdekin River at Blue Range. Streamflow data recorded at the three gauges ranged from 27 – 41 years.



**Figure 1.** Location of the three Upper Burdekin nested sub-catchments (pink) and the three Wet Tropics nested sub-catchments (green) used for calibration.

### 3. METHODOLOGY

#### 3.1 Source Catchment Configuration and Statistical Measures for Model Evaluation

Models were developed for the areas shown in Figure 1 in Source Catchments. Climate data was from SILO Data Drill (SILO 2009). Eleven land uses represented in the models were aggregated to represent three hydrologic response units: grazing, forestry land and cropping. Three rainfall-runoff models (Simhyd, GR4J and Sacramento) commonly used in Source Catchments were selected. The Laurensen non-linear flow routing model was chosen to simulate the flow routing processes for all links, however the two routing parameters of the Laurensen model were calibrated for each sub-catchment separately.

In order to compare the suitability of the three rainfall-runoff models for application in the GBR, identical Source Catchments projects for the wet (Herbert) and dry (Burdekin) catchments were populated with the relevant rainfall-runoff models and calibrated using the same observed data and objective function. Performance of each calibration was then assessed statistically and through visual analysis of high flow event characteristics. Statistical analyses included the coefficient of determination ( $r^2$ ) and Nash-Sutcliffe coefficient of efficiency (NSE) for daily flow volumes. The  $r^2$  value describes the proportion of the variance in the observed flow by the simulations. Published studies suggest that any value of  $r^2$  greater than 0.5 for daily flow comparison as an acceptable threshold for hydrology calibration (Moriassi *et al.*, 2007). The NSE indicates how well the simulated flow versus the observed against the 1:1 line. A NSE value of 0.5 for daily flow comparison is also regarded as an acceptable value in other hydrological studies (McCloskey *et al.*, 2011; Moriassi *et al.*, 2007). Other statistical measures used for model evaluation in this study include the flow volume ratio (percent bias) which is calculated as the percentage of the difference between the predicted and observed flow in the total observed flow volume. In this study, the flow volume ratio was calculated for different ranges of flow regime to systematically assess the model performance in simulating average flows: the volume ratios for the total flow (TVR), high flow (HVR, 0-10% of flow records), middle flow (MVR, 11-30% of flow records), and low flow (LVR, 31-100% of flow). A volume ratio value of no greater than  $\pm 20\%$  of observed flow volume is typically regarded as acceptable 'bias' in most hydrological studies (e.g. Moriassi *et al.*, 2007). In the GBR project, 20% or less of volume ratio was considered acceptable in this study.

#### 3.2 Calibration Objective Function

PEST optimises model parameters by repeatedly running Source Catchments and comparing the simulated flow predictions with observations through the use of a user-defined objective function. During the optimization process, the model selects new parameter sets to obtain lower objective function values until a minimum value is found. A typical objective function is comprised of the sum of weighted squared of

differences between certain model outcomes and their corresponding field-measured observations (Doherty, 2009).

An objective function that had previously been applied in several GBR catchments comprised of log-transformed daily flows, monthly flow volumes, and flow duration curves (Stewart, 2013; McCloskey *et al.*, 2011). Preliminary investigations for this study compared the calibrations using log-transformed daily flow in the objective function with those from non-transformed daily flow and found the former achieved poorer calibration statistics, especially when considering the desire to represent high flows more accurately. Therefore the log-transformation of daily flow is not used in this study. Due to the manner in which PEST calculates exceedance curves for calibration there is also a need to weight the exceedance values in the objective function to achieve the equal contribution from the exceedance curve to the whole objective function. A specific weighting factor was derived and applied for the exceedance duration curve when included in the objective function. The final objective function used in this study comprises of daily flow, monthly flow volume and weighted exceedance duration curves.

