Tools and Approaches for Evaluating Uncertainty in Streamflow Predictions in Ungauged UK Catchments

Neil McIntyre

Hyosang Lee

Howard Wheater

Andrew Young

Tools and Approaches for Evaluating Uncertainty in Streamflow Predictions in Ungauged UK Catchments

<u>Neil McIntyre</u>*, Hyosang Lee*, Howard Wheater*, Andrew Young[+]

**Imperial College London, UK. Email <u>n.mcintyre@imperial.ac.uk</u>*

[+] *Center for Ecology and Hydrology, Wallingford, UK*

Abstract: A main limitation of using conceptual models for predicting flow in ungauged catchments is the errors in identified relationships between calibrated conceptual parameters and known (or estimated) catchment descriptors. It is hypothesised here that these errors may be reduced if the modeller does not explicitly identify relationships, but applies all feasible models within a Bayesian averaging scheme. This maintains the information about parameter inter-dependencies obtained as part of local calibration, and also provides a strong basis for integrating various sources of uncertainty into the predicted average flow and associated confidence intervals. A case study of UK catchments provides encouraging results.

***Keywords**:* Rainfall-runoff; ungauged catchments; regression; Bayesian averaging

1. INTRODUCTION

A priori, without considering the nature of a catchment, there is a large range of conceptual models that might be used to model its rainfall-runoff response. Therefore, modellers generally try to constrain this model space by either (in the case of adequately gauged catchments) conditioning the model on observed rainfall-runoff data, or (in the case of ungauged or poorly gauged catchments) by regionalisation of models of similar well-gauged catchments. Regionalisation has commonly been approached by (a) selecting an appropriate model structure according to general catchment type and the experience of the modeller, and (b) regression of the gauged catchment parameters[1] against multiple catchment descriptors which are known, or may be estimated, for the ungauged catchment (e.g. Wagener 2002). In this paper, we will review the limitations of the regression approach to regionalisation, and propose that Bayesian model averaging is potentially a more sensible and powerful approach. A case study of 36 UK catchments is presented.

2. PARAMETER REGRESSION

Regression of model parameters against catchment descriptors (CDs) is problematic because the CDs are mostly inter-correlated, and so it is not intuitive which should be used as independent variables in the regression. Often this is allowed for by selecting a number of relatively orthogonal CDs which are perceived to have most hydrological significance, and neglecting the rest. Another problem is that the parameters of rainfall-runoff models are inter-dependent in highly non-linear ways. These cannot be sufficiently encaptured by a regression model (Kokkonen et al. 2003), which will linearise the problem (even if the data are transformed first). This problem of parameter inter-dependency is either neglected and each gauged/ungauged catchment parameter is treated as independent (multiple univariate regression), or is nominally included using methods such as multivariate regression, canonical correlation analysis or sequential regression. Using multivariate regression, correlations between the gauged catchment parameters can be included in the regression procedure, allowing covariance of the ungauged catchment parameters to be derived. Although employed with some success by Tung et al. (1997) for a two-parameter unit-hydrograph model, multivariate regression would become extremely complex if applied to a conceptual rainfall-runoff model with several parameters. Canonical correlation analysis (see Young 2000) uses linear combinations of both the CDs and the parameter sets as variables, thus partially allowing for CD inter-correlations as well as parameter inter-correlations. Using sequential regression (Lamb et al., 2000), univariate regression is applied to the perceived most important parameter, its values are

[1] We use the term "gauged catchment parameters" to refer to the calibrated parameters of the rainfall-runoff model of the gauged catchment.

fixed for each gauged catchment using the regressed estimates, then all the other parameters are re-calibrated. This continues sequentially through all the parameters, removing the issue of inter-dependency by making parameter estimates conditional on regionalised values of higher parameters.

Regression generally assumes normal or transformed normal distributions of residuals in calculating the uncertainty in regression coefficients (Haan 2002). The information about non-linear parameter interactions, which may be obtained using Monte Carlo-based calibration (e.g. Wagener et al. 2001), is lost.

