Climate Variability and Change—Hydrological Impacts (Proceedings of the Fifth FRIEND World Conference held at Havana, Cuba, November 2006), IAHS Publ. 308, 2006.

Predicting low flows in ungauged basins: a hydrological response unit approach to continuous simulation

ANDREW R. YOUNG, VIRGINIE KELLER & JIM GRIFFITHS
Centre for Ecology and Hydrology, Wallingford, Oxfordshire OX10 8BB, UK
ary@ceh.ac.uk

Abstract An alternative approach to the problem of the regionalization of a rainfall–runoff model is explored in which the concepts of Hydrological Response Units (HRU) are used to define a flexible model structure where catchment descriptors are used to define relatively complex, specific model structures within each catchment from one generic structure. This model structure has been simultaneously calibrated, and hence regionalized across a large set of catchments within the UK. The simulation results obtained with this approach have been compared with those from a simple lumped model regionalized by relating prior calibrated model parameters to catchment descriptors. The paper concludes by suggesting that model structure and regionalization strategy may not be significant in limiting the performance of regionalized models and that they can be improved by incorporating sparse, at-site measurements.

Key words rainfall–runoff models; continuous simulation; hydrological response units; regionalization; water resources

INTRODUCTION

Access to daily streamflow data at the river reach scale is a central component of many aspects of water resource and water quality management. The requirement for these data within ungauged catchments has been one of the many drivers behind research into the regionalization of conceptual rainfall–runoff models. The theory behind the regionalization objective is that if the structural description of the rainfall–runoff model is correct, the parameters of the model are more likely to be related to physical descriptors of the catchment that can be measured. However, as a consequence of scaling issues and the many unknowns and sources of error there are significant problems in retaining the physical basis of model parameters as models are scaled from the plot to the catchment scale.

The problems in regionalizing rainfall–runoff models are manifold; some key issues are summarized here. The body of literature on model structure, calibration, effect of forcing data errors (including observed streamflow) on calibration, parameter identifiability and whether any model structure has truly identifiable parameters is immense (see, for example Beven, 1993) and is not the subject of this paper. Broadly speaking, identifiability can be increased by reducing the complexity, and hence the number of parameters within a model (if streamflow data are to be the sole source of calibration data), minimizing uncertainty within the input data and by the selection of appropriate objective functions for measuring model simulation performance. If parameters are non identifiable then the likelihood of expressing them as a function of catchment descriptors is reduced. It has been commonly proposed that parameter parsimonious, generally lumped model structures are best suited to regionalization (Seibert, 1999). The consequence of all of these factors is that model parameters may have little physical relevance. Additionally, the catchment descriptors that can be used as predictors are nearly always extrapolated from point measurements, which are inherently erroneous and may not be independent from one another. Furthermore, dependencies are rarely amenable to quantification through the use of classical statistics.

The published approaches to regionalizing the relationships between model parameters and catchment characteristics are varied; the most common approach is to calibrate a model structure, usually lumped, within a set of gauged catchments that are representative of the region in which the model is to be utilized and to subsequently develop relationships (normally statistical) between parameters and descriptors describing the physical and climatological structure of a catchment. The commonest examples of these are regression based relationships (e.g. Abdulla & Lettenmaier, 1997); other approaches have been based on nearest neighbour approaches based on spatial proximity (e.g. Merz & Blöschl, 2004). Young (2006) considered both the multivariate regression approach to the regionalization problem and a nearest neighbour approach based on similarity in catchment descriptors. The study was based on a data set of 260 United Kingdom catchments. The approaches were explored using a simple, four parameter rainfall–runoff model based on the PDM model of Moore (1985), the study concluded that the regression based approach gave the best regionalized model performance.

The study presented in this paper builds on this previous work to explore whether the approach of *a priori* calibration of lumped catchment models with the subsequent regionalization of model parameters is the optimal approach to regionalizing the rainfall–runoff process. The approach proposes a flexible hydrological response unit (HRU) based model structure in which catchment descriptors of vegetation and soil type are used to define more complex, specific model structures within each catchment than can be identified from streamflow data from a single catchment. To enable subsequent parameter identification from streamflow data, the model parameters are simultaneously optimized against streamflow data across all catchments to combine model calibration and regionalization in a single step process. The scheme was developed using a data set of 501 UK catchments. This paper compares the simulation results obtained using this new class of regionalized model with those obtained earlier by Young (2006) (using 260 catchments) over the common 163 catchments within the data set. The comparison focuses on predictive uncertainty and concludes with some comments about the importance of hydrological structure and the way forward for the reduction of predictive uncertainty within ungauged catchments.

