Prediction of rainfall–runoff model parameters in ungauged catchments

M. ZVOLENSKÝ[1], K. HLAVČOVÁ[2], S. KOHNOVÁ[2] & J. SZOLGAY[2]

1 *Slovak Hydrometeorological Institute, Jeséniová 17, 833 15 Bratislava, Slovakia*

2 *Department of Land and Water Resources Management, Slovak University of Technology, Radlinského 11, 813 68 Bratislava, Slovakia*

silvia.kohnova@stuba.sk

Abstract The paper presents a modification of a rainfall–runoff parameter regionalization procedure, which can be used for model parameter estimation in ungauged basins. The upper Hron River basin in Slovakia was selected as a study area. Data from 19 sub-catchments were collected and a lumped conceptual rainfall–runoff model was calibrated in each of them. The model accounts for snow accumulation, soil moisture and groundwater balance, and runoff generation in a daily time step and has 15 parameters to calibrate. The catchments were pooled into homogeneous groups with respect to selected physiographic catchment characteristics using non-hierarchic clustering. Multiple regression relationships between rainfall–runoff model parameters and physiographic catchment characteristics were sought separately within the pooled catchments groups. The performance of these was compared with the performance of such relationships derived for the entire Hron catchment. The performance of model parameter predictions improved through considering groups of similar catchments as a basis for predictions.

Key words model parameter regionalization; clustering; homogeneous regions

INTRODUCTION

The main problem in using conceptual water balance models in ungauged catchments is related to the determination of parameter values since streamflow data is missing in such cases and a watershed model cannot be calibrated in the traditional way. Although recently considerable experience has been gained with rainfall–runoff model parameter estimation methods for ungauged catchments, there is a continuing need to upgrade these and to test them against practical requirements, because the problem of regional parameter estimation still cannot be considered as satisfactorily solved.

The model parameter regionalization problem is relatively long standing, with notable early studies, e.g. by Nash (1960) and Ross (1970). Many researchers have followed the same methodological approach. First, a watershed model is calibrated based on data available at a number of sites in a region. This is followed by the application of a regionalization method which attempts to relate the calibrated model parameters to the catchment characteristics. The most common method that is used is bivariate and multivariate regression. Abdulla & Lettenmaier (1997), Sefton & Howarth (1998) and Xu (1999) provide reviews of such studies.

Recent improvements in the availability of spatial data, together with improved computational resources and a new interest in continuous simulation, have led to new

Copyright © 2007 IAHS Press

research activity in this field. Kokkonen *et al*. (2003) concluded that high significance of regression between model parameter values and catchment characteristics does not guarantee a set of parameters with a good predictive power when applied to ungauged catchments and that care must be taken when interpreting the physical meaning behind the model parameter–catchment attributes relationship identified. Merz & Blöschl (2004) and Parajka *et al.* (2005) used physiographic catchment characteristics to estimate the HBV model parameters in selected catchments of Austria with diverse approaches. In their complex studies they realized that the correlations between model parameters and catchment attributes were not high in general. In this paper, therefore, a modification of such methodology is proposed for the regionalization of watershed model parameters. It is based on our previous study (Szolgay *et al*., 2003), in which we proposed grouping of catchments into groups with a similar runoff regime for the regional estimation of parameters of a monthly water balance model. The model parameters were subsequently estimated in all catchments simultaneously by regional model calibration in a group of catchments. Kokkonen *et al*. (2003) also concluded that to get more confidence in regional estimation of model parameters, catchments should form hydrologically homogeneous groups. Therefore, in this study the gauged catchments in the region of interest were first divided into homogeneous groups with respect to selected physiographic characteristics, using clustering. Multiple regression relationships between model parameters and physiographic catchment characteristics were sought within the pooled catchments groups. The performance of these regional relationships was compared with the performance of relationships derived by multiple regression for the whole region.

