20th International Congress on Modelling and Simulation, Adelaide, Australia, 1–6 December 2013
www.mssanz.org.au/modsim2013

An evaluation of hydrological models for predicting mean-annual runoff and flood quantiles for water quality modelling

F. Karim [a], S. Wilkinson [a] and C. Dougall [b]

[a] *CSIRO Land and Water, Christian Laboratory, Canberra, ACT 2601; Australia*
[b] *Environment and Resource Management, Queensland Government, Emerald, QLD 4720, Australia*
Email: fazlul.karimt@csiro.au

Abstract: Predicting spatial variations in runoff is important for water resource assessment and understanding spatial variation in flood quantile is valuable for assessing flooding risk and sediment loads during floods. Sediment fluxes in river networks are disproportionally sensitive to runoff events, river bank erosion and floodplain deposition. Many spatially disaggregated models of erosion and sediment transport require inputs of long term runoff volume and daily flood quantiles for each link in a stream network. This paper presents an evaluation and comparison of two different hydrological models in predicting catchment water yields and flood quantiles throughout a stream network. Using data from tropical Queensland catchments, a simple regionalization model (SedNet) and a daily time step conceptual rainfall-runoff model (SIMHYD) were calibrated using a set of gauge data and evaluated for ungauged condition using another set of gauge data. Results showed that a daily timestep rainfall-runoff model could provide better calibration to observed water yield but a regionalised model was as good as or better than the daily time step model for calibration to flood quantile. For prediction (on ungauged condition), both models produced similar results in predicting mean annual water yield and flood quantiles, with slightly better prediction of daily flow variability by the regionalised model. Based on an analysis of sensitivity of bankfull recurrence interval, we found large uncertainty in predicting bankfull flow using the SIMHYD model. Results imply that a simple regionalized model is as good as a complex daily time step model for long-term sediment budgets in water quality modelling. Using a daily time step, however, could be necessary if event modelling of flood information and sediment is required.

Keywords: *Flood quantile, performance evaluation, erosion, regionalisation, Great Barrier Reef*

1. INTRODUCTION

There are two approaches commonly taken to modelling the spatial variations in runoff volumes and flood quantiles across large river basins. Firstly by calibrating models of the hydrologic response of runoff timeseries to upstream rainfall using conceptual or mechanistic representations of transient stores and flows and applying the models to hydrologically-similar catchments (Post, 2009). Secondly by fitting regression models commonly referred to as regionalisations across river stations between a runoff statistic of interest and spatial variations in catchment characteristics such as mean rainfall, relief and potential evapotranspiration (Surian and Andrews, 1999; Wilkinson et al., 2006). Modelling of sediment sources and fluxes across catchments is often used to provide a technical underpinning for erosion remediation investments including simulating alternate land management scenarios. Such modelling needs spatial information of runoff volume and flood quantiles as inputs and has specific requirements for the quality of hydrologic modelling predictions, including accurate estimates of spatial patterns in bankfull and median overbank discharges, which differ from those of engineering applications such as water resource assessment or hydraulic structure design. However, despite the plethora of erosion and sediment flux models, the relative suitability of hydrologic response modelling and regionalisation approaches for this type of application is not well understood.

This study compares two methods of predicting mean annual discharge and other flow metrics used in sediment budget modelling, including flood quantiles of defined recurrence intervals: i) Regional relationships between observed flow statistics and catchment characteristics including rainfall and evapotranspiration (SedNet model), and (ii) Daily conceptual models of the hydrological response using the SIMHYD model. The comparison is undertaken using data from the Burdekin River basin, which drains to the Great Barrier Reef (GBR) in tropical Queensland, Australia. The models were evaluated for both goodness of fit to a set of stream gauges used in calibration and a separate set of gauges that were excluded in calibration (i.e. prediction).

2. MATERIAL AND METHODS

2.1. Study Area

The Burdekin basin is the second largest catchment that drains into the GBR lagoon. It covers an area of 134,000 km^2. Much of the basin is characterised by dry-tropical savannah, and cattle grazing covers about 90%, cropping 5% and nature conservation 5%. Rainfall varies significantly across the basin, with the annual average being as low as 500 mm in the south-west (dry tropics) and exceeding 1600 mm in the north-east (wet tropics). Evaporation rates are high, and average annual evaporation across most of the area is in the order of 2000 mm indicating a substantial water budget deficit. The hydrological regime is ideal for evaluating model predictions of flood quantiles, since it is highly seasonal and variable between years (Petheram et al., 2008). More than 80% of runoff occurs during the summer months, between November and April. Most of the stream network is ephemeral. The mean annual discharge at the basin outlet has flow volume varying from a minimum of 400 GL, mean of 9300 GL and maximum of 53,000 GL; covering a range of two orders of magnitude (Post, 2009).

