

Incorporating vegetation time series to improve rainfall-runoff model predictions in gauged and ungauged catchments

Y.Q. Zhang^a, J. Vaze^a, F.H.S Chiew^a, and Y. Liu^b

^a CSIRO Water for a Healthy Country National Research Flagship, CSIRO Land and Water, P.O. Box 1666, Clunies Ross Street, Canberra, ACT 2601

^b School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia
Email: yongqiang.zhang@csiro.au

Abstract: Conceptual lumped rainfall-runoff models are widely used for surface runoff predictions in gauged and ungauged catchments because they are simple, easy to calibrate, and give relatively accurate runoff predictions. Lumped rainfall-runoff model inputs are generally daily precipitation and potential evapotranspiration (or air temperature). Vegetation processes are seldom considered in these models but they can play an important role in controlling runoff production in mid-latitude catchments. The use of surface vegetation information in rainfall-runoff modeling may allow better estimation of water balance components, evapotranspiration and water storage change resulting in more accurate estimates of runoff.

Most studies calibrate the rainfall-runoff models against gauged streamflow data and use regionalisation methods to specify parameter values to model runoff in the ungauged catchments. This study investigates the potential to improve runoff and soil moisture prediction by incorporating vegetation time series data into lumped rainfall-runoff modelling. The modelling experiments are carried out using a daily conceptual lumped rainfall-runoff model SIMHYD. Daily rainfall, meteorological and streamflow data, NOAA-AVHRR monthly remotely sensed leaf area and TRMM-TMI daily microwave surface soil moisture for 470 unregulated catchments (50-5,000 km²) across Australia over the period of 1981 to 2006 are used.

The SIMHYD model is adapted to incorporate leaf area index series and land cover types by modifying the evapotranspiration sub-model (called SIMHYD-ET), with an additional one parameter for SIMHYD (total of 10 parameters). The original and modified versions of the SIMHYD models are then calibrated against daily streamflow in each of the 470 catchments. The model's ability to predict runoff and soil moisture in 'ungauged' catchments is then assessed by using parameter values from the geographically closest gauged catchment. The calibration and prediction results of the SIMHYD and SIMHYD-ET models are then evaluated using the Nash-Sutcliffe Efficiency (NSE) of daily runoff, Water Balance Errors (WBE) percentage and correlation coefficient between modeled daily soil moisture and TRMM-TMI soil moisture.

The modelling results indicate that the daily runoff series and total runoff volume modelled by the SIMHYD-ET model are similar to (or only very marginally better than) those simulated by the original SIMHYD model. The SIMHYD-ET model, however, performs noticeably better than the SIMHYD model in soil moisture predictions for both gauged and ungauged catchments. It is possible that better prediction skills can be achieved by modifying the lumped SIMHYD-ET model into a gridded model to take advantage of gridded/spatial rainfall and remote sensing data (leaf area index and land cover types) inputs.

Keywords: *Rainfall-runoff models, vegetation, leaf area index, runoff, soil moisture, PUB*

1. INTRODUCTION

Conceptual lumped rainfall-runoff models are widely used for surface runoff predictions in poorly gauged or ungauged catchments because they are simple, easy to calibrate, and relatively accurate in runoff predictions. Lumped rainfall-runoff model inputs are generally precipitation and potential evapotranspiration (or air temperature). Vegetation processes are seldom considered in these models but they can play an important role in controlling runoff production for mid-latitude catchments (Yildiz and Barros, 2007; Zhang *et al.*, 2009). Due to lack of surface vegetation information in rainfall-runoff modeling inputs, calibrated lumped rainfall-runoff models may not estimate water balance components, evapotranspiration and water storage change accurately and this possibly limits their ability to estimate runoff in ungauged catchments.

Remotely sensed data provide temporally dynamic and spatially explicit information on land surface characteristics, such as fractional vegetation cover and leaf area index (LAI) which is defined as the ratio of total upper leaf surface of vegetation divided by the surface area of land on which the vegetation grows. Remote sensing fractional vegetation cover and LAI data have been widely used in distributed hydrological models (Andersen *et al.*, 2002; Garcia-Quijano and Barros, 2005; Yildiz and Barros, 2007; Zhang *et al.*, 2009).

