HYDROLOGICAL PROCESSES
Hydrol. Process. **21**, 435–446 (2007)
Published online 2 November 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/hyp.6253

Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments

J. Parajka,*,† R. Merz and G. Blöschl
Institute for Hydraulic and Water Resources Engineering, Vienna University of Technology, Karlsplatz 13/222, A-1040 Vienna, Austria

Abstract:

We examine the value of additional information in multiple objective calibration in terms of model performance and parameter uncertainty. We calibrate and validate a semi-distributed conceptual catchment model for two 11-year periods in 320 Austrian catchments and test three approaches of parameter calibration: (a) traditional single objective calibration (SINGLE) on daily runoff; (b) multiple objective calibration (MULTI) using daily runoff and snow cover data; (c) multiple objective calibration (APRIORI) that incorporates an *a priori* expert guess about the parameter distribution as additional information to runoff and snow cover data. Results indicate that the MULTI approach performs slightly poorer than the SINGLE approach in terms of runoff simulations, but significantly better in terms of snow cover simulations. The APRIORI approach is essentially as good as the SINGLE approach in terms of runoff simulations but is slightly poorer than the MULTI approach in terms of snow cover simulations. An analysis of the parameter uncertainty indicates that the MULTI approach significantly decreases the uncertainty of the model parameters related to snow processes but does not decrease the uncertainty of other model parameters as compared to the SINGLE case. The APRIORI approach tends to decrease the uncertainty of all model parameters as compared to the SINGLE case. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS multiple objective calibration; parameter uncertainty; water balance modelling

Received 20 May 2005; Accepted 4 November 2005

INTRODUCTION

Knowledge of the regional variability of water balance components is important for solving a range of problems in water resources management and planning. Simulations of the water balance components are, however, fraught with a range of problems, including uncertainty in inputs, model parameters and model structure. These problems are particularly acute in Alpine regions, where data are sparse and the spatial variability of the hydrological environment is enormous.

Although modellers have always been aware of model parameter uncertainty it is only in the last few decades that explicit efforts have been made towards assessing this uncertainty. Because of multiple optima, non-linear interactions between model parameters and data errors, it may be difficult, if not impossible, to identify a unique parameter set from runoff data. Methods for assessing parameter uncertainty in hydrologic models include the generalized likelihood uncertainty estimation methodology (e.g. Beven and Binley, 1992; Beven and Freer, 2001), multi-normal approximations to parameter uncertainty (e.g. Kuczera and Mroczkowski, 1998), and Markov Chain Monte Carlo methods (e.g. Kuczera and Parent, 1998).

Although these methods are useful in identifying the degree of confidence one can attribute to a calibrated parameter set they do not reduce their uncertainty. For reducing parameter uncertainty the general line of thought has been that information additional to runoff data needs to be used to constrain the parameters over what can be achieved from calibrations to runoff alone. Various types of catchment response data can be used depending on the application at hand. Seibert (2000) and Madsen (2003) used runoff data and groundwater level data jointly to calibrate model parameters. They found that the groundwater data reduced the uncertainty of the parameters representing groundwater dynamics significantly. Beldring (2002) found that the inclusion of groundwater levels reduced the uncertainty of most of the model parameters. Other types of data that can be used in constraining the uncertainty of model parameters are snow cover and soil moisture data. The value of snow cover data in distributed hydrologic simulations has been demonstrated by Blöschl *et al.* (1991) and others. Grayson and co-workers (Grayson and Blöschl, 2000; Grayson *et al.*, 2002) summarize numerous examples of using snow and soil moisture data in addition to runoff and suggest that these response data are particularly useful if available as spatial patterns. Isotopes and geochemical characteristics, such as stream chloride, have also been used for identifying model parameters

* Correspondence to: J. Parajka, Institute for Hydraulic and Water Resources Engineering, Vienna University of Technology, Karlsplatz 13/222, A-1040 Vienna, Austria. E-mail: parajka@hydro.tuwien.ac.at
† On leave from the Institute of Hydrology, Slovak Academy of Sciences, Bratislava, Slovakia.

