Developing Similarity Measures for Predicting Ungauged Streamflow within a Model Averaging Framework

Reichl, J.P.C.[1,3], A.W. Western[1,3] and F.H.S. Chiew[2,3].

[1] Department of Civil and Environmental Engineering, The University of Melbourne, Victoria
[2] CSIRO Land and Water, Canberra, Australia
[3] eWater CRC, Australia
Email: j.reichl@civenv.unimelb.edu.au

Keywords: *Ungauged, streamflow, model averaging, similarity, SimHyd.*

EXTENDED ABSTRACT

The problem of estimating runoff in ungauged catchments remains an important but elusive one. Previous studies suggest that there is an important property common to rainfall-runoff models: over-parameterisation, leading to parameter covariance and the existence of multiple parameter sets which reproduce the streamflow adequately. This reduces parameter identifiability, impeding identification of relationships between model parameter values and catchment characteristics that would otherwise be useful for regionalisation. This study continues the development of a model averaging framework to circumvent this problem for application in Australia.

The model averaging framework is based on a selection of gauged catchments which fall within some threshold of "similarity" to a target ungauged catchment. A number of streamflow predictions are generated, using forcing data for the ungauged catchment and the parameter sets from the "similar" gauged catchments. These streamflow predictions are then combined, via a weighting based both on each catchment's physical similarity to the target ungauged catchment, and on each catchment's calibration quality. This combination is the predicted streamflow time series for the ungauged catchment.

The aim of this study is to inform development of multiple-attribute similarity measures. A number of catchment attributes are collected, which are considered by the hydrological community to be important in catchment-scale rainfall-runoff processes. The 29 attributes fall into four broad groups: Geomorphic; Soil; Climatic; and Vegetation. The results of model averaging experiments are examined, using each of the attributes individually as indicators of catchment similarity. Ungauged prediction results are presented here in terms of the Nash-Sutcliffe Efficiency (E). The correlation matrix between each of the attributes is used to reduce the list of catchment attributes as far as possible from the original 29. The purpose of this is to attempt to remove a double-dipping effect in the more complex similarity measures. For example, the use of mean winter precipitation in addition to mean annual precipitation, usually strongly correlated, would not add much useful information, and if anything would 'smear' the results and make them difficult to interpret. Strongly correlated attributes are eliminated from the study based on the E value they return in the model averaging runs.

The study is carried out using the conceptual daily rainfall-runoff model SimHyd, calibrated to a selection of 95 catchments across Australia.

A very small deterioration is seen when increasing the number of contributing catchments; however it is minor and does not affect the conclusions.

Regional effects are also discussed: when using catchments from the whole of Australia, 12 attributes returned median E values greater than the 0.42 for random catchment selection (considered uninformative and the basis for comparison). In the Köppen Climate Type Cfa region 27 attributes exceeded the random value of 0.46, and in the Type Cfb region 23 attributes exceeded the random value of 0.31. These differences justify the development of separate similarity measures for the three regions.

Deterioration from calibration remains significant for each attribute, however, based on the results of previous studies (Reichl *et al.* 2006, McIntyre *et al.* 2005) it is expected that when used to develop multiple-attribute similarity metrics the model averaging method will be able to provide good estimates of ungauged streamflow.

1. INTRODUCTION

The ability to estimate streamflow time series in ungauged catchments is important to natural resource management, but has to date been elusive. Attempts at applying 'regional' information to infer ungauged hydrological behaviour have met with little success. Recent focus has been on finding correlations between conceptual rainfall-runoff model parameters and catchment physical characteristics, be they measured or estimated (Chiew and Siriwardena, 2005; Merz and Bloschl, 2004). This approach (referred to as 'parameter regression') has been thwarted, largely due to the many sources of error and uncertainty in the modelling of complex natural systems, and the implications of this for estimation of meaningful parameter values. Errors typically arise from input data, model structure, parameter choice and observed output data. Together these uncertainties combine to hamper attempts at predicting ungauged streamflow.

This paper continues the development of a model averaging approach to the problem. Rather than searching for a single relationship between optimal parameter values and catchment physical characteristics, the model averaging approach relates gauged catchments to ungauged based on measures of physical similarity (McIntyre *et al.*, 2005; Reichl *et al.*, 2006). This paper investigates the usefulness of catchment attributes in assessing catchment similarity within the model averaging context. The implication of the selection is that these attributes can be considered important in controlling catchment-scale rainfall-runoff processes. The study is designed to inform the next stage of the development of the model averaging framework, namely the development of multiple-attribute spaces, enabling reliable assessment of catchment similarity to be used within this framework.