Zhang *et al.* (2008, 2010) used the Terra MODerate resolution Imaging Spectrometer mounted on the polar-orbiting Terra satellite (MODIS) - LAI with the Penman-Monteith-Leuning (PML) model to estimate 8-day land surface actual evapotranspiration across Australian continent at 5-km resolution for the period of 2000 to 2008. Their results show that (1) the PML model can be easily calibrated using catchment long-term water balances (precipitation minus runoff); (2) the actual evapotranspiration estimated from the PML model compares well with measurements at four eddy covariance flux sites and is better than the estimates from a daily lumped rainfall runoff model - SIMHYD.

Based on the finding of Zhang *et al.*, (2008), Zhang *et al.* (2009) modified the SIMHYD model by incorporating MODIS LAI time series and land cover types. The original and modified models were calibrated in 120 catchments in south-east Australia for the period 2001 to 2005, and then the modeling results for ungauged catchments are assessed using the optimised parameter values from the geographically nearest gauged catchments. The results showed that the modified SIMHYD model that used the MODIS-LAI perform slightly better than the original model, in terms of prediction of daily and monthly runoff.

This paper complements the work of Zhang *et al.* (2009) by using larger and longer-period dataset for the original and modified SIMHYD models, including

- (1) use of 470 unregulated catchments widely distributed across Australia (Figure 1);
- (2) use of NOAA-AVHRR LAI data which cover the period of 1981 to 2006;
- (3) use of micro-wave soil moisture data to evaluate the performance of models in simulating soil moisture.

Specifically, we investigate whether incorporation of NOAA-LAI time series data into the modified SIMHYD model can benefit: (1) daily runoff estimation for ungauged catchments; (2) daily soil moisture estimation for gauged and ungauged catchments.

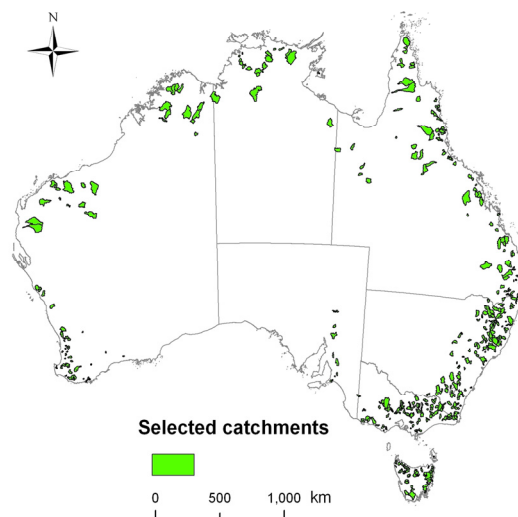


Figure 1. Spatial coverage of the 470 unregulated catchments across Australia.

2. METHODS AND DATA

2.1. Original and modified SIMHYD models

SIMHYD is a lumped conceptual daily rainfall-runoff model. The inputs into the model are daily rainfall and daily potential ET (ET_p), and the model estimates daily runoff.

The version of the SIMHYD model used here has 9 parameters. The structure of the SIMHYD model and the model parameters and algorithms that describe water movement into and out of the storages are shown in Figure 2. SIMHYD has been extensively used for various applications across Australia (Chiew, *et al.*, 2009; Vaze, *et al.*, 2010; Zhang, *et al.*, 2008).

To use RS-LAI data in the SIMHYD model, the ET sub-model is replaced with the PML model to calculate actual ET directly (see Figure 2).

The Penman-Monteith (PM) equation can be written as:

$$ET = \frac{1}{\lambda} \frac{\Delta A_e + \rho_a C_p D G_a}{\Delta + \gamma(1 + G_a/G_s)} \quad (1)$$

where λ is the latent heat of vaporization, $\Delta = de^*/dT_a$ is the slope of the curve relating saturation water vapour pressure to temperature, $D = e^*(T_a) - e_a$ is the vapour pressure deficit of the air, $e^*(T_a)$ is the saturation vapour pressure at air temperature, e_a is the actual vapour pressure, γ is the psychrometric constant, ρ_a is the air density, C_p is the specific heat capacity of air, A_e is the available energy, the difference of the net radiation to the soil heat flux (assumed to be zero here), G_a is the aerodynamic conductance and G_s is the surface conductance.