CATCHMENT DATA SETS

The catchment data set used was an expansion of that used for the previous study which was restricted to the selection of natural, high hydrometric quality catchments. Daily time series of 648 catchment average precipitation and potential evaporation were derived for all catchments using the methods described by Young (2006). The catchment data set was reduced to 501 by excluding those catchments in which a water balance assumption, explicit within the model, was obviously violated. This data set was subdivided into 217 catchments used for model calibration and regionalization, and 284 catchments retained for evaluation.

The catchment descriptors used within the model structure are: a hydrologically referenced 50 m resolution, Digital Terrain Model (DTM) (Morris & Heerdegen, 1988); the Hydrology of Soil Types (HOST) 29 class hydrological response classification of soils across the United Kingdom (Boorman *et al*., 1995) and a classification of land cover based on five broad vegetation classes: Deciduous, Coniferous, Arable, Grassland, and Upland, derived from the CEH Land-cover Map 2000 classification system mapped at a resolution of 50 m (Smith *et al*., 2001).

MODEL STRUCTURE

The model structure is based around two sub model components; the loss module that generates hydrologically Effective Precipitation (EP) and the routing module that subsequently routes the EP to the catchment outlet. The basic model structure for the loss module is a hydrological response unit consisting of an interception sub-module and a treatment of transpiration losses based on the *FAO 56* soil moisture accounting procedures for determining crop water requirements (Allen *et al*., 1998). The interception model was regionalized for inclusion within a rainfall–runoff model by Young (2006). The model has one parameter; the maximum depth of water that can be held by the vegetation, γ. The conceptual structure of the FAO transpiration module is presented within Fig. 1. The module describes vegetation as a function of maximum root depth, Z_r, and “moisture depletion fraction”, p, for a range of vegetation and soil types. The Total Available Water (TAW), the amount of water available to plants after a soil has drained to its field capacity is defined as the product of the difference between field capacity (FC) and wilting point (WP) (properties of the soil class) and Z_r. Plants freely transpire until Soil Moisture Deficits (SMD) exceed the threshold defined by pZ_r; beyond this threshold the plants become increasingly stressed and evaporation reduces below the potential rate in proportion to the depth of threshold exceedence. EP is generated by the module when the SMD within the module is zero.

HRUs were defined by combining the HOST classes and reduced land-cover classes to yield a potential 140 combinations of HOST and land cover classes plus an open water class. In practice

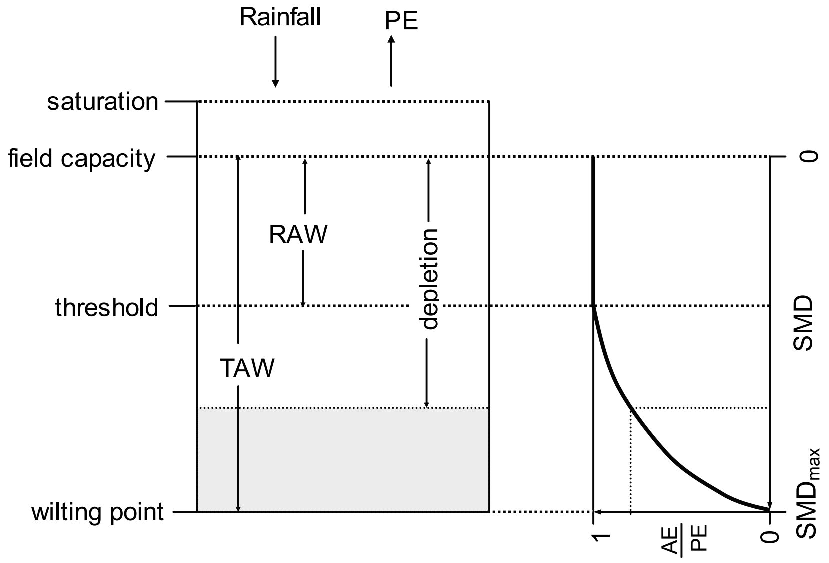

Fig. 1 Conceptual structure of the soil moisture accounting within the loss module.

the number of actual combinations was significantly less as some land-cover/soil class combinations do not occur in practice. At the catchment level the individual cells within the HRUs represented within the catchment are amalgamated to form HRUs with a fractional extent that is not necessarily contiguous within the catchment. The response of each HRU is controlled by the vegetation parameters of γ, Z_r and p and the *FC* and *WP* parameters for the soil class. *FC* and *WP* parameters were defined for each soil class based on extracting the average percentages of sand, silts and clays within each HOST class from the UK National Soil Resource Institute's SEISMIC data set. This process leaves Z_r and p, for each vegetation class as the free parameters for the HRUs within the loss module which equates to 10 parameters in total.