(e.g. Holko and Lepistö, 1997; Mroczkowski *et al.*, 1997).

As an alternative to measurements, the use of 'soft data' or qualitative information from field surveys has been suggested to constrain model parameters. 'Soft' information is widely used in practical applications of catchment models where parameters are selected based on all sources of information available to the analyst (e.g. Blöschl, 2005). A recent contribution of using this type of qualitative expert knowledge has been provided by Seibert and McDonnell (2002).

Formal methods of incorporating information in addition to runoff in the calibration process are usually referred to as multiple objective calibration. The various objectives (related to runoff, groundwater levels, snow cover, etc.) can be combined in various ways. The most straightforward combination is by a weighted sum, as in the weighted least-squares method used for calibrating groundwater models (e.g. Peck *et al.*, 1988: 50). An alternative is the use of fuzzy logics, where the components of the objective function are combined based on membership functions that indicate the relative degree of satisfaction of each fuzzy objective (Seibert, 1997; Franks *et al.*, 1998; Yu and Yang, 2000; Cheng *et al.*, 2002). A third method is based on the concept of Pareto optimality. A parameter set is considered Pareto optimal if there is no other parameter set that performs at least as well on every objective and strictly better on at least one objective. That is, a Pareto-optimal solution cannot be improved upon without hurting at least one of the objectives (Miettinen, 1999). Madsen (2003) proposed a Pareto-based approach that emulates the ability of manual expert calibration of using a number of complementary ways in evaluating model performance. The method provided generally better simulations of runoff compared with manual expert calibration but virtually similar performance for groundwater level simulations. Vrugt *et al.* (2003b) proposed an optimization technique termed the multi-objective shuffled complex evolution Metropolis (MOSCEM-UA) algorithm that provides an estimate of the Pareto solution space within a single optimization run. They found that the MOSCEM-UA algorithm generates a fairly uniform approximation of the entire set of Pareto parameter combinations for their problem. Although various methods exist for making use of multiple data sources, the actual merits of additional information have, to our knowledge, never been identified for a large number of catchments in the context of regional water balance modelling.

The aim of this paper, therefore, is to assess the value of snow data in addition to runoff data as well as the value of expert judgement in multiple objective calibration of a catchment model. Specifically, we address two science questions: (a) How will the multiple objective calibration change the model performance over single objective calibration? (b) To what extent does multiple objective calibration reduce model parameter uncertainty over single objective calibration? One would expect that the use of additional snow data will improve the snow cover simulations and decrease the model performance with respect to runoff. The aim of this paper is to assess the extent of such change. We use daily hydrologic data from 320 catchments over a period of 22 years, which will likely allow us to draw more generic inferences than has been possible in most previous studies that used smaller data sets.

DATA

This study was carried out in Austria using data from the period 1976–1997. Austria is flat or undulating in the east and north, and Alpine in the west and south. Elevations range from 115 to 3797 m a.s.l. Mean annual precipitation is less than 400 mm in the east and almost 3000 mm in the west. Land use is mainly agricultural in the lowlands and forest in the medium elevation ranges. Alpine vegetation and rocks prevail in the highest catchments. The dataset used in this study includes measurements of daily precipitation and snow depths at 1091 stations and daily air temperature at 212 climatic stations. To calibrate and verify a catchment model, daily runoff data from 320 gauged catchments were used with areas ranging from 10 to 9770 km^2 and a median of 196 km^2. Of these, 97 catchments range in area between 10 and 100 km^2, 106 catchments between 100 and 300 km^2, 64 catchments between 300 and 1000 km^2 and 55 catchments have areas of more than 1000 km^2. In preliminary analyses we carefully screened the runoff data for errors and removed all stations with significant anthropogenic effects. We also removed stations where we were not able to close the long-term water balance. The spatial distribution of the climate stations and the boundaries of the gauged catchments are shown in Figure 1.