The terms A_e , Δ , γ , ρ_a and D in Eq. (1) can be calculated from the basic daily meteorological time series and the term G_a in Eq. (1) can be calculated from land cover data (Zhang, *et al.*, 2008).

The surface conductance, G_s , is the only physiological variable in the PM equation. It is calculated using the algebraic, biophysical two-parameter surface conductance model, the Leuning model (Leuning, *et al.*, 2008). The daily input data required for the model are LAI and basic meteorological variables. The model has two parameters, the maximum stomatal conductance g_{sx} and the fraction of equilibrium evaporation at the soil surface f . The soil evaporation factor f is directly dependent on moisture status, and the soil wetness modelled by the SIMHYD model is used as the estimate for f . The g_{sx} term is considered as a parameter that is optimized together with the other SIMHYD model parameters.

The modified SIMHYD model has 10 parameters (one additional parameter, g_{sx}). The ET in the evapotranspiration submodel is calculated using the PM model. The parameter, f , in the PM equation is calculated as the soil wetness, the ratio of soil moisture storage (SMS) to soil moisture storage capacity (SMSC) (Figure 2). To distinguish between the revised and original models, the revised model will be referred to as 'SIMHYD-ET'.

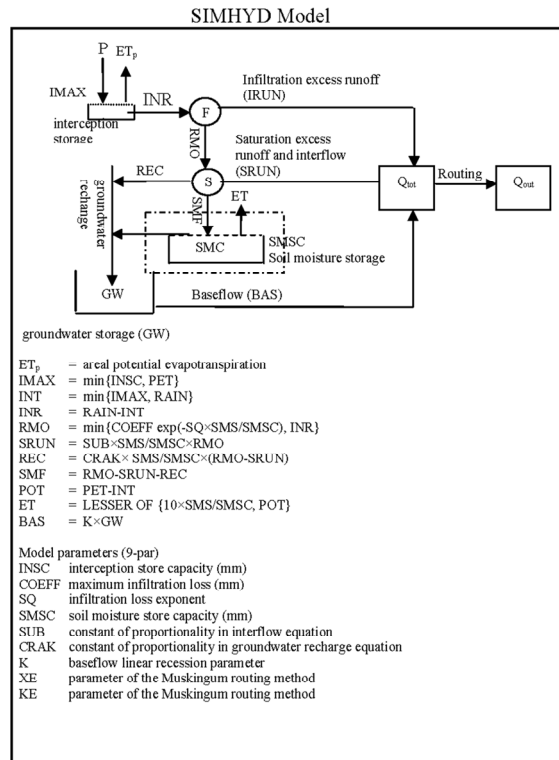


Figure 2. Model structure of SIMHYD and its revised version (the dash-dots show the evapotranspiration submodel, which is modified).

2.2. Data

2.2.1. Streamflow data

Daily time series of streamflow data for the selected 470 unregulated catchments (50 to 5000 km²) come from the respective state water agencies. The streamflow data used here are quality assessed using quality codes and spike control methods (Zhang *et al.*, 2011). Data from 1981 to 2006 are used in this study.

2.2.2. Meteorological data

Daily time series of maximum temperature, minimum temperature, incoming solar radiation, actual vapour pressure and precipitation from 1981 to 2006 at 0.05° × 0.05° (~ 5 km × 5 km) grid cells from the SILO Data Drill of the Queensland Department of Natural Resources and Water (www.nrw.gov.au/silo) are used. The SILO Data Drill provides surfaces of daily rainfall and other climate data interpolated from point measurements made by the Australian Bureau of Meteorology. The 0.05° × 0.05° spatial data is averaged across all the grid cells within a catchment to produce a catchment average timeseries for use in this study. The rainfall data is required as input to the rainfall-runoff models. The other meteorological data are used to calculate ET_p using the Priestley-Taylor model and to calculate PM-ET for the revised rainfall-runoff models.