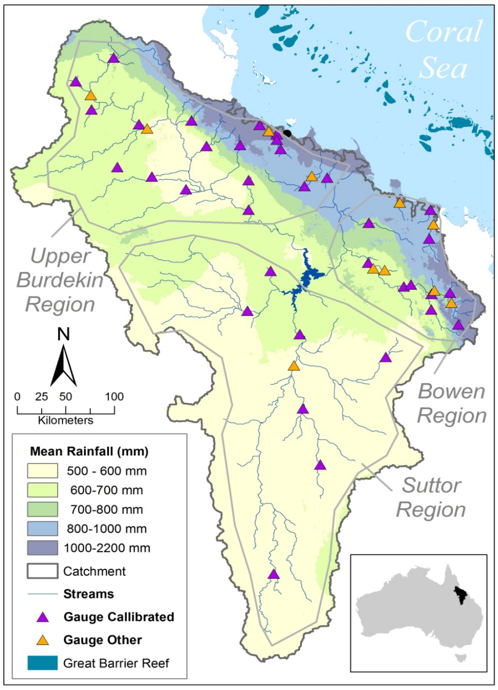

Figure 1. The Burdekin catchment and its three physiographic regions showing shaded rainfall, location of gauging stations and streams having catchment area of 500 km^2 or more. Inset shows the location of the catchment.

2.2. Hydrological variables

Spatial variations in mean annual flow (*MAF*) across river networks are used in SedNet to calculate suspended sediment trapped in reservoirs, and in many other applications. Secondly, bankfull discharge (Q_{bf}), defined as a flood of given recurrence interval on the annual maximum series, is used to estimate stream bank erosion and overbank flow. Based on local observations, we applied a recurrence interval of 5 years in this study (Q_5). Thirdly, floodplain sediment deposition is estimated using the median value of overbank discharges (Q_{mo}); being the median amount of flow above bankfull flow ($Q - Q_5$, where Q is the daily flow rate) of a flow series containing only days where flow is greater than Q_{bf}. Fourthly, following Prosser et al. (2001), bed material sediment transport capacity of river flow is assumed proportional to daily flow raised to a power 1.4. A non-dimensional measure of daily flow variability commonly known as sigma-daily flow (σ_d) is used to represent the effect of flow on spatial patterns in bedload sediment transport capacity:

$$\sigma_d = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Q_i}{Q_m}\right)^{1.4} \tag{1}$$

where n is the numbers of data, Q_i is the flow on i^{th} day and Q_m is the mean flow over the data period. Each of these statistics represents the integrated effects of discharge variability on the erosion, transport or deposition process, as an alternative to running the model at daily time-steps (Wilkinson et al., 2006).

2.3. SedNet regionalisation

To calculate each flow statistic for every link in the SedNet model, they are first calculated at river gauging stations within the basin that have long term records of daily discharge. The discharge statistics are then extrapolated or regionalized to ungauged river links using least-squares regression models based on catchment characteristics. Mean annual discharge is predicted using a physically based regionalisation of mean annual runoff coefficient (Wilkinson et al., 2006).

$$MAF = R_C P A \tag{2}$$

where P and A are rainfall and catchment area respectively (known parameters) and R_C is runoff coefficient, which is regionalised as a function of rainfall and evaporation based on the assumption that losses to changes in soil moisture and deep drainage are small in the long term (Wilkinson et al., 2006).

Bankfull and overbank discharges are regionalised as a power function of *MAF*. Compared with regionalizing *MAF*, the physical basis for regionalizing the Q_{bf} and the Q_{mo} is weaker and some variation in approaches between river basins may be justified. Following Wilkinson et al. (2009) we used the following relationships:

$$Q_{bf} = g((MAF))^h \tag{4}$$

$$Q_{mo} = j(MAF)^k \tag{5}$$

where g, h, j and k are calibration parameters that are estimated based on regression analysis.

There is also little theoretical basis on which to select variables for regionalizing σ_d. Based on studies around Australian catchments (Wilkinson et al., 2006) we applied an exponential relationship against P:

$$\sigma_d = mP^n + 1 \tag{6}$$

where m and n are calibration parameters that are estimated based on regression analysis.