The inputs to the water balance model were prepared in two steps. First, the daily values of precipitation, snow depth and air temperature were spatially interpolated by methods that use elevation as auxiliary information. External drift kriging was used for precipitation and snow depths, and the least-squares trend prediction method was used for air temperatures (Pebesma, 2001). The spatial distribution of potential evapotranspiration was estimated by a modified Blaney–Criddle method (Schrödter, 1985; Parajka *et al.*, 2003) using daily air temperature and potential sunshine duration calculated by the Solei-32 model (Mészároš *et al.*, 2002; http://www.ih.savba.sk/software/solei/) that incorporates shading by surrounding terrain. In a second step, a digital elevation model with a 1 km × 1 km grid resolution was used for deriving 200 m elevation zones in each catchment. Time-series of daily precipitation, air temperature, potential evaporation and snow depth were then extracted for each of the elevation zones to be used in the water balance simulations.

METHODS

The model used in this paper is a semi-distributed conceptual rainfall-runoff model, following the structure

DOI: 10.1002/hyp

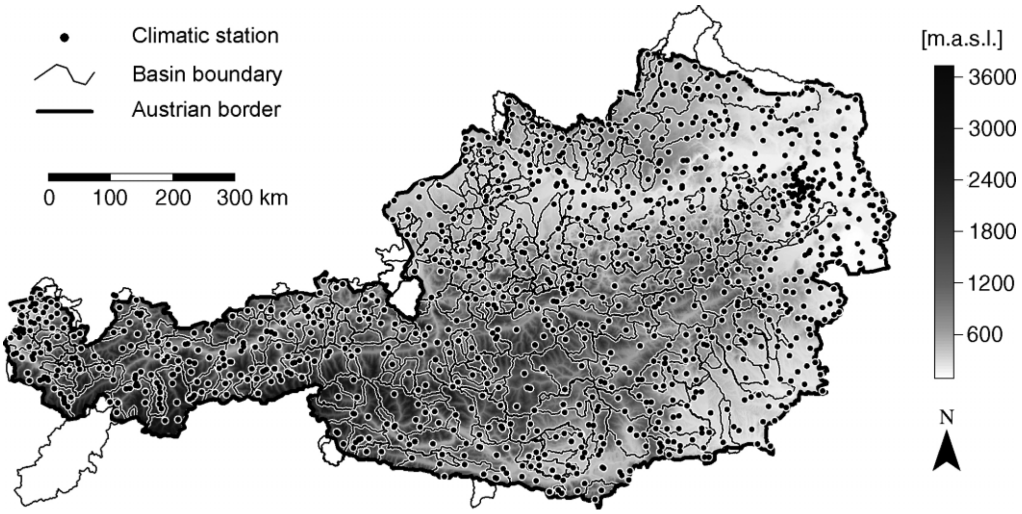

Figure 1. Map of Austria, including boundaries of gauged catchments and stations with precipitation and snow depth measurements