2.2.3. Remote sensing data

Daily time series of TRMM-TMI top soil moisture content data from 2000 to 2006 at 0.25° × 0.25° (~ 25 km × 25 km) are used (Liu *et al.*, 2007). The data were retrieved using the land parameter retrieval model and X-band brightness temperature. The retrieved soil moisture represents an estimate of the moisture content of roughly the top 2-5 centimeters of soil.

The RS-LAI data required to calculate PML-ET in the revised rainfall-runoff model are NOAA-AVHRR monthly leaf area index data at ~8-km resolution, obtained from Boston University.

Land cover data required to estimate G_a in Eq. (1) are obtained from the MODIS land cover product, the yearly Land Cover classification product (MOD12Q1) (<http://edcdaac.usgs.gov/modis/mod12q1v4.asp>). The dataset has 17 vegetation classes defined according to the International Geosphere-Biosphere Programme.

The albedo data required to calculate A_c in Eq. (1) are obtained from 8 day MODIS MCD43B bidirectional reflectance distribution function product at 1-km resolution.

All the remote sensing and meteorological data are reprojected and re-sampled to obtain 1-km gridded data. The gridded data in each catchment are then extracted and averaged to obtain aggregate daily data series for use in the modelling.

2.3. Model calibration

The two models are calibrated against daily runoff data from 1981 to 2006, with the first two-year (1981-1982) data used for model warm up.

A widely used global optimization method, the genetic algorithm, is used to optimise the parameters in the two models by minimising the objective function (Obj)

$$Obj = 1 - NSE + 0.1WBE \quad (2)$$

$$NSE = 1 - \frac{\sum_{i=1}^N (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^N (Q_{obs,i} - \overline{Q_{obs}})^2} \quad (3)$$

$$WBE = \left| \frac{\sum_{i=1}^N Q_{sim,i} - \sum_{i=1}^N Q_{obs,i}}{\sum_{i=1}^N Q_{obs,i}} \right| \quad (4)$$

where NSE is the Nash-Sutcliffe Efficiency of daily runoff (Nash & Sutcliffe, 1970), WBE is the absolute Water Balance Error, Q_{sim} and Q_{obs} are the simulated and observed daily runoff, respectively, $\overline{Q_{obs}}$ is the arithmetic mean of the observed daily runoff, i is the i th days and N is the total days sampled. The Equation 2 uses two criteria, Nash-Sutcliffe efficiency and water balance errors, to optimise model parameters. Obj is

close to 0 for parameters which give both a good agreement between the computed and recorded runoff, and a small volume error. Lindström (1997) found that $0.1WBE$ normally gives a good compromise between NSE and WBE.

NSE measures the agreement between the modelled and observed daily values, with $NSE = 1.0$ indicating perfect agreement between the modelled and observed daily runoff for a given catchment. *WBE* measures the water balance error between modelled and observed mean annual values, with $WBE = 0$ indicating a perfect agreement.

2.4. Predictions in ungauged catchments

A widely used regionalisation approach, the nearest neighbour approach, is used for predictions in ungauged catchments. Each of the 470 catchments is left out in turn and considered as an ' ungauged ' catchment, and the entire set of parameter values from the nearest donor catchment are used to model daily runoff and soil moisture in these ' ungauged ' catchments.

2.5. Evaluation criteria

Three criteria are used to evaluate model calibration and prediction results, NSE of daily runoff, WBE percentage and correlation coefficient (R) between modelled daily soil moisture and TRMM-TMI soil moisture. It is noted that the TRMM-TMI soil moisture data represent soil moisture storage in the top soil layer (2 ~ 5 cm) while modelled soil moisture data are for the whole soil column. Nevertheless, the correlation coefficient provides an indicator to evaluate modelled soil moisture.

It is also noted that the TRMM-TMI soil moisture data had large gaps (missing data) in some of the 470 catchment. As such, 301 catchments where the TRMM-TMI soil moisture data length is more than one year are selected for soil moisture evaluation.