The output from the loss module within a catchment is an EP time-series for each 1 km cell within the catchment. The routing module routes these effective precipitation time series to the catchment outlet via a semi-distributed routing scheme. Within the UK, the dominant influences on the routing of water through the land surface are soils, hydrogeology and topography. In the absence of an appropriate resolution digital hydrogeological classification of the UK, the 29 HOST classes were amalgamated into 11 hydrogeological routing HRUs based on substrate geology. The EP time series for each cell enters the routing HRU corresponding to the HOST class of the cell. A probability distributed storage model is used to represent the free water in the soil column for the routing HRU. This storage model is assumed to be uniformly distributed with a maximum storage depth of 75 mm (determined by preliminary individual catchment model applications), drainage takes place from the base of the store and is proportional to the depth of water held in storage. The constant of proportionality, *Kg*, is a free parameter. The runoff from the store is routed though a topographically defined routing path whilst the drainage is routed through a linear reservoir, with a time constant *Kb*, representing the baseflow for the routing class.

The previous regionalization work on lumped models, in which quick flow routing was via a linear reservoir, demonstrated a strong dependency of the reservoir time constant on both catchment size and soil storage. For this study the quick flow routing within the routing HRUs was subdivided into a topographic component and a component representing transient soil storage along the routing path. The topographic routing of the quick flow from the individual cells within a routing HRU to the catchment outlet was based upon the flow path defined from the DTM and the cell level topographic gradients along the path, β. Total travel time to the catchment outlet, T, is calculated for each cell as:

$$T = \sum_{i=1}^{N} \frac{x_i}{\beta_i v} \tag{1}$$

where x_i is the distance between the centroids of adjacent cells within the flowpath and v is a routing HRU dependent velocity which conceptually is linked to bulk, lateral hydraulic conductivity. The total topographically routed quick flow for the routing HRU is calculated as the sum of the EP time series for the constituent cells lagged by the corresponding cell travel times. The resultant summed time series is then passed through a linear reservoir of time constant, Kq, to represent the transient storage along the flow path lengths. For each routing HRU the model has four free parameters: Kg, V, Kb and $K1$, and with 11 routing class this gives a total of 44 routing parameters within the model which combined with the 10 free loss module HRU parameters yields a total of 54 model parameters for calibration.

MODEL CALIBRATION

The parameters were simultaneously optimized across all catchments within the calibration sets using data from the 15 year period from 1987–2001. This process combines both model calibration and regionalization into a one-step process as the model parameters are a function of the HRUs and hence the descriptors of the catchments. The loss module HRUs were initially calibrated to make the process of simultaneous calibration computationally tractable. The mean and variance of the distribution of the bias between simulated mean EP and the corresponding observed mean runoff, expressed as a percentage of the observed mean runoff (BIAS) over the calibration period was minimized as the objective function. The routing module parameters were subsequently optimized based on the trade off between maximising the Nash-Sutcliffe Efficiency Criterion (NSE) and minimizing the sum of squared deviations between observed and simulated streamflow over the lowest third of the flow distribution (LF_OBJ) and the bias error at the Q95 flow (BEQ95) across all catchments within the calibration data set. The NSE was used as a general measure of fit whilst the latter functions are a measure of fit at low flows. The earlier periods of flow data within the calibration catchments and the evaluation catchments were used to ensure the model was not over fitted within the calibration catchments.

EXAMPLE RESULTS

To facilitate comparison with Young (2006) the common catchments between the studies were grouped into the following classes: R1-catchments that were used in the regionalization of both models; R2-catchments that were not used in the regionalization of either model; R3-catchments that were used in the regionalization of the current model but not used in the development of the regionalization of the previous study and, finally, R4-catchments that were used in the regionalization work within the previous study but not used for the regionalization of the model in the current study. Example simulation results are presented for each catchment class in Table 1

Table 1 Summary of simulation results across catchment classes.

Class	c.i. limits	Bias		NSE		BE_Q95		lf_obj	
R1 n = 85	68% u.l.	**6**	*4*	**0.80**	*0.81*	**69**	*152*	**78**	*104*
	median	**–1**	*–2*	**0.73**	*0.73*	**–11**	*26*	**54**	*51*
	68% l.l.	**–10**	*–8*	**0.59**	*0.59*	**–52**	*–26*	**34**	*35*
R2 n = 31	68% u.l.	**8**	*8*	**0.79**	*0.75*	**17**	*58*	**63**	*69*
	median	**–7**	*–6*	**0.67**	*0.67*	**–16**	*5*	**49**	*50*
	68% l.l.	**–13**	*–12*	**0.45**	*0.49*	**–41**	*–32*	**36**	*39*
R3 n = 37	68% u.l.	**5**	*6*	**0.81**	*0.74*	**82**	*133*	**95**	*102*
	median	**–6**	*–5*	**0.74**	*0.69*	**–8**	*4*	**55**	*56*
	68% l.l.	**–10**	*–9*	**0.58**	*0.51*	**–43**	*–42*	**41**	*42*
R4 n = 73	68% u.l.	**9**	*13*	**0.75**	*0.80*	**27**	*60*	**61**	*66*
	median	**–2**	*0*	**0.64**	*0.70*	**–33**	*9*	**47**	*44*
	68% l.l.	**–8**	*–8*	**0.46**	*0.55*	**–51**	*–30*	**34**	*33*
All n = 226	68% u.l.	**8**	*8*	**0.80**	*0.79*	**55**	*100*	**76**	*80*
	median	**–2**	*–2*	**0.69**	*0.70*	**–17**	*13*	**48**	*49*
	68% l.l.	**–10**	*–9*	**0.54**	*0.55*	**–45**	*–32*	**33**	*35*