Another limitation of commonly used regression approaches is that equal weight is generally given to all gauged catchments, irrespective of the quality of their data or how successfully they have been modelled using the chosen model structure. Wagener (2002) used weighted regression to reduce the influence of gauged catchment parameters that were poorly identified during calibration, but this does not introduce weights to allow for the model's local performance or data quality. Another source of uncertainty, not accountable using regression, is the uncertainty in the CDs themselves. For example, BFIHOST is a commonly used CD in the UK, but is an approximation of the base flow index which may carry considerable uncertainty. By the nature of regression, this uncertainty is neglected. In this paper, we propose a regionalisation scheme, within which all these sources of doubt can be easily integrated.

3. BAYESIAN AVERAGING

An alternative to regression is Bayesian averaging, where all potentially viable models (structures and parameter sets) are assigned prior probabilities which are updated based on the various sources of evidence. All models with non-zero probability are applied to the ungauged catchment, providing an ensemble distribution of flow forecasts. By eliminating the regression model, this maintains the full information content of the local models' parameter sets through to the ungauged catchment predictions. This methodology also has the potential to integrate many sources of uncertainty into the ungauged catchment model.

The Bayesian averaging idea has already been employed for many hydrological modelling applications. Bayesian averaging is central to the Generalised Likelihood Uncertainty Estimation (GLUE) framework of Beven and Binley (1992), which has been widely applied in hydrology, and Neuman (2003) has recently presented a formal framework for hydrologic model averaging based on maximum likelihood. Shamseldin et al. (1997) employ various schemes to identify averaging weightings for five different rainfall-runoff model structures applied to eleven gauged catchments, and they provide a concise review of other applications. However, previous application of Bayesian averaging to prediction in ungauged catchments seems to be very limited. The region-of-influence approach used in the UK Flood Estimation Handbook (Institute of Hydrology 1999), is comparable. For a poorly gauged catchment, a number of well-gauged catchments with CDs within a threshold of similarity (measured by proximity in the CD space) are identified and their data are integrated with equal weights, giving an estimate of the flood statistics in the poorly gauged catchment. The region-of-influence method has also been applied to estimation of low flow statistics (Holmes et al. 2002).

The general concept which we will use is illustrated simplistically in Figure 1. In Figure 1, θ is the only shown model parameter, although the notion is applicable to multi-parameter models, and only one CD is shown whereas there would normally be several. Figure 1(a) illustrates the commonly used regression approach where θ is regressed against the CD. In Figure 1(b), the weights W assigned to each gauged catchment model represent the perceived relative probability with which that gauged catchment model could be applied to the target gauged catchment. All models with non-zero W would be applied without averaging or interpolation of their parameters. W would necessarily relate to the similarity of that gauged catchment to the ungauged catchment, but also could relate to, for example, the identifiability of that model, and/or the quality of the data used to identify that model. The concept would easily be coupled with the output of a GLUE analysis, whereby a large sample of parameter sets each with their own weight are available for each gauged catchment, and could equally well encompass the notion of equifinality of model structures (Perrin et al. 2001). The simplest approach would be "nearest-neighbour" approach where the gauged catchment most similar to the target ungauged catchment is given W=1 and all other catchments neglected (W=0), although Young (2001) found this inferior to regression. Alternatively, using a region-of-influence approach, all models of gauged catchments which have a CD value within a prescribed deviation from that of the target ungauged catchment would be given equal W values, and all others would be neglected.

The averaging method would only work if there is at least one non-zero *W*, and (presumably) many would be preferred, and therefore regression might intuitively be more sensible if considerable extrapolation or interpolation between gauged catchments was required. However, in many countries there are a large number (hundreds) of gauged catchments (Perrin et al. 2001, Young 2001).