3. RESULTS AND DISCUSSION

3.1. Calibration results

The model calibration results are summarised in Figures 3 and 4 and Table 1. The calibrations for SIMHYD and SIMHYD-ET models are satisfactory with NSE of daily runoff values greater than 0.65 and WBE percentage values less than 5% in about 50 percent of the 470 catchments. The calibration results for streamflow are similar to most rainfall-runoff modelling studies for Australian catchments (Post and Jakeman, 1999; Viney *et al.*, 2008). The simulated soil moisture from the calibrated models shows a reasonable agreement with the microwave soil moisture, with R values greater than 0.45 in about 50% of the catchments.

Table 1. Statistical summary of model calibration and prediction results for SIMHYD and SIMHYD-ET models

Criteria	Calibration/ Prediction	Model	Percentile				
			10	25	50	75	90
NSE	Calibration	SIMHYD	0.41	0.55	0.65	0.74	0.80
		SIMHYD-ET	0.46	0.56	0.67	0.75	0.81
	Prediction	SIMHYD	-0.03	0.31	0.50	0.62	0.71
		SIMHYD-ET	-0.02	0.33	0.51	0.64	0.71
WBE	Calibration	SIMHYD	0.1	0.6	4.9	17.9	33.6
		SIMHYD-ET	0.0	0.3	3.9	11.7	24.4
	Prediction	SIMHYD	4.6	9.7	23.3	41.7	82.2
		SIMHYD-ET	3.8	8.8	22.2	40.7	79.4
R	Calibration	SIMHYD	0.17	0.30	0.46	0.63	0.68
		SIMHYD-ET	0.25	0.37	0.49	0.64	0.69
	Prediction	SIMHYD	0.16	0.30	0.45	0.63	0.69
		SIMHYD-ET	0.24	0.37	0.49	0.64	0.69

Comparison between SIMHYD and SIMHYD-ET results show that the calibration results for SIMHYD-ET model are slightly better than those for SIMHYD model. This is indicated by (1) NSE of daily runoff for SIMHYD-ET model is about 0.01-0.05 higher than that for SIMHYD model, (2) WBE percentage for SIMHYD-ET model is about 0-9% lower than that for SIMHYD model, and (3) R of soil moisture for SIMHYD-ET model is about 0.01-0.08 higher than that for SIMHYD model. The SIMHYD-ET model results are better than the SIMHYD results especially for poorly simulated catchments where low NSE of daily runoff, low R of soil moisture and high WBE values are observed (Figures 3 and 4).

3.2. Predictions of daily runoff

The model prediction (cross-validation) results are summarised in Figures 3 and 4 and Table 1 as well. The SIMHYD-ET model gives a slightly better daily runoff simulation than the SIMHYD model because it incorporates LAI time series and land cover type data. The NSE values for the SIMHYD-ET model are about 0.01-0.02 higher than those for the SIMHYD model. The WBE percentage values for the SIMHYD-ET model are about 0.1-3% lower than those for the SIMHYD model.

Comparison between model calibration and model prediction shows that use of LAI time series improves estimate accuracy of daily runoff more significantly in model calibration than model prediction.

3.3. Predictions of daily soil moisture

Figure 5 and Table 1 summarise model predictions of daily soil moisture. For both models, soil moisture obtained from model calibration performs similar to that obtained from model prediction. The SIMHYD-ET model simulates soil moisture noticeably better than the SIMHYD model. The improvement in R of daily soil moisture is about 0.07 at the 25 percentile or less, and is about 0.04 at 25-50 percentiles. But, it is not significant above the 60 percentile. The further analysis shows that the improvements are mainly observed in south-east Australian forest catchments, indicating that it is more possible to improve soil moisture predictions in forest catchments using the SIMHYD-ET model (data not shown).