2. UNGAUGED STREAMFLOW PREDICTION

Various approaches have been used to predict hydrological behaviour in ungauged catchments. Most are based on the idea of common physical properties. Merz and Bloschl (2004) conducted a thorough study of 308 catchments in Austria, comparing 8 regionalisation techniques, including the use of parameter sets from the closest upstream and downstream catchments, a parameter regression approach and the use of a 'global' parameter set. They found that, "apparently, spatial proximity is a better surrogate of unknown controls on runoff dynamics than catchment characteristics", since the upstream/downstream approach performed significantly better than the parameter regression approach.

Peel *et al*. (2000) in a study of 331 catchments in Australia found statistically significant (but not strong) correlations between most SimHyd model parameters and catchment characteristics. Chiew and Siriwardena (2005) investigated the potential for regionalisation of SimHyd model parameters based on these correlations, and found that the simulations were not consistently better than those using parameter values from the nearest gauged catchment.

Vogel (2005), in addition to a thorough review of regionalisation techniques, described a variation on the regression approach, whereby the calibration of individual catchments aimed at optimising local streamflow reproduction while at the same time optimising the regional relationships between parameter values and catchment characteristics. This approach was also unsuccessful.

McIntyre *et al.* (2005) and Reichl *et al.* (2006) describe an alternate approach to the problem of regionalisation. This approach uses intact parameter sets from gauged catchments to infer hydrological behaviour at ungauged sites. Each gauged catchment is assigned a prior likelihood of accurately predicting streamflow in the ungauged catchment, based on physical similarity. This allows for retention of information about parameter inter-relationships contained in an intact parameter set. This approach shows some success, achieving regionalisation results which are far better than the regression approach.

3. SIMILARITY

A review of the literature produced a list of catchment attributes that are considered important by the hydrological community. These attributes, in addition to some considered important (and convenient to collect) by the authors (Table 1), are used individually as similarity measures in model averaging experiments.

In Table 1, Cos of the Aspect (CosA) is used to avoid wrap-around effects; ERR is the Elevation-Relief Ratio (equal to the hypsometric integral); X and Y are projections onto the x-y plane of the unit normal vector to the surface; X-Y is a combination of the two; Links is the number of network links (1:250k blue lines map); MWP, MSP and MAP are the mean winter (April-September); summer (October-March) and annual precipitation; Aridity is the Mean Annual Areal Potential Evapotranspiration divided by the MAP; Fr refers to fraction; PAWHC is plant-available waterholding capacity; and A_KSAT is the A-horizon saturated hydraulic conductivity.

Table 1. Attributes to be used in development of similarity measures

Geomorphic	Climatic	Vegetation	Soil
MeanCosA	MWP	FrNativeWoody	SoilDepth
StDevCosA	MSP	FrPlantation	PAWHC
MinElevation	MAP	FrOrchard	A_KSAT
MaxElevation	Aridity	FrWoody	Transmissivity
MeanElevation			
ERR			
MeanSlope			
StDevSlope			
MeanX			
StDevX			
MeanY			
StDevY			
Links			
Area			
LinkDensity			
MeanX-Y			
Proximity			

4. MODEL AVERAGING STUDY

4.1. Data

The precipitation, potential evapotranspiration (PET) and streamflow data used for this study are a subset of the NLWRA dataset (Peel *et al.*, 2000). Since the full set of NLWRA data are of varied lengths and periods a common period was sought in order to eliminate potentially confounding effects of using disparate time periods, for example differences in climate signals from catchments in close proximity. Following the report of Jakeman *et al.* (1993) on the data requirements of models of varying degrees of parameterisation, it is considered that 10-15 years of monthly data are adequate for determination of SimHyd's parameters, allowing for significant instances of missing data. A requirement of this data period is that it contain, for the majority of catchments, both wet and dry periods, in order to adequately exercise the model parameters (Gupta *et al.*, 1983). The period 1972 – 1985 fulfils this requirement, whilst maximising the number of available catchments. This reduction comes at the risk of a loss of generality of conclusions, however the authors consider this to be a small risk, given that (i) the spatial extent of the reduced lists are not less than that of the full data set (Figure 1); and (ii) the reduction in number of catchments is not overly significant. The full data set has 329 catchments. This is split into two, one set for development and one for testing of the method (although the Test set is not used in this preliminary study). The Development set is reduced from a possible 165 to 95 and the Test set from 164 to 89. This is the result of both reducing the time period and removing catchments due to inconsistent data (approximately 10 catchments). Figure 2 shows the distributions of the areas of the Development and the Test catchment sets.