SedNet regionalisations were implemented firstly across the Burdekin River basin as a whole, and secondly independently across 3 physiographic regions (Upper Burdekin, Bowen and Suttor; Figure 1), as defined in a previous regionalisation study (Post, 2009).

2.4. Conceptual rainfall-runoff modelling

We used the SIMHYD model (Chiew et al., 2002), which is a conceptual model that uses 7 model parameters to simulate daily runoff based on daily catchment average rainfall and potential evapotranspiration (PET). The parameters represent 3 runoff components: overland flow (infiltration excess runoff), interflow (saturation excess) and base flow. The SIMHYD parameters were calibrated using PEST by minimising an objective function defined as the sum of squares of the differences between modelled and observed daily and monthly flows as described in McCloskey et al. (2011). The calibration objective was to meet the following criteria for each calibration gauge. These were daily Nash-Sutcliffe coefficient of efficiency (*NSE*) of greater than 0.5, monthly *NSE* of greater than 0.8 and volume error of less than 20%. Each gauge had an independent parameter set, but all gauges were calibrated together to achieve the optimum objective function across all gauges. The contributions of each time-scale and each gauge to the objective function were weighted equally. Each gauging station represented its own calibration region, and calibrated parameters were applied when simulating time-series for all links within that region. For river links outside calibration catchments, parameters were applied using the nearest neighbour approach (Zhang and Chiew, 2009). Simulated daily flow was then processed to calculate *MAF*, Q_{bf}, Q_{mo}, and σ_d.

2.5. Model implementation and performance evaluation

Streams and subcatchment boundaries were generated based on a 270 m hydrologically corrected digital elevation model (DEM) using ArcInfoTM routines. We constrained first order stream by using an area threshold of 50 km^2. The network has a node-link structure, where links route water downstream between network nodes. Each link has a subcatchment that drains directly to it, and out flow from each link is calculated using water balance analysis. The stream network consists of 1812 links having average length of 8.7 km, average channel width of 81.5 m and average link slope of 0.024. Major stream links having catchment area of 500 km^2 or more are shown in Figure 1. The models were run for the period of 1970 to 2011.

For SedNet modelling, we used grid based (250 × 250 m) mean annual rainfall and evaporation derived based on the Bureau of Meteorology ground-based data. SIMHYD takes daily rainfall averaged across each subcatchment. Rainfall data were obtained from the Bureau of Meteorology for the 45 rain gauges located across the Burdekin catchment and mean areal rainfall for each subcatchment was estimated using Thiessen method.

Daily streamflow data were obtained from the Queensland Department of Environment and Resource Management (DERM). A total of 47 gauges were used in model calibration and prediction (Figure 1). These were selected from the 110 current and historical gauges in the basin using several criteria. Gauges having a minimum catchment area of 50 km^2 and at least 8 years of data record were considered and gauges that were heavily regulated were excluded from the analysis. Records from gauges which had been relocated during the course of operation but had contributing area <5% difference were merged into 1 continuous dataset. Both the SedNet and SIMHYD models were calibrated using 35 stream gauges located across the catchment representing the spatial heterogeneity of hydrological properties. The remaining 12 gauges were used to assess predictive performance.

Based on recent reviews of relevant literature on hydrological model evaluation (e.g. Moriasi et al., 2007), we selected two main quantitative statistics, the Nash-Sutcliffe efficiency (*NSE*) index and Percent bias (*PBias*) error index. In addition to above performance criteria we evaluated a discrepancy ratio (D_R), being a scale independent measure of absolute error in model prediction of each flow variable which equals the factor by which the predicted metric differs from that observed (Rustomji et al., 2008).

$$D_R = 10^{|\log(pre) - \log(obs)|} \quad (7)$$

A D_R of 1 indicates perfect agreement. D_R =1.2 means that the prediction is within ± 20% of the observation and D_R = 2 means that the prediction is within a factor of 2 of the observation (either higher or lower). This measure scales overpredictions and underpredictions evenly.

3. RESULTS

3.1. Mean Annual Flow

Across the calibration gauges both models produced similar results for the catchments with runoff < 200 mm/year and SIMHYD appeared to provide better goodness of fit to observed *MAF* than SedNet regionalisations. The gauge discrepancy ratios D_R also indicate that SIMHYD performed better, with the main difference being that SedNet regionalisations had several gauges with large errors (Figure 2a). For SIMHYD, 100% of predictions were within a factor of 2, compared with 90% for sub-basin SedNet regionalisations and 82% for the whole-basin regionalisation. Both SIMHYD and the sub-basin regionalisation had 80% of gauges with discrepancy ratio of < 1.3. The evaluation metrics are consistent with the SIMHYD model having a better fit to observed data across the calibrated gauges. Both model slightly overestimated *MAF*, with SIMHYD being within 1.5% on average (Table 1).