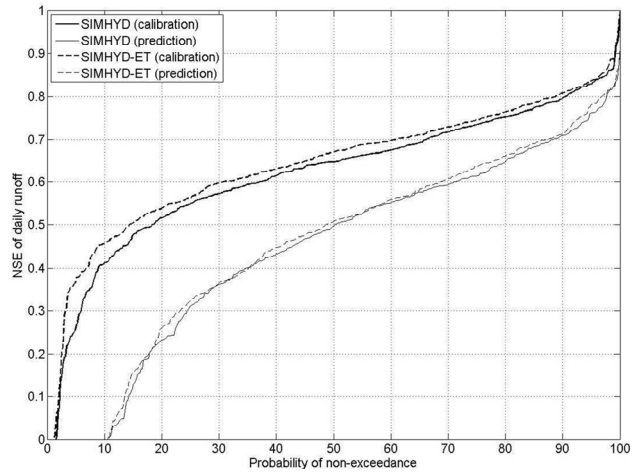


Figure 3. Summary of calibrated and validated Nash-Sutcliffe Efficiency (NSE) of daily runoff for SIMHYD and SIMHYD-ET models of the 470 catchments. Large NSE values indicate a good model performance.

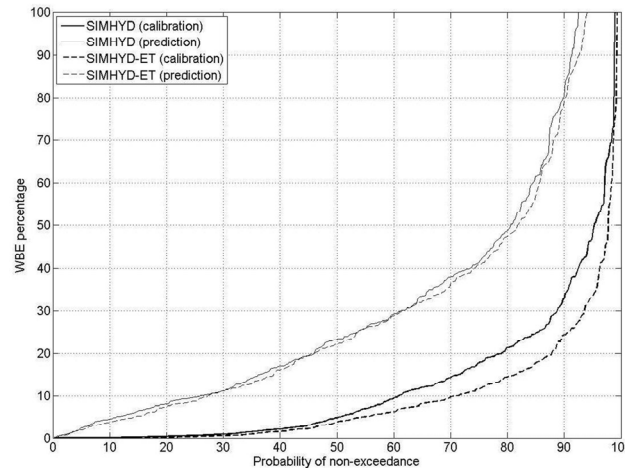


Figure 4. Summary of calibrated and validated Water balance Errors (WBE) percentage for SIMHYD and SIMHYD-ET models of the 470 catchments. Small WBE values indicate a good model performance.

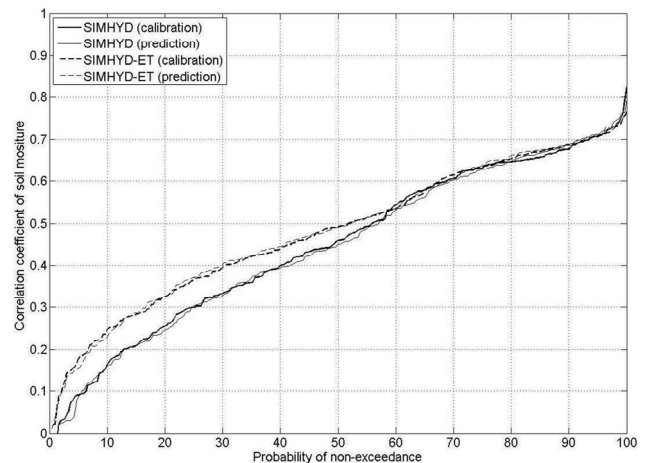


Figure 5. Summary of predicted correlation coefficient of daily soil moisture for SIMHYD and SIMHYD-ET models of the 301 catchments. Large correlation coefficient values indicate a good model performance.

4. CONCLUDING REMARKS

This study explores the benefits of incorporating remote sensing leaf area index time series and land cover type data into a traditional rainfall runoff model, SIMHYD. The model calibration and prediction results for 470 unregulated Australian catchments show that for both model calibration and prediction the modified model incorporating remote sensing leaf area index performs slightly better than the original model, in terms of higher NSE of daily runoff and lower WBE. It, however, performs noticeably better for soil moisture predictions for both gauged and ungauged catchments.