of the HBV model (Bergström, 1976; Lindström *et al.*, 1997). The model runs on a daily time step and consists of a snow routine, a soil moisture routine and a flow routing routine (Merz and Blöschl, 2004). A flow chart of the model is presented in Merz (2002). The snow routine represents snow accumulation and melt by a simple degree-day concept, using a degree-day factor DDF and a melt temperature parameter T_M. The catch deficit of precipitation gauges during snowfall is corrected by a snow correction factor SCF. A threshold temperature interval $T_R - T_S$ is used to distinguish between rainfall, snowfall and a mix of rain and snow. The soil moisture routine represents runoff generation and changes in the soil moisture state of the catchment and involves three parameters: the maximum soil moisture storage FC, a parameter representing the soil moisture state above which evaporation is at its potential rate, termed the limit for potential evaporation LP, and a parameter in the non-linear function relating runoff generation to the soil moisture state, termed the non-linearity parameter β. Runoff routing on the hillslopes is represented by an upper and a lower soil reservoir. Excess rainfall enters the upper zone reservoir and leaves this reservoir through three paths: outflow from the reservoir based on a fast storage coefficient K_1; percolation to the lower zone with a constant percolation rate C_P; and, if a threshold of the storage state LS_{UZ} is exceeded, through an additional outlet based on a very fast storage coefficient K_0. Water leaves the lower zone based on a slow storage coefficient K_2. The outflow from both reservoirs is then routed by a triangular transfer function representing runoff routing in the streams, where C_R is a free parameter. More details on the model are given in Appendix A.

The model is run for all 320 gauged catchments in Austria. Inputs (precipitation, air temperature and potential evapotranspiration) are allowed to vary with elevation within a catchment, so the soil moisture accounting and snow accounting is performed independently in each elevation zone of the 200 m altitudinal range. However, the same model parameters are assumed to apply to all elevation zones of a catchment. These parameters (14 in total) are estimated by model calibration.

We tested three cases of model calibration in this paper. The first case conforms to the most widely used procedure in hydrology, where the model parameters are adjusted in a way that runoff simulations closely match measured runoff. This case is termed SINGLE in this study. The runoff objective function Z_Q follows the relation proposed by Lindström (1997), which combines the Nash–Sutcliffe efficiency ME and the relative volume error VE and is defined as

$$Z_Q = (1 - \mathrm{ME}) + w\mathrm{VE} \quad (1)$$

where

$$\mathrm{ME} = 1 - \frac{\sum_{i=1}^{n}(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i})^2}{\sum_{i=1}^{n}(Q_{\mathrm{obs},i} - \overline{Q_{\mathrm{obs}}})^2} \quad (2)$$

$$\mathrm{VE} = \frac{\sum_{i=1}^{n} Q_{\mathrm{sim},i} - \sum_{i=1}^{n} Q_{\mathrm{obs},i}}{\sum_{i=1}^{n} Q_{\mathrm{obs},i}} \quad (3)$$

$Q_{\mathrm{sim},i}$ is the simulated runoff on day i, $Q_{\mathrm{obs},i}$ is the observed runoff, $\overline{Q_{\mathrm{obs}}}$ is the average of the observed runoff over the calibration (or verification) period of n days, and the weight $w = 0{\cdot}1$ was found in test simulations to give the most plausible results.

The second case, referred to as MULTI in this study, uses snow data in addition to runoff. In this study we did not directly compare the observed snow depths with simulated snow water equivalent, because the snow density measurements necessary for such a comparison were not available. Observed snow cover, therefore, was estimated from daily grid maps constructed from the observed snow depth data. If the catchment zone average

DOI: 10.1002/hyp

of snow depth was greater than 0·5 mm then the zone was considered as snow covered, otherwise it was considered as snow free. Simulated snow cover was derived from the snow water equivalent simulated by the model: a zone was considered snow covered if the water equivalent was greater than 0·1 mm; otherwise, it was considered snow free. The thresholds of 0·5 and 0·1 mm were set a little above zero to avoid snow cover overestimation that may result from the spatial interpolation. For local snowfall events the interpolation of snow depth measurements into neighbouring regions, where no snow was observed, may lead to small values of catchment (or elevation zone) snow depth average that, however, should be considered as no snow. Snow simulations on a particular day were considered to be poor if the absolute difference between simulated and observed snow cover was greater than 50% of the catchment area. The 50% threshold was determined in sensitivity analysis (not shown here) taking into account different areal arrangements of elevation zones in different catchments, where the sensitivity was assessed on the basis of model performance. The snow objective function Z_S was then defined as the ratio of the number of days with poor snow cover simulation n_{ps} to the total number of days in the simulation period n:

$$Z_S = \frac{n_{ps}}{n} \tag{4}$$

The third calibration arrangement, termed APRIORI, accounts not only for the runoff and snow cover objectives, but also incorporates an *a priori* expert judgement about the expected distribution of each model parameter. In calibration procedures, the parameter values are usually bounded between two limits (Duan *et al.*, 1992) and otherwise no *a priori* assumptions are made about the parameters. This implies that the *a priori* distribution of the parameters is a uniform distribution. We believe that it is possible to make a more informed guess about the shape of the *a priori* distribution and introduced a penalty function Z_P:

$$Z_P = \sum_{j=1}^{k} \frac{f_{\max,j} - f_j\left(\dfrac{p_j - p_{l,j}}{p_{u,j} - p_{l,j}}\right)}{f_{\max,j}} \tag{5}$$

$$f_{\max,j} = f_j\left(\frac{p_{\max,j} - p_{l,j}}{p_{u,j} - p_{l,j}}\right) \tag{6}$$

where p_j is the model parameter j to be calibrated, p_l and p_u are the lower and upper bounds of the parameter space respectively, p_{max} is the parameter value at which the *a priori* distribution is at a maximum and k is the number of parameters to be calibrated. The probability density function of the Beta distribution is

$$f(x|U,V) = \frac{1}{B(u,v)} x^{u-1}(1-x)^{v-1} \quad \text{for } 0 < x < 1, u > 0, v > 0 \tag{7}$$

with

$$B(u,v) = \int_0^1 x^{u-1}(1-x)^{v-1}dx = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)} \tag{8}$$

We assumed values of u, v, p_l, p_u and p_{max} for each parameter j based on our own assessment of the hydrologic characteristics of the study region and on literature values (Bergström, 1992; Seibert, 1997). The order of magnitude of the p_{max} was consistent with values found by modelling studies in the region (Merz and Blöschl, 2004). In the absence of more detailed information we have chosen the same values of u, v, p_l, p_u and p_{max} (Table I) for all catchments. If more detailed information was available (e.g. from catchment attributes or from field studies), then the limits and parameters of the Beta distributions for each model parameter could be assigned differently from catchment to catchment. The resulting Beta distribution functions are shown in Figure 2.

We calibrated the rainfall-runoff model for 320 catchments using two automatic calibration methods: the MOSCEM-UA (Vrugt *et al.*, 2003a) and the SCE-UA

Table I. *A priori* distribution of parameter values. u and v are parameters of the Beta function (Equation (7)), p_l and p_u are the lower and upper bounds of the parameter space and p_{max} is the parameter value at which the Beta distribution is at a maximum (Equation (5))

Parameter	Model part	u	v	p_l	p_u	p_{max}
SCF	Snow	1·2	4·0	1·0	1·5	1·03
DDF (mm °C^{-1} day^{-1})	Snow	2·0	4·0	0·0	5·0	1·25
T_R (°C)	Snow	2·0	4·0	1·0	3·0	1·5
T_S (°C)	Snow	2·0	4·0	−3·0	1·0	−2·0
T_M (°C)	Snow	3·0	3·0	−2·0	3·0	0·5
LP/FC	Soil	4·0	1·2	0·0	1·0	0·94
FC (mm)	Soil	1·1	1·5	0·0	600	100
β	Soil	1·1	1·5	0·0	20	3·4
K_0 (day)	Runoff	2·0	4·0	0·0	2·0	0·5
K_1 (day)	Runoff	2·0	4·0	2·0	30	9·0
K_2 (day)	Runoff	1·05	1·05	30	180	105
C_P (mm day^{-1})	Runoff	2·0	4·0	0·0	8·0	2·0
C_R (day^2 mm^{-1})	Runoff	1·05	1·05	0·0	50	25
LS_{UZ} (mm)	Runoff	3·0	3·0	1·0	100	50