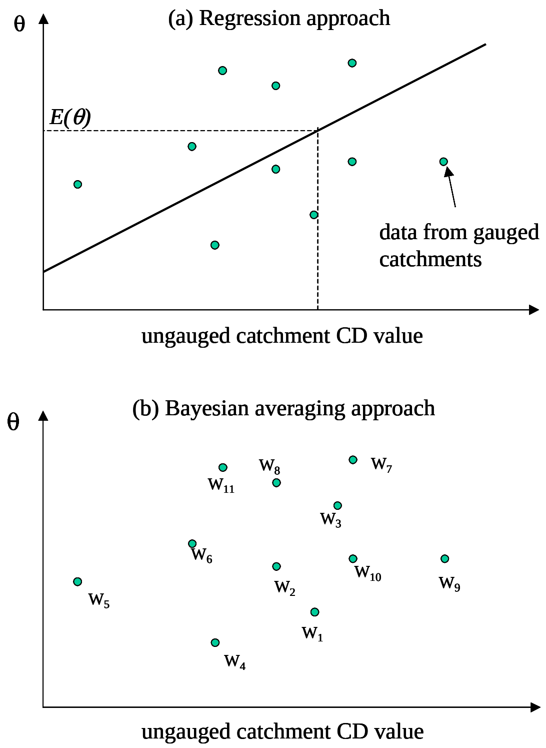

Figure 1. Regression and Bayesian averaging regional models

4. CASE STUDY

4.1 Gauged catchment models

In this case study we limit the analysis to a sample of 36 gauged catchments. The 36 catchments have been chosen from a database of gauged UK catchments kept by the Center for Ecology and Hydrology, Wallingford, UK (see Young, 2001). The 36 catchments have been selected to have a wide variation in catchment descriptors (within the UK context) and to all have high quality daily rainfall-runoff data in a common period. 30 of the catchments are used to formulate the regional model, and the other six are used for testing its performance. The 5-year period 1989-1994 was used for all catchments. The CDs of these catchments were obtained from the Flood Estimation Handbook (Institute of Hydrology 1999). As a summary description of the six test catchments, Table 1 gives their base flow indices (*B*), standardised average annual rainfalls (*R*), and catchment areas (*A*).

The modelling is done using the Rainfall-Runoff Modelling Toolkit of Wagener et al. (2001), and for simplicity a single lumped model structure is employed – the probability distributed model of Moore (1985) which calculates effective rainfall, coupled with a routing model consisting of two linear stores in parallel (representing fast and slow routing). This model structure has been found to perform relatively well over a range of UK catchment types (Young 2000), using a two-parameter Pareto distribution to define the spatial distribution of soil moisture capacity. It has six parameters that are considered to be uncertain here; the maximum soil moisture over the catchment, the Pareto distribution shape parameter, the proportion of rainfall which bypasses the soil stores, the split of the effective rainfall between fast and slow routing stores, and the fast and slow routing time constants.

Table 1. Summary of the six test catchments

Catchment	*B* [-]	*A* [km^2]	*R* [mm/yr]
Thet at Bridgham	0.68	276	640
Ithon at Disserth	0.43	359	1130
Bela at Beetham	0.54	132	1298
Coquet at Morwick	0.39	578	884
Coquet at Rothbury	0.40	346	951
Medway at Chafford	0.44	252	852

B = base flow index, *A* = catchment area, *R* = standardised annual average rainfall

Firstly, the models were calibrated using plain random sampling from within prescribed parameter bounds. 10000 random samples were taken, and the best-performing parameter set was identified. As well as this optimum parameter set, the next best nine parameter sets for each catchment were retained, so that parameter equifinality can be later integrated into the model averaging. The Nash-Sutcliffe Efficiency was used as the fit criterion.

4.2 Regression

Firstly, we use multiple univariate regression for the regional model (see Haan (2002) for the theory behind this method). To decide which CDs to use as regressors, rank correlations between all CDs and optimum model parameters were calculated. From this correlation analysis, the following CDs were chosen as regressors, on the basis of high correlation with one or more parameters and low correlations with previously selected regressors:

- Base flow index.
- Standardised average annual rainfall.
- Catchment area.
- Standard period potential evaporation (from Young 2000).
- Variation of the drainage network distances.
- Mean direction of all 50m slopes.
- Extent of urban land cover in 1990.
- Flood attenuation by reservoirs and lake.

Each of the six model parameters was regressed independently against these CDs. The optimum parameter values and associated CDs were used as the regression data (i.e. 30 data points). The covariance matrix of the regression coefficients was calculated so that uncertainty in the ungauged parameters could be estimated.

This regional model was then applied to the six test catchments that were not included in the regional model data. Mean values and variances of all parameters for each test catchment were calculated. Assuming the parameters to be normally distributed, Monte Carlo simulation (1000 samples) was used to simulate 90% confidence limits on the flow time series for each test catchment. An example time-series is given in Figure 2(a), for the Thet at Bridgham catchment. The NSE performances obtained using the mean streamflow obtained from the Monte Carlo simulation are in Table 2.