This is an ongoing study. The lumped SIMHYD-ET model is being modified into a gridded model to take advantage of gridded remote sensing data (leaf area index and land cover types) inputs. In the current version of SIMHYD-ET model, only one parameter value is given to the maximum stomatal conductance. The gridded SIMHYD-ET model treats the parameter separately for different land cover types. It is expected that the gridded model can obtain more benefit of use of remote sensing leaf area index time series and land cover type data.

ACKNOWLEDGMENTS

This study is carried out in the CSIRO Water for Healthy Country National Research Flagship and is supported by two CSIRO projects: the eWater CRC Catchment Water Yield Estimation Tools (CWYET) project and the Water Information Research and Development Alliance (WIRADA) between CSIRO and the Australian Bureau of Meteorology.

REFERENCES

- Andersen, J., Dybkjaer, G., Jensen, K. H., Refsgaard, J. C., and Rasmussen, K. (2002). Use of remotely sensed precipitation and leaf area index in a distributed hydrological model. *Journal of Hydrology*, 264, 34-50.
- Chiew, F.H.S., Teng, J., Vaze, J., Post, D.A., Perraud, J.M., Kirono, D.G.C., Viney, N.R., 2009: Estimating climate change impact on runoff across southeast Australia: Method, results, and implications of the modeling method. *Water Resources Research*, 45.
- Garcia-Quijano, J. F. and Barros, A. P. (2005). Incorporating canopy physiology into a hydrological model: photosynthesis, dynamic respiration, and stomatal sensitivity. *Ecological Modelling*, 185, 29-49.
- Lindström, G. (1997). A simple automatic calibration routine for theHBV model. *Nordic Hydrology*, 28, 153-168.
- Liu, Y., de Jeu, R. A. M., van Dijk, A., and Owe, M. (2007). TRMM-TMI satellite observed soil moisture and vegetation density (1998-2005) show strong connection with El Nino in eastern Australia. *Geophysical Research Letters*, L15401,10.1029/2007gl030311.
- Nash, J. E. and Sutcliffe, J. V. (1970). River forecasting using conceptual models, 1. A discussion of principles. *Journal of Hydrology*, 10, 280-290.
- Post, D. A. and Jakeman, A. J. (1999). Predicting the daily streamflow of ungauged catchments in SE Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model. *Ecological Modelling*, 123, 91-104.
- Vaze, J., F. H. S. Chiew, J. M. Perraud, N. Viney, D. A. Post, J. Teng, B. D. Wang, J. Lerat, and M. Goswami (2011). Rainfall-runoff modelling across southeast Australia: datasets, models and results. *Australian Journal of Water Resources*, 14, 101-116.
- Viney, N., Vaze, J., Chiew, F., and Perraud, J. 2008: Regionalisation of runoff generation across the Murray-Darling Basin using an ensemble of two rainfall-runoff models. *Water Down Under 2008, April 2008*, Adelaide: Engineers Australia, 1700-1711.
- Yildiz, O. and Barros, A. P. (2007). Elucidating vegetation controls on the hydroclimatology of a mid-latitude basin. *Journal of Hydrology*, 333, 431-448.
- Zhang, Y. Q., Chiew, F. H. S., Zhang, L., and Li, H. X. (2009). Use of remotely sensed actual evapotranspiration to improve rainfall-runoff modelling in southeast Australia. *Journal of Hydrometeorology*, DOI: 10.1175/2009JHM1061.1.
- Zhang, Y. Q., Chiew, F. H. S., Zhang, L., Leuning, R., and Cleugh, H. A. (2008). Estimating catchment evaporation and runoff using MODIS leaf area index and the Penman-Monteith Equation. *Water Resources Research*, doi:10.1029/2007WR006563.
- Zhang, Y., Leuning, R., Hutley, L. B., Beringer, J., McHugh, I., and Walker, J. P. (2010). Using long-term water balances to parameterize surface conductances and calculate evaporation at 0.05 degrees spatial resolution. *Water Resources Research*, W05512,10.1029/2009wr008716.
- Zhang, Y.Q., Viney, N., Chen, Y., Li, H. Y. 2011. Collation of streamflow data for 719 unregulated Australian catchments. CSIRO: Water for a Healthy Country National Research Flagship, 67pp. (in review)