Model predictive performance (based on 12 gauges excluded from calibration) was poorer than across the calibrated gauges for SIMHYD, in terms of D_R (Figure 2e) and *NSE* (Table 2). In contrast SedNet returned a similar profile of discrepancy ratios (Figure 2e), and similar *NSE* in prediction mode (Table 2) as across the calibration gauges. This is related to SedNet predictions being less prone to poor model performance at individual gauges than SIMHYD, though this difference in behaviour was small considering the number of gauges.

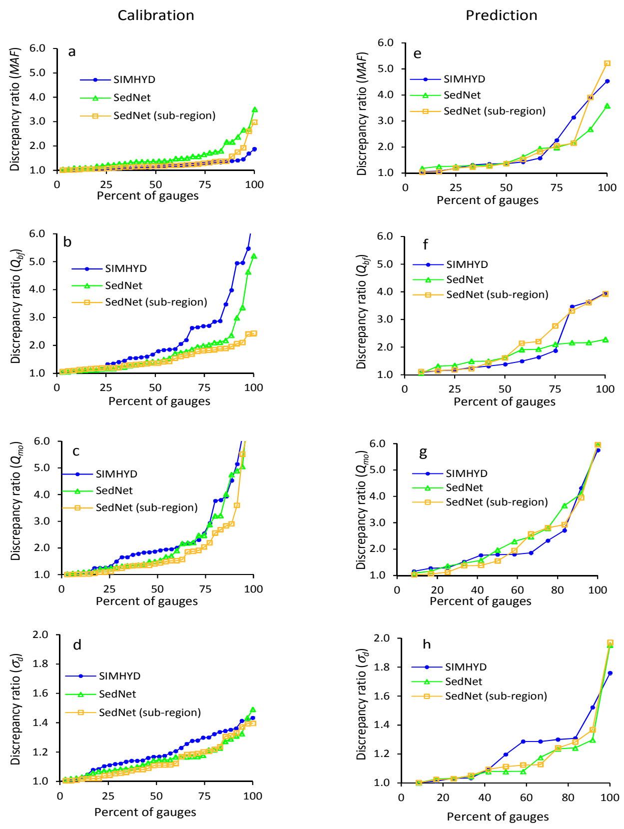

Figure 2 Comparison of discrepancy ratio between observed and predicted *MAF*, Q_{bf}, Q_{mo} and σ_d for (a-d) Calibrated and (e-h) predicted conditions.

3.2. Bankfull Discharge

Goodness of fit to observed Q_5 across calibration gauges was poor relative to *MAF*. For the majority of the gauges SedNet overestimated and SIMHYD underestimated bankfull flow (Table 1). In contrast to *MAF*, the SedNet regionalisations generally produced smaller errors than SIMHYD, with the sub-basin regionalisation performing best and being within a factor of 2 for 90% of gauges (Figure 2b). In terms of *NSE* SIMHYD results are slightly better but it greatly underestimated bankfull flow by 33% comparing with 7% overestimate by SedNet (Table 1). Results also show relatively large D_R for SIMHYD comparing with SedNet (Figure 2b). Simulated results were investigated to explore these contrasting results and found SIMHYD underpredicted Q_5 for majority of the gauges that produced large *PBias* while SedNet both overestimated and underestimated Q_5 keeping *PBias* small.

At prediction, the SIMHYD results were slightly better than the regionalisations for the majority of the gauges however D_R was > 3 for 3 gauges. As for *MAF*, SedNet errors were similar in prediction mode as across the calibration gauges, with 80% of gauges having D_R < 2.1 at calibration < 2.2 at prediction. For the same number of gauges SIMHYD model produced D_R of 2.8 or less at calibration and 2.0 for prediction. In terms of *NSE* and *PBias* SIMHYD results are slightly better comparing with SedNet. *PBias* is only 2% for SIMHYD comparing with large *PBias* in SedNet (26%).