The soil data are from McKenzie *et al.* (2000), the vegetation data are from ANZLIC (2005), and the geomorphic data are from ANZLIC (2002).

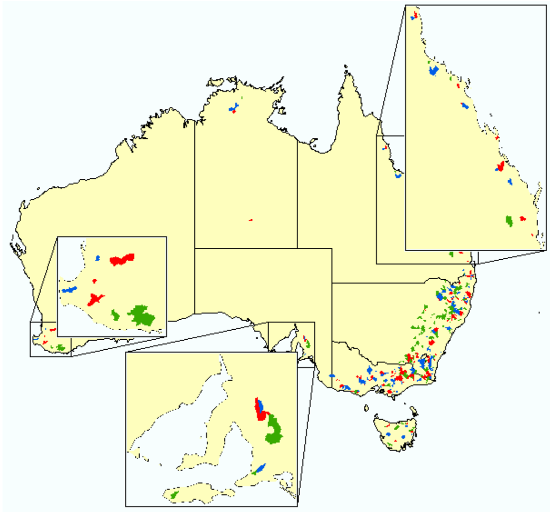

Figure 1. Map of Australia, showing the spatial distributions of catchments from the Development set (red), the Test set (blue) and the full set (green)

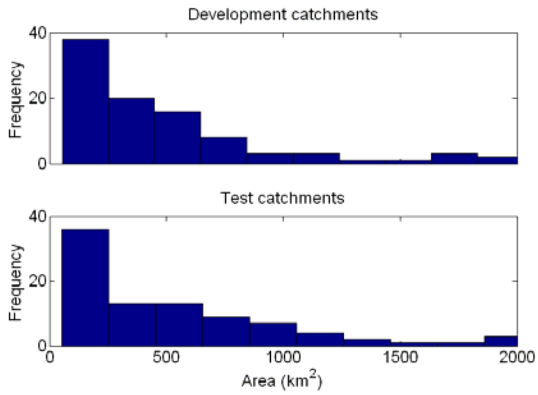

Figure 2. Histograms showing the distribution of the areas of the Development catchments and Test catchments

4.2. SimHyd rainfall-runoff model

SimHyd is a lumped conceptual daily rainfall-runoff model. It is driven by daily precipitation and PET, and simulates daily streamflow and ET. It has been tested and used extensively across Australia. See Chiew *et al.* (2002) for a full description of the seven model parameters and the model algorithms. SimHyd is calibrated against monthly streamflow using a quasi-Newtonian optimisation routine. The standard 3-way bootstrap cross-validation was attempted, however many of the validation results were quite poor, leading to the conclusion that 2/3rds (9 or 10 years) of the available period is not enough to adequately calibrate the model. This conclusion was confirmed when the second attempt at validation was carried out, this time using the full period for calibration and a separate but contiguous 10 year period for validation. This time the results were more satisfactory. Apart from a few outliers, validation deterioration is quite small.

4.3. Model Averaging

In this study each catchment is treated as being ungauged in turn, with information from the other 94 catchments used to infer hydrological behaviour.

McIntyre *et al.* (2005) and Reichl *et al.* (2006) demonstrated the potential for using the model averaging framework for estimating ungauged streamflow. The framework involves these 5 steps: similarity definition; streamflow time series generation; weighting; averaging; and assessment.

Similarity:

The similarity of each catchment to other available catchments is assessed in terms of some metric. In this study this will simply be individual catchment attributes, however in future work these will be combined to form some multiple-attribute space, such as described by the Euclidean Distance Metric (Equation 1):

$$D_{i,j} = \left[\sum_{a=1}^{N} w_a \left(A_{a,i} - A_{a,j} \right)^2 \right]^{1/2} \qquad (1)$$

where $D_{i,j}$ is the dissimilarity of catchment j to catchment i, $A_{a,j}$ is the value for attribute a of N for catchment j, and w_a is some weight associated with attribute a. Note that N is equal to 1 in this preliminary study, but in general is more than one.

Once assessed, the catchments' similarities can be subjected to some form of threshold, whereby those that are most similar to the target ungauged catchment are selected for the following steps.

Streamflow time series generation:

Forcing data for the target catchment are used to generate streamflow time series using model parameter sets from those catchments which fall within the aforementioned threshold.