DATA USED IN THE STUDY

The pilot basin selected for this study is located above the Brehy gauging station of the Hron River basin (3821 km^2) in the middle of Slovakia. The digital elevation model of the basin, the river network and gauging stations used in this study are shown in Fig. 1. The minimum elevation of the basin is 195 m a.s.l., the maximum elevation is 2043 m a.s.l. and the mean elevation is 690 m a.s.l. Nineteen sub-catchments (for their location see Fig. 2) were selected, where daily flow, precipitation and air temperature time series needed for calibration of the rainfall–runoff model were available for the period 1980–2000.

The following physiographic catchment characteristics were derived from existing meteorological, geological and soil maps and from hydro-meteorological records for the period 1931–1980 for each sub-catchment: catchment area [km^2] (denoted by F), mean elevation of the catchment [m a.s.l.] (E), mean catchment slope [%] (SL), mean aspect of the catchment's slopes [°] (A), shape parameter of the catchment computed as the ratio between the catchment area and the power of main stream length [–] (Sh), density of the river network [km km^{-2}] (RND), mean slope of the main river in the catchment [%] (MSG), time of flow concentration computed according to the Kirpich formula [h] (Tc), hydrogeological transmissivity index, characterizing the permeability and water bearing capacity of the subsurface layer [–] (HgTI), index of the infiltration capacity of the upper soil level [–] (STI), percentage of forested area in the catchment

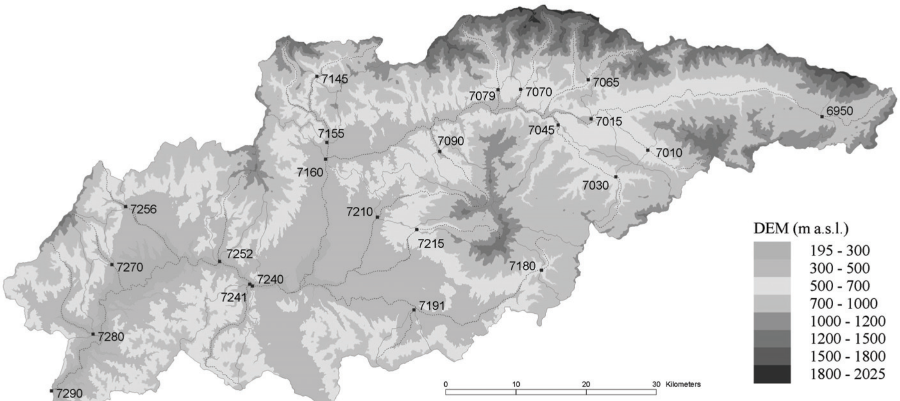

Fig. 1 Digital elevation model, river network and location of flow gauging stations in the Hron catchment.

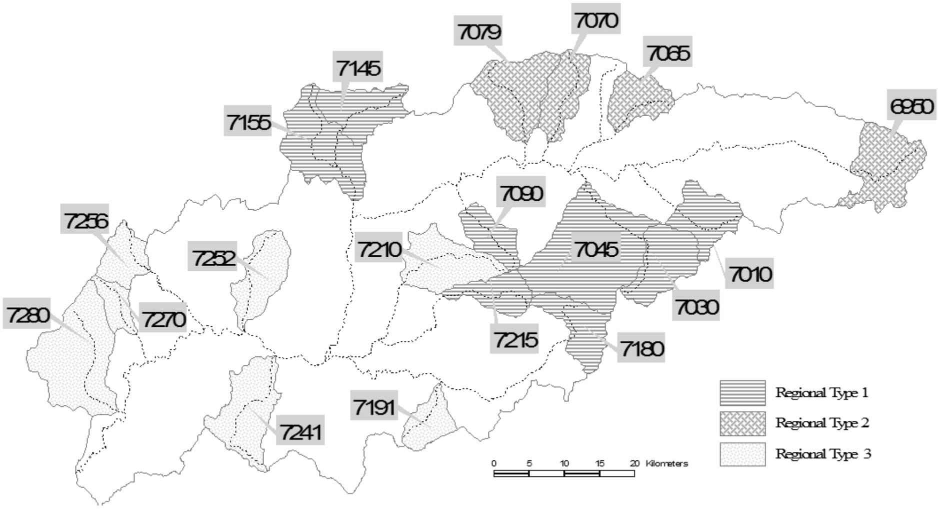

Fig. 2 The regional types of the Hron River basin.