### 3.3 PEST

To calibrate a large number of hydrological parameters, a ‘regularisation’ process is required that allows a large number of parameters to be simultaneously estimated without incurring numerical instability. As a solution, PEST has implemented Singular Value Decomposition-Assist (SVDA) technique for parameter optimisation (Doherty, 2009). By using SVDA, PEST actually estimates ‘super parameters’ which are linear combinations of ‘real’ or base parameters, thus significantly reducing the number of parameters being estimated and dramatically reducing the number of model runs required. For this reason, we applied the SVDA technique in this study.

PEST offers several solutions to completing required calibration model runs in a faster time frame including the BEOPEST. BEOPEST involves the ‘slaving’ of model runs to as many other processors as is available to the modeler. Combining SVDA and BEOPEST with PEST’s unique approach to objective function minimization achieves significant efficiencies when calibrating complex hydrology models with PEST.

## 4. RESULTS AND DISCUSSION

### 4.1 Model Performance at Long-Term Simulation

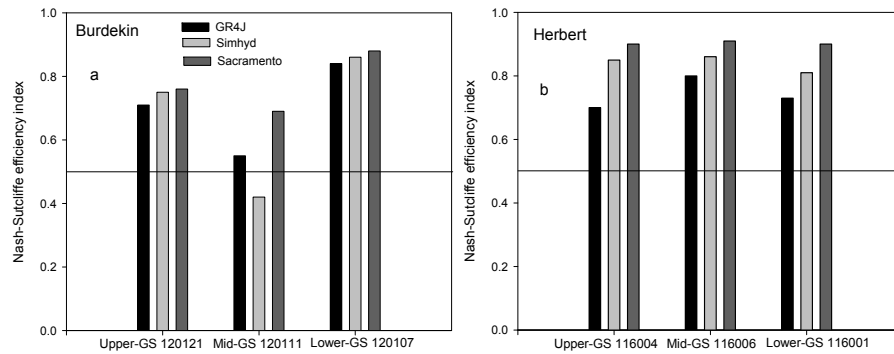
The results in this study have shown that in the Burdekin Catchment, the  $r^2$  value for the Sacramento model is equal to those of Simhyd at GS 120121 (most upstream gauge), but higher than that of GR4J. At the mid sub-catchment GS 120111 and the outlet GS 120107, the  $r^2$  values achieved by Sacramento model are the highest among the three models. The NSE values for the three models follow the similar trend at this gauge (Figure 2a). For the GS 120111 located the mid-section of the catchment, the  $r^2$  value of Sacramento is greater than that of GR4J and the latter greater than that of Simhyd. The NSE values are following the same pattern and it is noted that the NSE of Simhyd at GS 121111 is lower than 0.5.

In the Herbert Catchment, both  $r^2$  and NSE values of the three models at all three gauges are above the acceptable level, with  $r^2$  value between 0.85-0.95 and NSE between 0.70-0.91 (Figure 2b). The  $r^2$  and NSE values in this Catchment have shown a clear pattern: Sacramento performs better than Simhyd, and Simhyd performs better than GR4J.

In the Burdekin catchment, all three rainfall-runoff models simulated less than 20% of total and high flow volume ratio at all gauges, except that Simhyd simulated greater than 20% high flow volume ratio at GS 120111. For the mid and low range flows, GR4J model simulated greater than 20% flow volume ratio at all three gauges. Comparatively, Simhyd and Sacramento model produced variable results across mid and low flows.

In the Herbert catchment, again all three rainfall-runoff models have simulated total flow volume ratio and high flow ratio within 20%. For the mid-range flow, Simhyd simulated less than 20% of volume ratio at all three gauges. Sacramento similarly estimated less than 20% of mid flow ratios at two gauges but greater than 20% at one gauge. Comparatively, GR4J simulated greater than 20% of middle flow ratios at two gauges. For the low range flow, GR4J model simulated greater than 20% of flow ratio at all three gauges. Both Sacramento and Simhyd models predicted less than 20% of low flow ratios at two gauges but greater than 20% at one gauge.