Weighting:

Those catchments which fall within the similarity threshold are assigned weights. These are based both on the similarity to the target catchment (Equation 2) and the quality of the gauged catchments' calibration (Equation 3):

$$Ws_{i,j} = \frac{1}{\sum_{i=1}^{M} 1.3 - \frac{D_{i,j}}{D_{i,j\max}}} \left[1.3 - \frac{D_{i,j}}{D_{i,j\max}} \right] \qquad (2)$$

where $Ws_{i,j}$ is the similarity weight assigned to the i^{th} gauged catchment, and $D_{i,jmax}$ is the maximum dissimilarity of M catchments falling within the threshold; and

$$Wc_i = \frac{E_i}{\sum_{i=1}^{M} E_i} \qquad (3)$$

where Wc_i is the calibration weight assigned to the i^{th} gauged catchment, and E_i is the Nash-Sutcliffe Efficiency (Equation 5). In Equation 2 the value 1.3 was used instead of 1 so that the least similar of the M selected catchments would not receive a weight of zero. The value 1.3 is a neutral choice, resulting in the differences in weightings from least to most similar catchments being reasonably separated, but not severely.

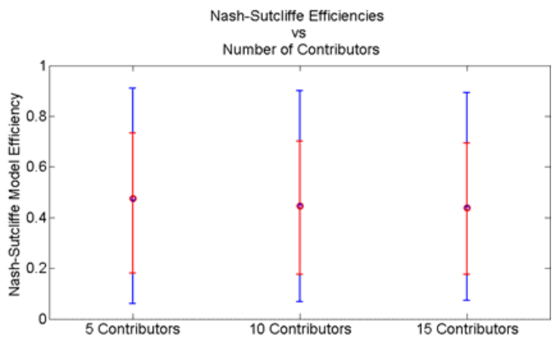

Figure 3. 5^{th}, 25^{th}, 75^{th} and 95^{th} percentiles (bars) and median (marker) E for all attributes, with increasing numbers of contributing catchments.

Averaging:

The streamflow time series are combined to form the estimation for the ungauged catchment (Equation 4):

$$S_{t,j} = \sum_{i=1}^{M} \frac{Ws_{i,j} Wc_{i,j}}{\sum_{i=1}^{M} Ws_{i,j} Wc_{i,j}} h(\theta_i, X_{t,j}) \qquad (4)$$

where S_t is the output streamflow at time t and $h(\theta_i,X_{t,j})$ is the model output given parameters θ and forcing data X.

Assessment:

In this paper the model performance is presented in terms of the Nash-Sutcliffe Model Efficiency (Equation 5):

$$E = \frac{\sum_{t=1}^{n}(Y_t - \overline{Y})^2 - \sum_{t=1}^{n}(h_t - Y_t)^2}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2} \qquad (5)$$

where E is the model efficiency and Y_t is the recorded response data at time step t of n. An E value of 1 indicates that the estimated response is the same as the recorded response for all time steps, and smaller values (negative values are possible) indicate greater disparity between recorded and estimated responses.

5. RESULTS AND DISCUSSION

5.1. Number of Contributing Catchments

A brief investigation into threshold effects was carried out in this study, in order to ascertain how the choice of threshold would affect the conclusions. Three groups of model averaging experiments were carried out, using 5, 10 and 15 contributing catchments for each ungauged catchment. Reichl *et al.* (2006) found a very slight deterioration in performance when the number of contributing catchments was increased, and this is seen again in this study. Figure 3 shows the 5th, 25th, 50th, 75th and 90th percentiles of the Nash-Sutcliffe Efficiency for every catchment for every attribute. There is a slight deterioration in the median when increasing the number of contributing catchments; however, the variance also decreases slightly. It may be that for different applications it would be desirable to decrease the variance at the cost of a slight loss of overall performance. More important for the present study is whether changing the number of contributing catchments affects which attributes are retained for the next stage of the research. There are some changes seen, and, although few and minor, the results of all three groups of experiments are used instead of any one experiment group in determining attribute performance.

5.2. Individual Attribute Performance

The performance of each attribute when used individually in model averaging experiments is shown in Figure 4. All attributes show large variance, with mean and median E values close to 0.5. The distribution of E values for the calibrated catchments is shown for reference.

Of note is that 15 of the attributes have median E values lower than that achieved when catchments are chosen at random, with no weighting ("Random" in Figures 4 and 5).

This result appears to indicate that there are attributes which actively lead to inappropriate catchments, rather than merely being uninformative.