[%] (FP), catchment average long-term mean annual temperature [ºC] (T), maximum daily precipitation totals [mm day^{-1}] (P_{max}), catchment average long-term mean annual precipitation totals [mm] (P), maximum snow cover depth with return period of 100 years [cm] (SC), long-term mean annual runoff [L s^{-1} km^{-2}] (Q).

RAINFALL–RUNOFF MODEL CALIBRATION

The Hron rainfall–runoff model, which was developed at the Slovak University of Technology in Bratislava (Hlavčová *et al.*, 2005) was used in this study. The hydrological balance in this conceptual model is based on the principles used in the HBV model (Bergström & Forsman, 1973). The model contains three basic storage components with 15 calibrated parameters. Surface and subsurface processes can be modelled separately for elevation zones and/or sub-catchments. The snow sub-model uses the degree-day method with refreezing for snow accumulation and snowmelt calculations. The sub-model for soil moisture simulation contains four parameters and calculates the soil water storage, input to the groundwater storage and actual evapotranspiration from the soil profile using a recognised relation between the water content in the soil profile and the field capacity value.

The runoff sub-model simulates both quick and slow runoff components (surface and subsurface runoff and baseflow). The basin runoff is calculated as the sum of all the partial runoffs, and it is routed by a nonlinear routing scheme consisting of a cascade of linear reservoirs with a discharge-dependent time parameter (multilinear cascade model). Input data needed for runoff simulation with a daily time step are the catchment's average mean daily precipitation values, catchment's average mean daily air temperature values, long-term mean monthly potential evapotranspiration, long-term mean monthly air temperature values and mean daily flows in the closing section of the catchment. A more detailed description of the model is given in Hlavčová *et al.* (2005).

The Hron model was calibrated in each watershed using data from the time period 1991–2000 using a genetic algorithm and the Nash-Sutcliffe model efficiency (ME hereafter) criterion. The time period 1981–1990 was left for validation of the model. The model efficiencies (ME) during the calibration period varied between 0.85 and 0.72. The best simulations (ME = 0.85) of observed runoff were achieved in the catchments located in the upper part of the basin. The validation period was best represented with the values 0.77, again in the upper part of the catchment; the lowest value was 0.51. The results mainly depended on the quality of the input data (mainly spatial distribution of raingauges). In general, lower values of ME were noticed in catchments where precipitation data were extrapolated from raingauge stations located outside the catchments.

REGIONALIZATION OF RAINFALL–RUNOFF MODEL PARAMETERS

With respect to rainfall–runoff model parameter regionalization it is assumed here, that to get more confidence in regional estimation of model parameters, catchments should form hydrologically homogeneous groups (Kokkonen, 2003). Therefore, in this study the gauged catchments in the region of interest were first divided into homogeneous groups with respect to selected physiographic characteristics using clustering assuming that similarity based on physiographic characteristics will result in similarity of model parameters and hydrological response. Hierarchic and non-hierarchic methods of clustering were tested; K-means clustering with Euclidean metrics with the same weight assigned to each characteristic in the clustering process was finally selected as suitable. The condition of non-colinearity between physiographic characteristics was controlled during clustering. Regionalization experiments based on logical hydrological reasoning with different combinations of physiographic characteristics were performed. Due to the relatively small number of catchments (19), emphasis was put on keeping the number of regional types and regionalisation variables small (max 3).

Finally catchments were divided into three regional types (clusters). Figure 2 presents the location of the sub-catchments splitting the Hron catchment into three regional types according to the mean catchment elevation (E), since this variable

Table 1 Summary statistics of the regional classification.

Variable	Between cluster mean square	Within cluster mean square	F-ratio
Elevation (E)	2.7528×10^5	2.2804×10^3	120.7115

Table 2 Statistical characteristics of the discriminating variable in the regional types.