When calibration performance is assessed by  $r^2$  and NSE, the Sacramento and GR4J models performances are similar to Vaze *et al.* (2010) results in southeast Australia, where the models outperformed 4 other models in re-producing the daily runoff series in 232 catchments. When the flow volume ratios were used to assess the models performance together with  $r^2$  or NSE, Sacramento still performed best of the three models; however GR4J achieved poor statistics especially in simulating middle and low range flows. Overall, all three models performed better in the wet catchment than in the dry catchment, which may be caused by many regular no flow periods recorded in the drier Burdekin catchment.



**Figure 2.** Nash-Sutcliffe efficiency comparison for the three Rainfall-Runoff models at the upper, mid and lower gauging stations in (a) Burdekin Catchment, and (b) Herbert Catchment. The vertical bar in the plots represents NSE value and the horizontal straight line is the 0.5 NSE. Any NSE bar above the line is deemed to be acceptable

#### 4.2 Model Performance at High Flow Simulation

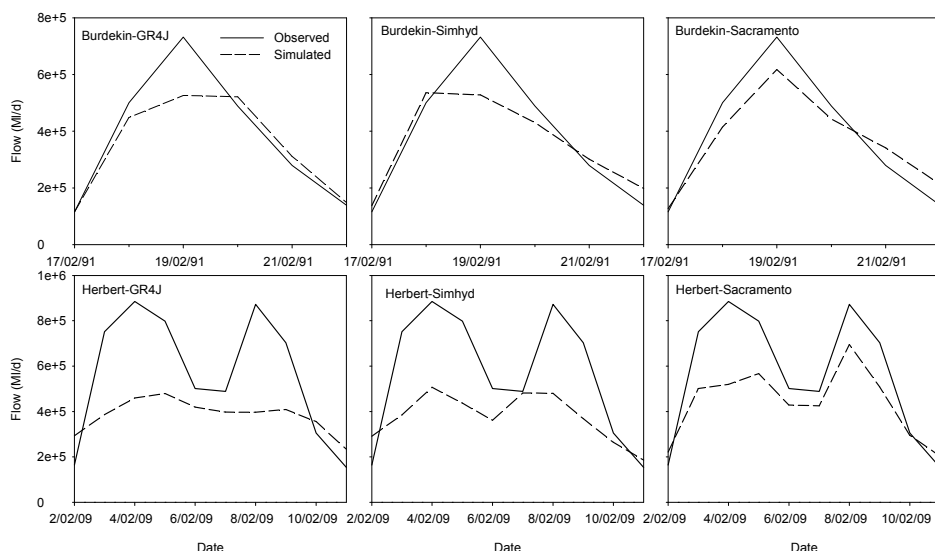
High flow events are very important for the water quality in the GBR as high flows contribute much of the pollutant load to the reef. In this study, to assess model performance we have selected the three high flow events recorded at the outlet gauge (GS 120107) of the Burdekin catchment, and at GS 116001 in the Herbert catchment. In the Burdekin the flow volumes for the three biggest events at GS 120107 were 376,104 ML/d (17-22, February 1991), 341,328 ML/d (3-10, February 2009) and 210,445 ML/d (10-14, January 1991) and total volumes of the three events were 194%, 234% and 90% of average annual flow volume recorded at the gauging station. For the Herbert at GS 116001, the flow volumes of the three biggest events were 606,469 ML/d (3-11, February 2009), 534,522 ML/d (10-16, January 1998) and 298,919 ML/d (18-22, February 1991) and total volumes of the three events were 182%, 89% and 70% of average annual flow volume. We then examined the performance of the three rainfall-runoff models in simulating the high flow events. Table 1 shows that although all three models under-estimated the three biggest flows in the Burdekin catchment, the volume ratios are within 20%, except Simhyd which simulated 23.1% of volume ratio for event 3. Among the three rainfall-runoff models, Sacramento simulated the smallest volume ratio values (percent bias) across the three events. In the Herbert catchment, both GR4J and Simhyd simulated greater than 20% of volume ratios for all three events. However, Sacramento simulated less than 20% of the volume ratios of event 2 and event 3 and a 24.1% ratio for event 1.