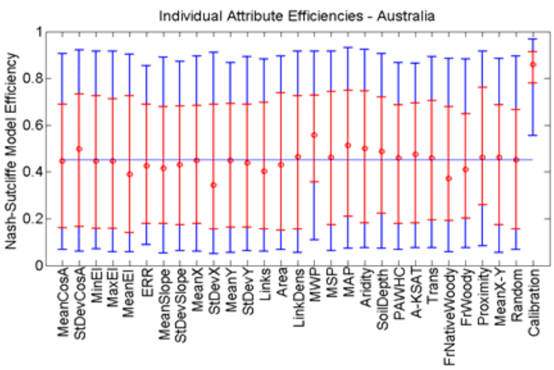

Figure 4. E percentiles for all individual attributes, as well as the random approach and the calibration results

5.3. Regional effects

It is probable that, given the large variation in climate, vegetation, soil and geomorphology across Australia, better results would be attained if similarity measures were developed within specific regions. An obvious way to distinguish regions is to use a Köppen climate map, which delineates regions based on temperature and precipitation.

A recently updated Köppen-Geiger Climate Map (Peel *et al.*, 2007) has been made freely available in digital form. This map is used here to group catchments into climate types.

27 of the 95 catchments fall into the Cfa climate type, 56 into Cfb, 4 into Aw, 1 into each of Am and Cwa, and 6 into Csb. Of these, only the Cfa and Cfb types have enough catchments for meaningful interpretation of a model averaging

experiment. These are temperate, without a dry season, and are distinguished by being hot and warm climates respectively. Tasmania and Victoria are predominantly of Cfa type, and coastal NSW and southern Queensland of Cfb type.

The results of the model averaging experiment using only Cfa-type catchments are, unsurprisingly, far better than those using catchments across the country. Those using the Cfb-type catchments return results of similar quality to those from the whole country, although with different dominant attributes.

The Cfa-type catchments all show far better results than the random approach (Figure 5 top), but for the Cfb-type catchments four attributes performed worse than the random approach (Figure 5 bottom).

One possible reason for the poorer results from the Cfb region is that there is a tail (~25 %) of catchments that did not calibrate well for that region, compared to the Cfa region; another is that among the attributes collected, the important runoff controls for the region were not captured as well.

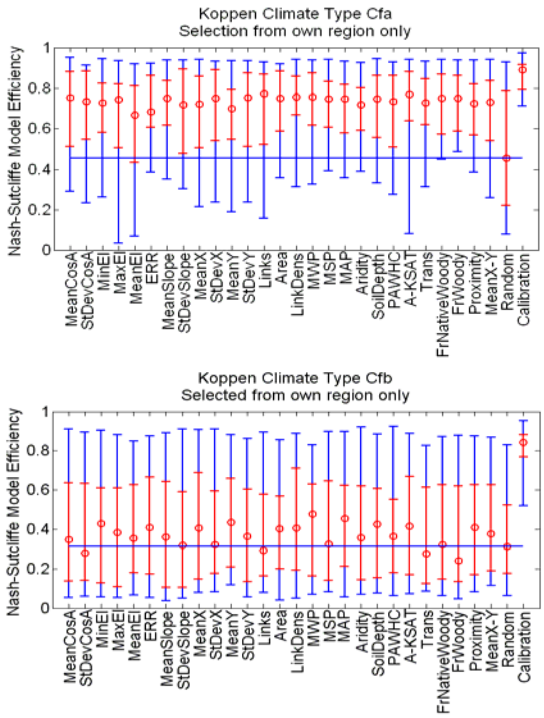

Figure 5. *E* percentiles for Köppen climate type Cfa (top) and Cfb (bottom), for all attributes. The horizontal line is the median Random *E*.

The difference between the results for the country and those for the two climate regions justifies separate development of similarity measures for the three regions.

5.4. Correlation Between Attributes

The list of attributes available for construction of multiple-attribute space must be reduced (Table 2). First, those attributes with poorer performance than the random approach are removed. Next, those attributes which correlate strongly (correlation coefficient>0.5) with another, but which show poorer performance than it, are removed. In order to restrict the computational demands of the problem, the five best performing attributes for each region were chosen of those remaining, plus proximity, since proximity is a commonly good performer for all three regions, and that it is reported in several studies as being as good as more complex regionalisation techniques.