Regional type	Mean	Range	Std deviation	Variation coeff.
1	848	156	45.5	0.054
2	1109	83	30.1	0.027
3	646	127	48.3	0.075

reflects the governing altitudinal zonality of some parameters influencing the runoff conditions in the region. The use of two and more discriminating variables has not led to regional types that were as clearly distinguishable, or were based on combinations with weak logical relationships to runoff generating conditions. Summary statistics of the K-means clustering classification is presented in Table 1, which compares the between-cluster mean square values to the within-cluster mean squares of the discriminating variable, and reports the F-ratio. Range, standard deviation, coefficient of variation of the discriminating variable within each of the regional type can be found in Table 2.

In this study, linear regression models were considered and stepwise multiple regression was used to determine the relationships between the climatic and physiographic basin characteristics and the model parameter values in the regional types. Regional regressions for model parameter estimation from physiographic characteristics were derived within each of the three regional types and for the whole Hron catchment. Different starting variables and variable sets were selected in a trial and error manner. Attention was paid to the minimisation of the effect of multicolinearity by choosing predictors with low interdependence. Several sets were considered, each run starting with a simple regression with the independent variable with the strongest influence on the dependent variables. Pearson correlation coefficients were chosen as a measure of dependence between the predictors. Several subjectively chosen and hydrologically reasonable starting combinations of the independent variables were used as seeds in the discrimination process. The values of the multiple correlation coefficients were, on average, about 0.85 for the resulting combinations of predictors. Different combinations of independent variables gave statistically comparable results. For the regional estimation of model parameters in selected regional types, finally the formulae with lowest standard error of estimate and highest multiple regression coefficient were selected and are summarized in Table 3.

Cumulative distribution curves of the values of the Nash-Sutcliffe model efficiency criterion (ME) for the calibration and validation period for at-site and regional parameter estimates are presented in Fig. 3. Using the regional multiple regression relationships for the regional types, ME values ranging between 0.8 and 0.43 for the calibration period, and between 0.76 and 0.33 for the validation period, were achieved. However, we must take into account that independent validation (e.g. jack-knife) was not used in this case due to the limited data available. Average model efficiencies for equations valid for the whole Hron region reached 0.56 for the

Table 3 Multiple regression coefficients, standard errors of estimate and regression relationships valid for regional types of the Hron River basin.