4.3 Model averaging

We now conduct a preliminary evaluation of whether model averaging has the potential to improve the forecasts in these six ungauged catchments. Three separate averaging schemes are used. Firstly, only the optimum gauged catchment parameter sets are included, and the *W* values are calculated using the similarity of the respective gauged catchment to the target ungauged catchment:

$$W_i = \frac{1}{\sum\left(1 - E_i / E_{\max}\right)}\left(1 - E_i / E_{\max}\right) \qquad (1)$$

$$E_i = \sqrt{\frac{1}{2}\left(\frac{\ln A_i - \ln A'}{\sigma(\ln A')}\right)^2 + \left(\frac{\ln R_i - \ln R'}{\sigma(\ln R')}\right)^2 + \left(\frac{B_i - B'}{\sigma(B')}\right)^2} \qquad (2)$$

where subscript *i* refers to the i^{th} of the 30 gauged catchment models, *E* is a measure of catchment dis-similarity, and E_{max} is the maximum value of *E* for that ungauged catchment. *A*, *R* and *B* are, respectively, the catchment area, standardised annual average rainfall, and base flow index (as estimated by the HOST model) of the gauged catchments, and *A*', *R*' and *B*' are the same for the target ungauged catchment. See Institute of Hydrology (1999) for more detail about this *E* measure and the associated CDs.

Secondly, only those parameter sets for which E is below a threshold value are assigned non-zero weights.

$$W_i = \frac{1}{\sum_S\left(1 - E_i / E_{\max}\right)}\left(1 - E_i / E_{\max}\right) \quad \text{for } E_i < 1 \qquad (3a)$$

$$W_i = 0 \quad \text{for } E_i \geq 1 \qquad (3b)$$

where *S* is the number of gauged catchment models under the threshold $E_i < 1$. Finally, using this threshold scheme, all 10 of the best-performing parameter sets for the *S* gauged catchments are included, and the weight for each one is a combination of its calibration performance (as defined below) and the catchment similarity. 10 is assumed here to be the number of parameter sets that might be feasible, given input data errors, but is open to review and revision in future work. For this analysis, there are $S\times10$ values of *W* for each test catchment defined by,

$$W_{i,j} = \frac{1}{\sum\left(W1 \times W2\right)} W1_i \times W2_{i,j} \qquad (4)$$

where subscript *i* refers to the i^{th} of *S* similar gauged catchments, *j* refers to the j^{th} of 10 parameter sets, *W1* is the *W* defined in Equation 3 and *W2* is the NSE value obtained from each parameter set during calibration. Therefore, in this case the influence of the poorer performing parameter sets is weighted down, as are the influences of the gauged catchments which did not yield well performing models. Using the NSE in this way has the potential to reduce the effects of unreliable data sets (and poorer model structures if more than one is used), as this would tend to cause lower NSE values.

5. RESULTS

For the six test catchments Table 2 gives the following NSE values: the best fit using local model calibration (a); the averaged flow from Monte Carlo simulation of the regressed parameter distributions (b); the three alternative Bayesian averaging schemes (c-e); and the fit using the optimum model of the most similar gauged catchment (f).

Performance using local calibration defines the benchmark against which to evaluate the success of the other methods. The Bayesian averaging including all 30 gauged catchment models has generally done worse than the regression method, while using only the more similar catchments with an E threshold of 1 has done generally better. Using the weighted average of all S×10 retained parameter sets has done consistently better than regression. In four cases using only the optimum model of the most similar gauged catchment was better than regression, but in no cases was it the best method, and in two cases it was found to be the worst approach, marginally. Although no single averaging method was consistently better, taking the best out of them comes close to matching local calibration performance. There was no evidence that any one type of approach was especially amenable to one type of catchment.

Figure 2 shows time-series results for the Thet at Bridgham catchment. Figure 2(a) shows the 90% confidence intervals calculated using Monte Carlo sampling of the regressed parameter distributions. Figure 2(b) shows the ensemble of time-series obtained using each of the 30 optimum parameter sets from the gauged catchment models. This ensemble represents our knowledge prior to considering the nature of our ungauged catchment. Figure 2(c) shows the smaller ensemble of the four optimum parameter sets from the 'similar' catchments, representing our posterior knowledge as constrained by the similarity threshold. Figure 2(d) shows the weighted average flow from the same four optimum parameter sets (weights from Equation 1), illustrating the success of this regionalisation scheme, at least for this catchment. While the NSE is a useful summary of performance, visualisation of the time-series in this way gives us greater insight into how successfully our criteria have constrained our knowledge for various flow regimes.