Table 2. Reduced list of attributes for each region, in order of decreasing performance

Australia	Cfa	Cfb
MWP	Links	MWP
StDevCosA	A_KSAT	MeanY
SoilDepth	MWP	MinEl
LinkDens	LinkDens	SoilDepth
MSP	MeanCosA	ERR
Proximity	Proximity	Proximity

The three regions each have remaining at least one climatic, geomorphic and soil attribute. None have a vegetation attribute. Although vegetation is strongly affected by land use and is known to affect runoff (Brown *et al.*, 2005), it is usually strongly correlated to climate and certain geomorphic attributes, since these are important determinants of land use.

MeanCosA, StDevCosA and MeanY are all indicators of slope direction and/or magnitude. These influence both the exposure to the prevailing weather and the amount of radiation the slopes are exposed to, and thus the drying that the catchment will experience.

The density of network links and the soil properties are important in terms of the partitioning of flows, however, timing effects are not important when reproducing monthly data. More important is the influence of the groundwater on the streamflow time series. The elevation-type attributes and the proximity indicate climatic influences on the catchment, and possibly are surrogate indicators of the influence of such things as alluvial groundwater systems.

6. CONCLUSIONS

With the aim of informing selection of predictive attributes for development of multiple-attribute catchment similarity metrics, results of experiments using individual attributes within the model averaging framework are shown. A random experiment is used as the basis against which results are compared. Results of experiments are discussed using catchments from the whole of Australia, and for two Köppen climate types. The differences in results between regions justify development of separate similarity measures for different regions.

A process for selecting attributes for multi-dimensional catchment similarity measures is outlined. First, attributes which perform better than a random approach are retained. Next, attributes which correlate strongly to a better-performing attribute are eliminated. From those remaining, the best 5 attributes plus proximity are retained.

Deterioration from calibration remains significant for each attribute, however, based on previous studies (Reichl *et al.* 2006, McIntyre *et al.* 2005) it is expected that when used to develop multiple-attribute similarity metrics the method will be able to provide good estimates of ungauged streamflow.

7. REFERENCES

ANZLIC, (2002). GEODATA 9 Second Digital Elevation Model (DEM-9S) ANZCW0703005624. *www.ga.gov.au*.

ANZLIC, (2005). Land cover for the intensive use zone of Australia. *www.ga.gov.au*.

Brown, A.E., L. Zhang, T.A. McMahon, A.W. Western and R.A. Vertessy, (2005). A review of paired catchment studies with reference to seasonal flows and climatic variability. *Journal of Hydrology*, 310(1-4): 28-61.

Chiew, F.H.S., M.C. Peel and A.W. Western, (2002). Application and testing of the simple rainfall-runoff model SimHyd. *Mathematical models of small watershed hydrology and applications (Eds V. P. Singh and D. K. Frevert), Water Resources Publication*: 335 - 367.

Chiew, F.H.S. and L. Siriwardena, (2005). Estimation of SimHyd parameter values for application in ungauged catchments. *In Zerger, A. and Argent, R.M. (eds) ModSim 2005 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand, December 2005,* www.mssanz.org/modsim05/papers

Gupta, H.V. and S. Sorooshian, (1983). Uniqueness and observability of conceptual rainfall-runoff model parameters: The percolation process examined. *Water Resources Research*, 19(1): 269-276.

Jakeman, A.J. and G.M. Hornberger, (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, 29(8): 2637-2649.

McIntyre, N.R., H. Lee, H.S. Wheater, A.R. Young and T. Wagener, (2005). Ensemble predictions of runoff in ungauged catchments. *Water Resources Research*, 41: W12434, doi:12410.11029/12005WR004289.

McKenzie, N.J., D.W. Jacquier, L.J. Ashton and H.P. Cresswell, (2000). Estimation of soil properties using the atlas of Australian soils. *CSIRO Land and Water Technical Report*, 11/00.

Merz, R. and G. Bloschl, (2004). Regionalisation of catchment model parameters. *Journal of Hydrology*, 287: 95-123.

Peel, M.C., F.H.S. Chiew, A.W. Western and T.A. McMahon, (2000). Extension of monthly unimpaired streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments. *In Australian Natural Resources Atlas*.

Peel, M.C., B.L. Finlayson and T.A. McMahon, (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences Discussions*, 4: 439-473.

Reichl, J.P.C., F.H.S. Chiew and A. Western, (2006). Model averaging, equifinality and uncertainty estimation in the modelling of ungauged catchments. *A. Voinov (editor), iEMSs 2006 International Congress* www.iemss.org/summit.

Vogel, R.M., (2005). Regional calibration of watershed models. *V. P. Singh and D. F. Frevert (editors) In Watershed Models, CRC Press*: 549-567.