Parameter	Regional type	Multiple R	Std error of estimate	Regional regression formulae
	1	0.92	0.12	DDF = 2.90 – 1.189 RND + 0.0023 SC
DDF	2	0.98	0.08	DDF = –10.38 + 0.0104 E
	3	0.86	0.26	DDF = 5.25 – 0.1151 SL – 0.0025 E
	1	0.75	0.17	TT = 0.79 – 0.9975 FP – 2.1723 STI
TT	2	0.98	0.19	TT = 1.66 – 0.0392 F
	3	0.79	0.22	TT = 0.87 – 0.0462 Q – 0.0045 SC
	1	0.90	0.01	WHC = –0.01 + 0.0005P_{max} + 0.0005TC
WHC	2	1.00	0.00	WHC = 0.08 – 0.0001 P
	3	0.97	0.01	WHC = –0.45 + 0.0471 TC + 0.0603 T + 0.0018 Q
	1	0.66	0.07	RFC = 0.90 + 0.0114 SL – 0.1931 FP – 0.0007 P
RFC	2	0.87	0.12	RFC = –1.40 + 0.0082 SC
	3	0.79	0.05	RFC = 0.04 + 0.0011 F + 0.0007 SC
	1	0.86	0.09	SCF = 0.21 + 0.4776 FP + 0.0053 P_{max} – 0.0004 SC
SCF	2	0.89	0.12	SCF = –0.35 + 0.0142 P_{max}
	3	0.77	0.20	SCF = –1.28 + 0.0025 P
	1	0.63	50.98	FC = 40.45 + 195.8586 FP + 130.7068 Sh
FC	2	0.94	14.87	FC = 444.17 – 11.605 Q
	3	0.87	33.16	FC = 741.84 – 135.3094 RND – 0.3126 P
	1	0.89	0.08	LPE = –1.25 + 0.0023E – 0.3936 Sh
LPE	2	0.79	0.16	LPE = –0.51 + 0.0118 P_{max}
	3	0.93	0.12	LPE = –0.38 + 0.0023 P – 0.0109 P_{max}
	1	0.63	0.46	RC = 3.72 – 0.4146 Tc
RC	2	0.98	0.16	RC = 6.66 – 0.2063 Q
	3	0.92	0.40	RC = 8.45 – 0.3092 SL – 0.049 MSG
	1	0.65	0.36	EMP = 0.21 + 0.0018 A – 0.0761 MSG
EMP	2	0.99	0.16	EMP = 1.19 – 0.0213 F + 0.0081 A
	3	0.80	0.53	EMP = 3.05 + 0.1952 MSG – 0.0386 P_{max}
	1	0.75	0.22	K0 = 1.05 – 4.0914 STI + 1.1447 FP
K0	2	0.94	0.16	K0 = –1.20 + 0.0835 SL
	3	0.74	0.22	K0 = 0.82 + 0.0053 SC – 0.0095 P_{max}
	1	0.85	0.13	K1 = 0.88 – 0.521 Sh
K1	2	0.67	0.03	K1 = 0.23 – 0.0012 F
	3	0.87	0.08	K1 = 0.26 + 0.0422 SL – 0.0049 P_{max}
	1	0.78	0.04	K2 = 0.19 – 0.1104 Sh + 0.0075 MSG
K2	2	0.93	0.01	K2 = 0.11 – 0.0007 F – 0.0128 SH
	3	0.97	0.01	K2 = 0.24 – 0.0002 A – 0.0007 F
	1	0.61	20.41	UZL = 58.53 + 216.9779 STI – 0.3349 HgTI
UZL	2	0.91	11.76	UZL = 102.42 + 2.5281 SL - 0.4929 A
	3	0.96	11.65	UZL = 132.97 – 182.6016 SH + 7.6294 MSG – 0.7287 P_{max}
	1	0.97	0.57	PER = –6.76 + 0.3683 SL + 0.0248 HgTI
PER	2	0.92	1.04	PER = –3.86 + 0.4496 SL
	3	0.92	0.67	PER = –13.09 + 0.1295 HgTI – 0.0963 P_{max}
	1	0.71	0.39	MB = 3.08 – 0.005 F
MB	2	1.00	0.00	MB = 3.00
	3	0.94	0.16	MB = 3.31 + 2.1775 FP – 0.0194 P_{max}

Legend: DDF - degree-day factor [mm °C^{-1} day^{-1}], TT - threshold air temperature [ºC], WHC - water holding capacity [–], RFC - refreezing coefficient [–], SCF - snow correction factor [–], FC - maximum field capacity [mm], LPE - limit for potential evapotranspiration [–], RC - recharge coefficient [–], Emp - empirical parameter for evapotranspiration [-], K0 - recession coefficient for surface runoff [–], K1 - recession coefficient for subsurface runoff [–], K2 - recession coefficient for baseflow [–], UZL - limit for upper zone [mm], PER - percolation parameter [mm day^{-1}], MB - parameter of runoff retardation [day].