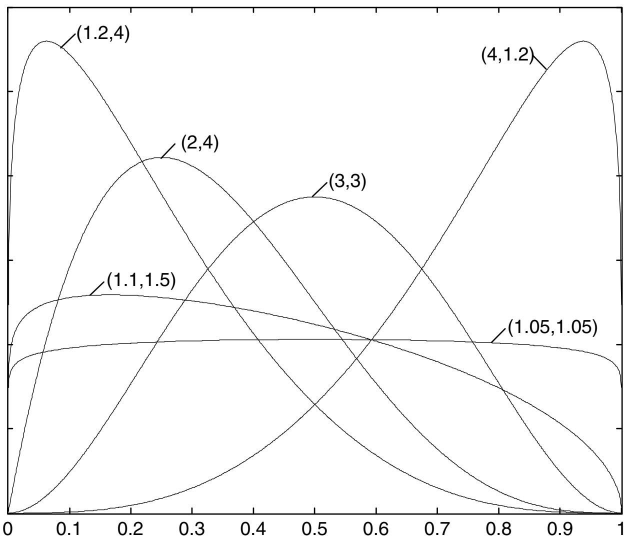

Figure 2. Shapes of the Beta functions used for defining the *a priori* distributions of the model parameters. Number in parentheses are u and v (see Table I)

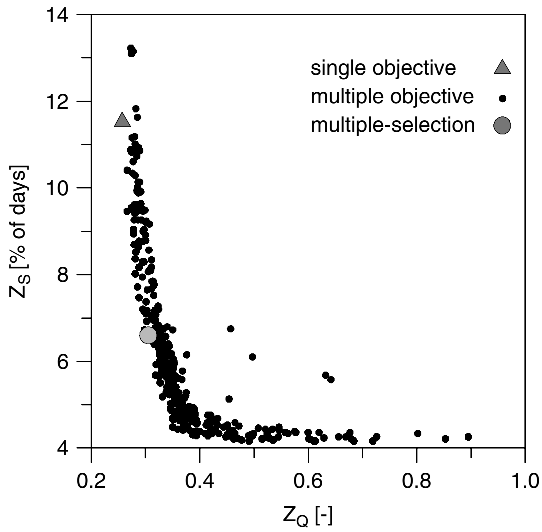

Figure 3. Pareto solutions in the two-dimensional objective function space obtained by the MOSCEM-UA method; Wienerbruck catchment is shown as an example. The small black dots correspond to 1500 Pareto solutions, the filled triangle represents the minimum runoff objective function (SINGLE) and the large filled circle shows the compromise solution (MULTI) between the runoff and snow objective functions, Z_Q and Z_S

(Duan *et al.*, 1992) methods. The MOSCEM-UA algorithm we applied to the SINGLE and MULTI calibration cases. As a result of the calibration, the MOSCEM-UA provides a discrete set of possible parameter combinations (a Pareto set) that represent tradeoffs between optimal ways of constraining the model to be consistent with observed daily runoff and snow cover data. From the Pareto set solutions we selected two parameter combinations. The single criterion end point in the Pareto set with respect to the minimum Z_Q represents the parameter combination used for the SINGLE calibration case. The second parameter combination represents a compromise solution between the runoff and snow cover objective functions and is used for the MULTI calibration case. We selected this parameter set in a way that yielded the minimum value of

$$Z_{\mathrm{MULTI}} = 0{\cdot}8Z_Q + 0{\cdot}2\frac{Z_{\mathrm{S}}}{Z_{\mathrm{S,MAX}}} \quad (9)$$

The weights (0·8 and 0·2) were determined by test simulations and gave a relative importance of 80% to Z_Q and 20% to Z_S on average over the 320 catchments.

An example of the Pareto solution space for the Wienerbruck catchment is presented in Figure 3. The triangle in Figure 3 represents the parameter set selected for the SINGLE case and the big black circle represents the parameter set selected for the MULTI case.