The recurrence interval of the flood quantile was thought to be an influential parameter in predicting bankfull and median overbank discharges. The Q_5 analysis was repeated for Q_2 and Q_{10}. Results of

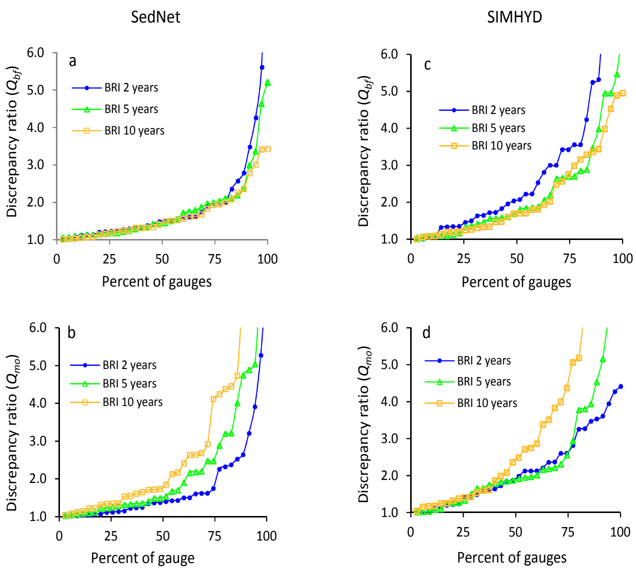

Figure 3 Sensitivity of bankfull recurrence interval (BRI) on predicting flood quantile for (a-b) SedNet model, and (c-d) SIMHYD model.

SIMHYD were found more sensitive to changes in the recurrence interval than SedNet for bankfull discharge and SedNet was found more sensitive for Q_{mo} (Figure 3).

3.3. Median Overbank Discharge

Overall estimates of Q_{mo} (calibrated condition) by both models were poorer than for *MAF* and Q_5. Both models produced similar results with very large discrepancies (>4.0) for about 10% of gauges (Figure 2c). In terms of *NSE* both model performed similar, although SIMHYD had a large positive bias indicating underestimate of mean Q_{mo} (Table 1). At prediction both models had similar profiles of D_R, but they were generally larger than for *MAF* and Q_5 (Figure 2g). It is interesting to note that SIMHYD results at predictions are relatively better than across the calibration gauges. In terms of *NSE* both model produced better prediction results comparing with *MAF* or Q_5. Both models produced similar *PBias* however SIMHYD underestimated and SedNet overestimated Q_{mo}.

3.4. Sigma-Daily Flow

In general SedNet overestimated and SIMHYD underestimated sigma-daily at the calibration gauges. In terms of discrepancy profiles SedNet predictions were generally better than SIMHYD but the difference was not large (Figure 2d). Again, sub-regional scale SedNet regionalisation produced slightly better results than the whole-basin regionalisation. However, model performances at predicting the observed spatial variations in sigma-daily were poor (*NSE* < 0.5, Table 1). Both models had similar performance at prediction (Figure 2h), with SedNet being closer to observed values in most cases.

Table 1 Statistical evaluation of model performance at the 35 gauges used in model calibration

	MAF		Bankfull flow		Median overbank		Sigma daily	
	NSE	*PBias (%)*	*NSE*	*PBias (%)*	*NSE*	*PBias (%)*	*NSE*	*PBias (%)*
SIMHYD	0.98	-1.52	0.85	32.70	0.66	43.03	0.06	11.33
SedNet	0.81	-6.24	0.69	-6.58	0.70	-4.98	0.13	-0.05
SedNet (sub-region)	0.93	-1.52	0.87	-6.67	0.82	-9.41	0.19	0.35

Table 2 Statistical evaluation of model performance for 12 gauges excluded from model calibration

	MAF		Bankfull flow		Median overbank		Sigma daily	
	NSE	*PBias (%)*	*NSE*	*PBias (%)*	*NSE*	*PBias (%)*	*NSE*	*PBias (%)*
SIMHYD	0.68	-36.10	0.73	-2.57	0.87	18.91	-0.44	4.65
SedNet	0.81	-20.56	0.68	-25.78	0.84	-21.14	-0.07	-5.11
SedNet (sub-region)	0.75	-32.48	-0.13	-82.35	0.67	-36.81	-0.27	-2.76

4. DISCUSSION

The major findings of this study is that SIMHYD (a daily time step conceptual rainfall-runoff model) produces better fit to observed flow statistics but it doesn't always produce better predictive capacity at ungauged locations within the basin than the SedNet regionalisations, particularly when those regionalisations are applied within hydrologically similar regions (e.g. SedNet produced slightly better results for *MAF* and sigma-daily flow). One of the reasons is SIMHYD calibration process takes account of spatial variations in catchment attributes for individual gauge using seven parameters while SedNet uses two parameters representing combined effects of all attributes at a regional scale. Also a three part objective function (i.e. daily *NSE*, monthly *NSE* and volume error) was used in SIMHYD calibration. These produced better calibration but didn't necessarily produce better results at prediction. In future applications of daily time-step modelling, regionalisation of rainfall-runoff parameters may be preferable than allowing parameter variation at individual gauges, to provide similar or slightly improved predictive capacity (Post, 2009).