6. DISCUSSION

At the outset of the paper we suggested that a Bayesian averaging scheme may be a more sensible method of regionalisation than parameter regression methods. Table 2 indicates that this may be true, although more work is required to substantiate this. The regression method used is univariate regression without any transformation of the CDs, and it may be expected that more sophisticated methods would perform better. Also, some more thought is needed in selecting regressors, and in designing similarity measures (a particular limitation of this paper is that more CDs were used in the regression than were used for defining similarity). A small population of gauged catchments has been used and a larger set is needed to make stronger conclusions. In particular, we would like to explore the extent to which the averaging approach relies on a having group of similar gauged catchments, and whether it is useful when extrapolation to an extreme type of catchment is required. The small number of ungauged catchment models has restricted us to representing posterior prediction uncertainty as an ensemble of trajectories in Figure 2(c). A higher number, as well as presumably leading to a more reliable average, would allow confidence limits to be presented. Another priority for further work is to develop the model evaluation beyond the NSE measure. This might include multi-objective evaluation to assess how we should design our regionalisation according to the objective of the modelling, e.g. whether it is directed at high flow or low flow applications.

Table 2. NSE performances of alternative regionalisation schemes for six test catchments

Catchment	(a)	(b)	(c)	(d)	(e)	(f)
Thet at Bridgham	0.84	0.75	0.68	**0.81** (4)	**0.77**	**0.80**
Ithon at Disserth	0.85	0.82	0.81	**0.83** (4)	**0.83**	0.81
Bela at Beetham	0.89	0.82	0.61	0.63 (4)	**0.89**	**0.87**
Coquet at Morwick	0.64	0.61	0.61	**0.63** (2)	**0.62**	0.60
Coquet at Rothbury	0.63	0.58	**0.62**	**0.62** (1)	**0.61**	**0.62**
Medway at Chafford	0.82	0.40	**0.81**	**0.80** (5)	**0.80**	**0.78**

Note: (a) Local calibration, (b) Regression, (c) Model averaging (using the optimum parameter set from all 30 gauged catchments), (d) Model averaging (using the optimum parameter set from S most similar catchments), with S given in parenthesis, (e) Model averaging (using the 10 best parameter sets from the similar catchments), (f) Using the model from only the most similar gauged catchment. Regionalisation performances better than regression are boldened.