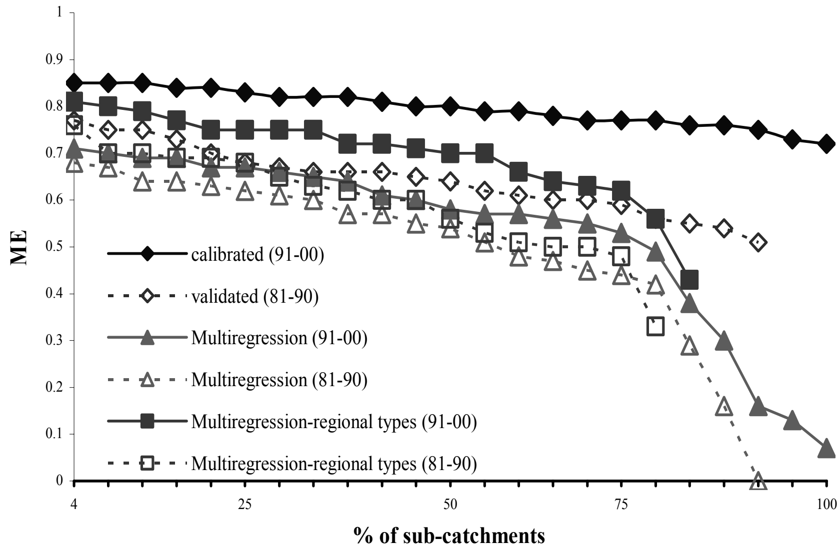

Fig. 3 Cumulative distribution curves of model efficiencies of calibrated and regionalized model parameter values (regional multiple regressions).

calibration period and 0.58 for the validation period. The mean ME values for the equations for regional types increased to 0.69 for the calibration period, and to 0.61 for the validation period. The model parameters derived from regressions for separate regional types lead to higher values of ME in all subcatchments when compared to model simulations achieved by parameters obtained for the whole catchment (Fig. 3).

CONCLUSIONS

Multiple regression methods for calculating the model parameter values according to catchment physiographic characteristics are less effective according to the results of comparisons published by Parajka *et al.* (2005) and Merz & Bloeschl (2004). The results in a similar study by Zvolensky (2006) supported these findings. The model parameters derived from separate regressions for regional types lead to higher values of ME in all subcatchments when compared to model simulations achieved by parameters obtained for the whole catchment. The explanation could be that parameters have weak physical meaning and the regression relationships may be accidental. The methods presented in this paper can be used to improve their performance. Improvements occurred when catchments were classified into regional types. Catchments within a regional type are supposed to have similar hydrological behaviour, but distinct from other regional types. Regional regression relationships for model parameter calculation derived separately for each of the regional types may thus better represent the real hydrological behaviour as was also suggested by Kokkonen *et al.* (2003).

Acknowledgements This work was supported by the Slovak Grant Agency under VEGA project nos 1/2032/05 and 1/4024/07, and the EU Commission project HYDRATE, no. FP6-2006-STREP-037024. The support is gratefully acknowledged.

REFERENCES

Abdulla, F. A. & Lettenmaier, D. P. (1997) Application of regional parameter estimation schemes to simulate the water balance of a large continental river. *J. Hydrol*. **197**, 258–285.

Bergström, S. & Forsman, A. (1973) Development of a conceptual deterministic rainfall-runoff model. *Nordic Hydrol*. **4**, 147–170.

Hlavčová, K., Kohnová, S., Zvolenský, M., Kubeš, R. & Szolgay, J. (2005) An empirical method for estimating future flood risk for flood warnings. *Hydrol. Earth System Sci*. **9**(4), 431–488.

Kokkonen, T. S., Jakeman, A. J., Young, P. C. & Koivusalo, H. J. (2003) Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. *Hydrol. Processes* **17**, 2219–2238.

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

Nash, J. E. (1960) A unit hydrograph study with particular reference to British catchments. *Proc. Inst. Civil Engrs* **17**, 249– 282.

Parajka, J., Merz, R. & Bloesch, G. (2005) A comparison of regionalisation methods for catchment model parameters. *Hydrol. Earth System Sci*. **9**, 157–171.

Ross, G. A. (1970) The Stanford watershed model: the correlation of parameter values selected by a computized procedure with measurable physical characteristics of the watershed. *Research Report 35*. Water Resources Institute, University of Kentucky, Lexington, USA.

Sefton, C. E. M. & Howarth, S. M. (1998) Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales. *J. Hydrol*. **211**, 1–16.

Szolgay, J., Hlavčová, K., Kohnová, S. & Danihlík, R. (2003) Regional estimation of parameters of a monthly balance model. *J. Hydrol. Hydromech*. **51**(4), 256–273.

Xu, C. Y. (1999) Estimation of parameters of a conceptual water balance model for ungauged catchments. *Water Resour. Manage*. **13**, 353–368.

Zvolensky, M. (2006) Estimation of rainfall–runoff model parameters in ungauged catchments of the Hron River Basin. PhD Thesis, SvF STU Bratislava, Slovakia.