**Table 1.** The volume ratio (in percent) of predicted to observed values for three high flow events at the outlet gauge in the Burdekin and Herbert catchments. The volume ratio is calculated as the percentage of the difference between the predicted and observed flow in the total observed flow volume and the negative value of volume ratio indicates underestimation of the modeled flow

| Model      | Burdekin model flow ratios |                     |                     | Herbert model flow ratios |                     |                     |
|------------|----------------------------|---------------------|---------------------|---------------------------|---------------------|---------------------|
|            | Event 1<br>(6 days)        | Event 2<br>(8 days) | Event 3<br>(5 days) | Event 1<br>(9 days)       | Event 2<br>(7 days) | Event 3<br>(5 days) |
| GR4J       | -8.2%                      | -19.4%              | -19.2%              | -35.2%                    | -28.4%              | -36.7%              |
| Simhyd     | -5.6%                      | -18.1%              | -23.1%              | -36.5%                    | -21.9%              | -31.8%              |
| Sacramento | -4.4%                      | -14.4%              | -14.8%              | -24.1%                    | -16.0%              | -10.7%              |

Hydrographs of event 1 in the Burdekin and Herbert catchments were examined to compare the three rainfall-runoff models ability to simulate the timing and magnitude of peak flow and the shape of recession curves of

high flow events (Figure 3). Visual assessment confirms that Sacramento simulated the better timing and magnitude of peak flow and the shape of recession curves of the high flow event, compared with GR4J and Simhyd in both models.



**Figure 3.** Hydrograph of predicted daily flows vs. observed flows during the period of the highest flow event in the record at the outlet gauge in Burdekin Catchment (upper plots) and Herbert Catchment (lower plots)

This study has shown that Sacramento outperformed Simhyd and GR4J model in both re-producing the long-term time series of daily flow (measured by  $r^2$ , NSE and flow volume ratios) and simulating single high flow events. One reason for the better performance of Sacramento may be attributed to its large number of parameters (22) that can be calibrated in contrast to Simhyd (9 parameters) and GR4J (4 parameters). However, where model calibration run time is an important consideration, the increased number of parameters associated with Sacramento will impact on model iterations and run time. Thus it is recommended that where PEST is used for calibration, SVDA or BEOPEST techniques should be applied.

## 5. SUMMARY AND CONCLUSIONS

This study has used a method of coupling Source Catchments with model-independent parameter optimisation tool, PEST and using the SVDA techniques to calibrate hydrological parameters of three widely-used rainfall-runoff models in two catchments that are located in different climate zones.

In the Burdekin catchment which is located in a dry and sub-tropical climate zone with low base flow, the highest  $r^2$  and NSE values between predicted flows and observed values are achieved using Sacramento. The  $r^2$  value and NSE achieved by GR4J are all greater than 0.55 at all three gauges but lower than those of Sacramento's. Simhyd achieved greater than 0.75  $r^2$  values at three gauges and NSE at two gauges but a lower than 0.5 NSE at one gauge. The Sacramento model also performed the best in predicting the total flow volume and high flow volume, with the lowest volume ratios (bias) compared with Simhyd and GR4J. Though simulating within 20% of the total flow and high flow ratios, GR4J simulated greater than 20% volume ratios for middle and low flows. For the wetter Herbert catchment all three models simulated within 20% volume ratios for total flow and high flow in this catchment. However, for the middle and low flow categories, GR4J again performed poorly. For simulating the three biggest flow events, in Herbert catchment, both GR4J and Simhyd were not able to achieve 20% of volume ratio of the three highest flow events but Sacramento achieved that at two gauges. Sacramento also achieved the lowest bias (volume ratio) in both Burdekin and Herbert Catchment.