Regionalisation metrics to consider will represent aspects of climate, geography, topography and vegetation (Bates et al., 1998), including the degree of seasonality and upstream channel length (Post, 2009). However, large improvements cannot be expected, since the exceedence curve of daily *NSE* shows similar performance to that of Post (2009). Considering other objective functions for calibration of hydrologic response model may provide small improvements in predictive performance.

5. CONCLUSIONS

Two methods for predicting daily flow variability and flood quantile were compared using data from tropical Queensland. Results show that daily time step model provide better calibration to *MAF*. However, both models had similar performance in predicting flood quantile but regionalized model is slightly better if calibrated in sub-region scale. SedNet model performed slightly better in predicting MAF and sigma-daily flow and SIMHYD model performed slightly better in predicting bankfull and median overbank flow. It has been concluded that a daily time-step model could be better for an event based water quality modelling but a regionalized model could provide similar results for long term modelling. This information would be useful to selecting an appropriate water quality model for verities of water quality issues.

ACKNOWLEDGMENTS

The authors would like to acknowledge funding for this research by eWater Ltd. The Queensland Government Department of Environment and Resource Management is gratefully acknowledged for providing streamflow data. Thanks to Anne Henderson of CSIRO and anonymous reviewers.

REFERENCES

Bates, B.C., Rahman, A., Mein, R.G., Weinmann, P.W. (1998). Climatic and physical factors that influence the homogeneity of regional floods in southeastern Australia. *Water resources research*, 34, 3369-3381.

Chiew, F.H.S., Peel, M.C., Western, A.W. (2002). Application and testing of the simple rainfall-runoff model SIMHYD. In: Singh, V.P., Frevert, D.K. (Eds.), Mathematical Models of Small Watershed Hydrology and Applications. *Water Resource Publication*, Colorado, pp. 335–367.

McCloskey, G.L., Ellis, R.J., Waters, D.K., Stewart, J. (2011). PEST hydrology calibration process for source catchments – applied to the Great Barrier Reef, Queensland. 19th International Congress on Modelling and Simulation, Perth, 2359-2366.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the American Society of Agriculture and Biological Engineers*, 50, 885-900.

Petheram, C., McMahon, T.A., Peel, M.C. (2008). Flow characteristics of rivers in northern Australia: Implications for development. *Journal of Hydrology*, 357, 93-111.

Post, D.A. (2009). Regionalizing rainfall-runoff model parameters to predict the daily streamflow of ungauged catchments in the dry tropics. *Hydrology Research*, 40, 433-444.

Prosser, I.P., Rutherfurd, I.D., Olley, J.M., Young, W.J., Wallbrink, P.J., Moran, C.J. (2001). Large-scale patterns of erosion and sediment transport in river networks, with examples from Australia. *Marine and Freshwater Research*, 52, 81-99.

Rustomji, P., Caitcheon, G., Hairsine, P. (2008). Combining a spatial model with geochemical tracers and river station data to construct a catchment sediment budget. *Water Resource Research*, 44, W01422.

Surian, N., Andrews, E.D. (1999). Estimation of geomorphically significant flows in alpine streams of the Rocky Mountains, Colorado (USA). *Regulated Rivers: Research & Management*, 15, 273-288.

Wilkinson, S.N., Young, W.J., DeRose, R.C., 2006. Regionalising mean annual flow and daily flow variability for sediment transport modelling. *Hydrological Processes*, 20, 2769-2786.

Wilkinson, S.N., Prosser, I.P., Rustomji, P., Read, A.M. (2009). Modelling and testing spatially distributed sediment budgets to relate erosion processes to sediment yields. *Environmental Modelling and Software*, 24, 489-501.

Zhang, Y.Q. & Chiew F.H.S. (2009). Evaluation of regionalisation methods for predicting runoff in ungauged catchments in southeast Australia, 18th World IMACS / MODSIM Congress, Cairns, http://mssanz.org.au/modsim09: 3442-3448.