For three objective functions, the MOSCEM-UA method can be numerically quite taxing as it evaluates the entire three-dimensional Pareto solution space. For the APRIORI case, therefore, we used the SCE-UA calibration algorithm instead, which is numerically more efficient, and minimized one compound objective function Z_C:

$$Z_{\mathrm{C}} = w_1Z_Q + w_2Z_{\mathrm{SC}} + w_3Z_{\mathrm{P}} \quad (10)$$

where the part representing the snow objective function Z_{SC} is defined as the ratio of number of days with poor snow cover simulation n_{ps} to the number of days with observed snow cover n_{os} expressed as

$$Z_{\mathrm{SC}} = Z_{\mathrm{S}}\frac{n}{n_{\mathrm{os}}} \quad (11)$$

The weights in Equation (10) were assigned in test simulations as $w_1 = 0{\cdot}7$, $w_2 = 0{\cdot}2$ and $w_3 = 0{\cdot}1$. These test simulations consisted of sensitivity analyses that showed that the model results were only moderately sensitive to the choice of weights. The selection of weights is arbitrary and always depends on subjective user requirements and expectations. In this paper we estimated the weights so that, on average, the runoff Z_Q, snow Z_{SC} and *a priori* penalty Z_{P} contribute to the final compound objective function Z_{C} with 40%, 40% and 20% respectively.

The evaluation of the calibration and verification efficiencies of the SINGLE, MULTI and APRIORI optimization approaches, along with the comparison of their parameter uncertainties, were performed in two steps. In a first analysis, we split the entire period of observations (1976–1997) into two 11-year periods: from 1 January 1976 to 31 December 1986 and from 1 January 1987 to 31 December 1997. Warm-up periods from January to October were used in both cases. For the efficiency estimation and comparison between different calibration procedures, we performed a split sample test in the terminology of Klemeš (1986). We used the 11-year periods in turn for calibration and validation, and compared the model performances from both arrangements. In a second analysis, we judged the parameter uncertainty for the SINGLE, MULTI and APRIORI methods by comparing the parameters calibrated for the 1976–1986 period with those calibrated for the 1987–1997 period. The comparison of parameter values obtained in two different periods

DOI: 10.1002/hyp

gives the total amount of uncertainty of these parameters, including uncertainty due to input data and model structure.

CALIBRATION AND VERIFICATION EFFICIENCIES

The performance of the single- and multiple-objective parameter optimization approaches is presented in terms of their efficiency to simulate runoff ME (Equation (2)), runoff volume errors VE (Equation (3)) and snow cover errors Z_S (Equation (4). For a favourable model performance, the ME runoff efficiencies should be large, the VE volume errors should be close to zero with a small scatter and the Z_S snow cover errors should be small.

Table II and Figures 4 and 5 (left panels) show the ME model efficiencies of the SINGLE, MULTI and APRIORI optimization approaches. The statistical evaluation in Table II includes the median and, as a measure of scatter, the differences of the 75th and 25th percentiles over all 320 catchments. The SINGLE optimization approach yields median model efficiencies of ME = 0·74 and 0·75 for the two periods when they are used for calibration. Incorporating snow cover data in the MULTI approach slightly decreases the median efficiencies to ME = 0·72 and 0·74. However, the incorporation of both snow cover data and an expert judgement about the expected parameter values in the APRIORI optimization scheme yields the same ME performance as the SINGLE approach. This is because, in the APRIORI objective function, less weight is given to snow than in the MULTI case. This result suggests that the use of *a priori* information on parameters does not necessarily decrease model efficiency. This is a typical result, which, of course, depends on the weights chosen in the objective function. For the verification cases, there is a slight decrease in runoff model efficiency for all methods compared with the calibration case, but the relative performance of the methods remains similar. It is interesting that in the APRIORI case some of the lowest model efficiencies improve over the SINGLE case (Figure 5, bottom left panel), whereas the opposite seems to be true of the MULTI case (Figure 5, top left panel).