In summary, model evaluation in this study found that Sacramento model has performed better than GR4J and Simhyd in simulating daily flow and high flow events in both dry sub-tropical catchments and wet tropical catchments in GBR region. The method of combining the SVDA and BEOPEST techniques implemented in PEST substantially reduced required model run numbers hence run time. This improvement

in model efficiency makes the adoption of complicated yet robust rainfall-runoff models, like Sacramento, more likely and useful in large projects like the GBR catchment modelling undertaken as part of the Paddock to Reef program.

## REFERENCES

- Carroll, C., Waters, D., Vardy, S., Silburn, D.M., Attard, S., Thorburn, P.J., Davis, A.M., Halpin, N., Schmidt, M., Wilson, B. and Clark, A. (2012). A paddock to reef monitoring and modelling framework for the Great Barrier Reef: Paddock and catchment component, *Marine Pollution Bulletin*, doi:10.1016/j.marpolbul.2011.11.022.
- Doherty, J. (2009). PEST: Model-Independent Parameter Estimation. *Watermark Numerical Computing*. <http://www.pesthomepage.org>
- Doherty, J. (2010). Windows BEOPEST. *Watermark Numerical Computing*. <http://www.pesthomepage.org>
- Ellis, R.J., Doherty, J., Searle, R.D. and Moodie, K. (2009). Applying PEST (Parameter ESTimation) to improve parameter estimation and uncertainty analysis in WaterCAST models, 18<sup>th</sup> International Congress on Modelling and Simulation, Cairns, Australia. 13-17 July 2009, pp3158-3164.
- McCloskey, G.L., Ellis, R.J., Waters, D.K. and Stewart, J. (2011). PEST hydrology calibration process for source catchments – applied to the Great Barrier Reef, Queensland, 19<sup>th</sup> International Congress on Modelling and Simulation, Perth, Australia. 12-16 December 2011, pp2359-2366.
- Moriassi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., and Veith, T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *American Society of Agricultural and Biological Engineers*, 50(3):885-900.
- Petheram, C., Rustomji, R., Chiew, F.H.S., and Vleeshouwer, J. (2012). Rainfall-runoff modelling in northern Australia: A guide to modelling strategies in the tropics, *Journal of Hydrology*, 28-41, doi:10.1016/j.jhydrol.2011.12.046.
- Post, D.A., Vaze, J., Viney, N. and Chiew, F.H.S. (2007). Regionalising the hydrologic response of ungauged catchments using the Simhyd, IHACRES, and Sacramento rainfall-runoff models, 17<sup>th</sup> International Congress on Modelling and Simulation, Christchurch, New Zealand. December 2007, pp2534-2540.
- SILO (2009). The SILO Data Drill. Queensland Department of Environment and Resource Management. Available from <http://www.longpaddock.qld.gov.au/silo>.
- Vaze, J., Chiew, F. H. S., Perraud, J.M., Viney, N., Post, D. A., Teng, J., Wang, B., Lerat, J., Goswami, M., 2010. Rainfall-runoff modelling across southeast Australia: datasets, models and results. *Australian Journal of Water Resources*, Vol 14, No 2, pp. 101-116.
- Vaze, J., Jordan, P., Beecham, R., Frost, A., Summerell, G. (eWater Cooperative Research Centre 2011) Guidelines for Rainfall-Runoff Modelling: Towards Best Practice Model Application. ISBN 978-1-921543-51-7. [http://www.ewater.com.au/uploads/files/eWater-Guidelines-RRM-\(v1\\_0-Interim-Dec-2011\).pdf](http://www.ewater.com.au/uploads/files/eWater-Guidelines-RRM-(v1_0-Interim-Dec-2011).pdf).
- Viney, N.R., Perraud, J., Vaze, J., Chiew, F.H.S., Post, D.A. and Yang, A. (2009). The usefulness of bias constraints in model calibration for regionalisation to ungauged catchments, 18<sup>th</sup> International Congress on Modelling and Simulation, Cairns, Australia. 13-17 July 2009, pp3421-3427.