together with the results over all catchments. The variation of each of the objective functions across the catchments within a class is summarized as the limits defining the 68% confidence interval for the simulation results and the median value within the class. The class membership is also presented in Table 1. The results for the current semi-distributed model are highlighted in bold whilst the results for the lumped model are presented in italics. The stability of both models is similar for the overall model BIAS and NSE. across all catchment classes. This indicates very similar model performances, in terms of water balance closure and general time-series fit, irrespective of model type, whether the catchments were used within the regionalization scheme or whether they are fully independent. The NSE. 68% c.i. for the semi-distributed model tends to be marginally wider than that for the lumped model, but the median values are very comparable. Considering the fit at low flows, the semi-distributed form tends to perform better than the lumped model, particularly with respect to the 68% u.l. for the BEQ95 statistic (representing an over-prediction at Q95). This is also reflected in the LF_OBJ with the values for the semi-distributed model generally being lower for all limits, thus indicating a better fit.

CONCLUSIONS

The approach of using catchment descriptors to define the model structure and the combination of model calibration and regionalization within a one-step process has many conceptual advantages over the traditional approach of seeking statistical relationships between *a priori* calibrated parameters for a lumped model and catchment descriptors. The comparison with the results of a previous study has demonstrated that the approach yields a better fit to low flows but marginal benefits in terms of the simulation of mean flow (BIAS) and overall time-series fit, as measured using NSE. In a UK context, the model structure and regionalization approach may not therefore be the limiting factors within this type of study. This suggest that other factors are limiting; these may include: input data errors, subtle violation of the assumption of a closed water balance at the catchment scale and the limitation of current catchment descriptors in aiding the differentiation of the hydrological regimes between river catchments. It is stressed that the results are preliminary, and research is ongoing to separate the aspects of model performance into those related to these factors and those attributable to model and parameterization error. Given the intractable nature of improving catchment descriptors, a fruitful future research direction is to investigate how sparse measurements, for example a short period of continuous measurement or a set of discrete measurements covering a range of flows, may be used to constrain model uncertainty within ungauged catchments.

REFERENCES

Abdulla, F. A. & Lettenmaier, D. P. (1997) Development of regional parameter estimation equations for a macroscale hydrologic model. *J. Hydrol.* **197**, 230–257.

Allen, R. G., Pereira, L. S., Raes, R. & Smith, M. (1998) Crop evapotranspiration – Guidelines for computing crop water requirements. *FAO Irrigation and Drainage Paper 56*, Food and Agriculture Organization of the United Nations, Rome, Italy.

Beven, K. (1993) Prophecy, reality and uncertainty in distributed hydrological modelling. *Adv. Water Resour.* **16**, 41–51.

Boorman, D. B., Hollis, J. M. & Lilly, A. (1995) *Hydrology of Soil types: A Hydrologically-based Classification of the Soils of the United Kingdom.* Institute of Hydrology Report no. 126. Wallingford, UK.

Merz, R. & Blöschl, G. (2004) Regionalisation of catchment model parameters. *J. Hydrol.* **287**, 95–123.

Moore, R. J. (1985) The probability-distributed principle and runoff production at point and basin scales. *Hydrol. Sci. J.* **30**, 273–297.

Morris, D. & Heerdegen, R. (1988) Automatically derived catchment boundaries and channel networks and their hydrological applications. *Geomorphology* **1**, 131–141.

Seibert, J. (1999) Regionalisation of parameters for a conceptual rainfall-runoff model. *Agric. Forest Met.* **98–99**, 279–293

Smith, G. M., Fuller, R. M., Sanderson, J. M., Hill, R. A. & Thompson, A. G. (2001) Land Cover Map 2000: a parcel-based approach from satellite images. In: *Proceedings of the RSPS meeting in Uncertainty and Remote Sensing and GIS*, 689–702.

Young A. R. (2006) Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model. *J. Hydrol.* **320**, 155–172.