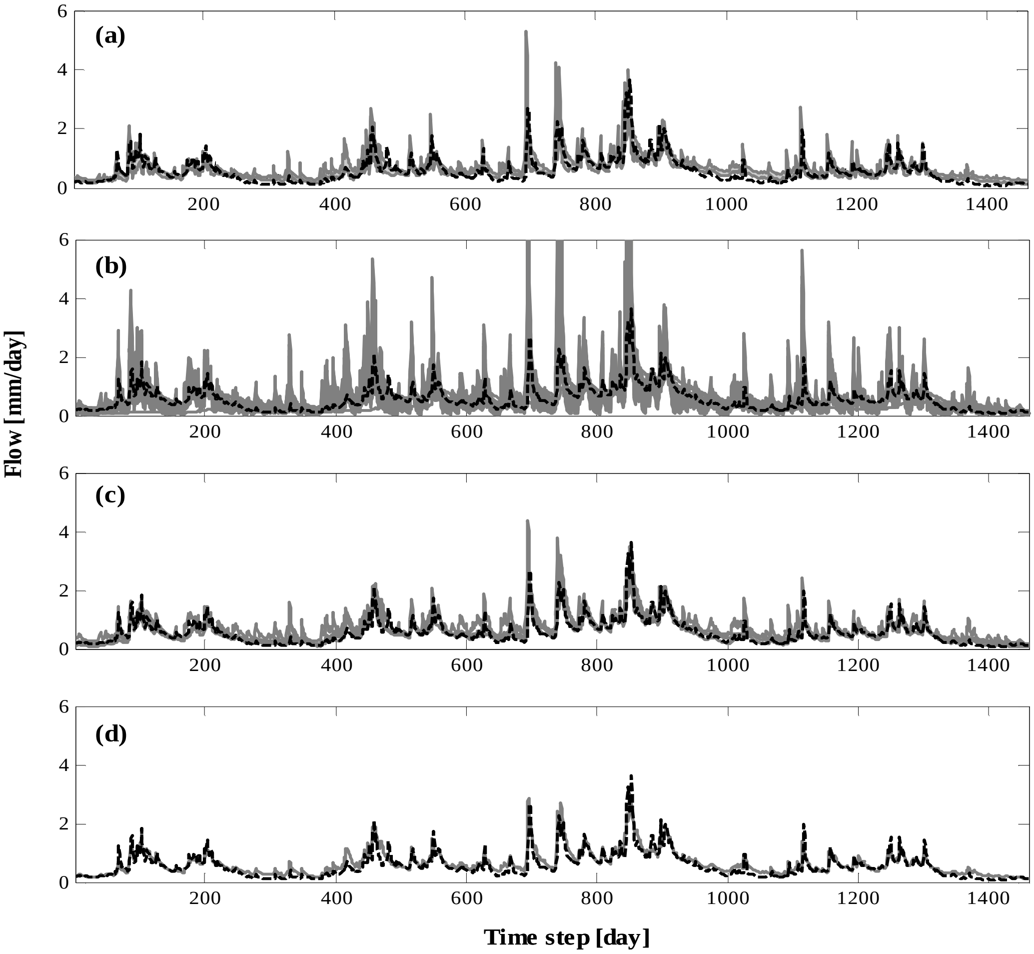

Figure 2. Comparison of observed and simulated flow time-series from Oct 1984-Sept 1989 for the Thet at Bridgham. (a) 90% confidence intervals derived using the regression, (b) ensemble of model results from using all 30 gauged catchment optimal models, (c) ensemble from using the optimal models of the 4 most similar gauged catchments, (d) the weighted average result of these 4 models.

7. REFERENCES

Beven, K., Binley, A. 1992. The future of distributed models; model calibration and predictive uncertainty. Hyd. Processes 6, 279-298.

Haan, C. 2002. Statistical Methods in Hydrology, 2nd Edition. Blackwell.

Holmes M., Young A., Gustard A., Grew R. 2002. A region of influence approach to predicting flow duration curves within ungauged catchments, Hyd. & Earth Sys. Sci., 6 (4): 721-731

Institute of hydrology (IH), 1999, Flood Estimation Handbook, Wallingford, UK.

Kokkonen T., Jakeman A., Young P., Koivusalo H. 2003. Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina, Hyd. Processes, 17(11) :2219-38.

Lamb, R., Crewett, J., Calver, A. 2000. Relating hydrological model parameters and Catchment properties to estimate flood frequencies from simulated river flows. BHS 7th National Hydrology Symposium, Newcastle-upon-Tyne.

Moore, R. 1985. The probability-distributed principle and runoff prediction at point and basin scales. Hydrological Sciences Bulletin, 30(2), 273-297.

Neuman, S. 2003. Maximum likelihood Bayesian averaging of uncertain model predictions. Stoch. Env. Res. & Risk Assessment, 17, 291-305.

Perrin, C., Michel, C., Andreassian, V. 2001. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. J. of Hyd., 242: 275-301.

Shamseldin, A., O'Connor, K., Liang, G. 1997. Methods for combining the output of different rainfall-runoff models. J. of Hyd., 197: 203-229.

Tung, Y., Yeh, K., Yang, J. 1997. Regionalization of unit hydrograph parameters: Comparison of regression analysis techniques. Stochastic Hydrology & Hydraulics, 11, 145-171.

Wagener, T., Boyle, D., Lees M., Wheater, H., Gupta, H., Sorooshian, S. 2001. A framework for the development and application of hydrological models. Hyd. & Earth Sys. Sci., 5(1), 13-26.

Wagener. T, 2002. Identification of parsimonious rainfall-runoff models for gauged and ungauged catchments. PhD Thesis, University of London.

Young, A. 2000. Regionalising a daily rainfall-runoff model within the United Kingdom. PhD Thesis, University of Southampton.

Young, A. 2001. River flow simulation within ungauged catchments using a daily rainfall-runoff model. National meeting of the British Hydrological Society, July 2001.