The runoff volume errors VE are shown in Table III and Figures 4 and 5 (top centre and bottom centre panels). The median VE values around 0% for the calibration periods indicate that the calibration is essentially unbiased for all calibration procedures. In the verification periods, the median biases are around ±5%, which is likely related to data issues. The scatter is around 3–4% in the calibration periods and increases to around 11%

Table II. Model efficiency ME of runoff according to Nash–Sutcliffe for the calibration and verification periods using the SINGLE, MULTI and APRIORI calibration approaches

	ME[a]		
	SINGLE	MULTI	APRIORI
Calibration 1976–1986	0.74/0.17	0.72/0.18	0.74/0.17
Calibration 1987–1997	0.75/0.12	0.74/0.12	0.75/0.12
Verification 1987–1997	0.70/0.13	0.70/0.14	0.71/0.13
Verification 1976–1986	0.68/0.20	0.64/0.20	0.68/0.19

[a] First value: median of ME efficiency over the 320 catchments. Second value: difference of the 75% and 25% quantiles of model efficiencies, i.e. a measure of scatter. High model performances are associated with large medians and a small scatter

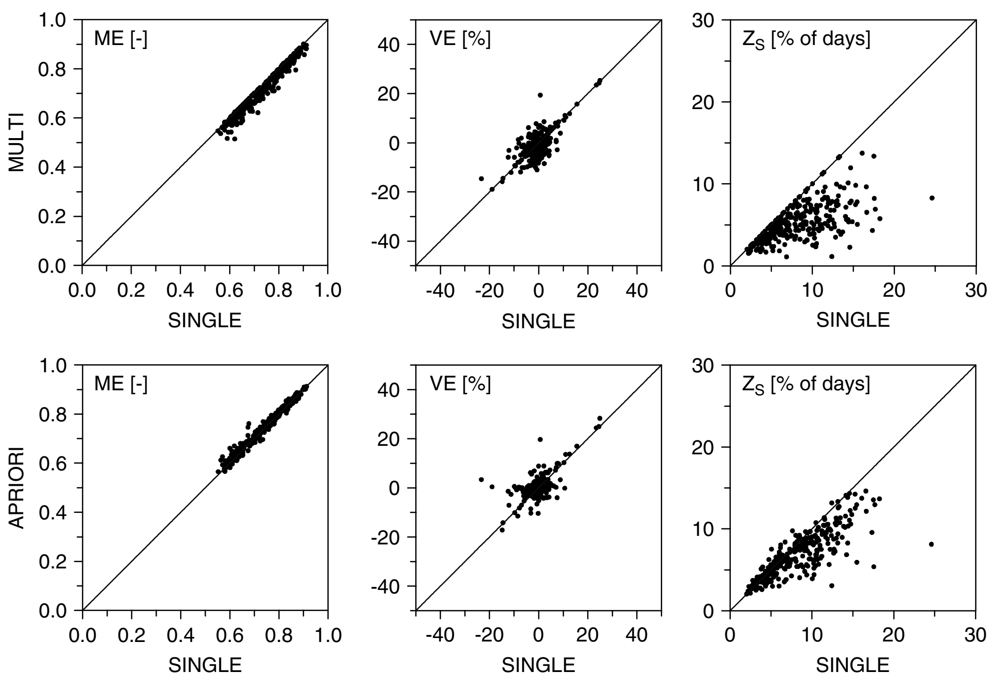

Figure 4. Comparison of the model efficiency of daily runoff ME, runoff volume error VE and snow over error Z_S estimated by the single objective (SINGLE) and the multiple objective (MULTI and APRIORI) calibration approaches. Each point in a panel relates to one out of 320 catchments for the calibration period 1976–1986

DOI: 10.1002/hyp

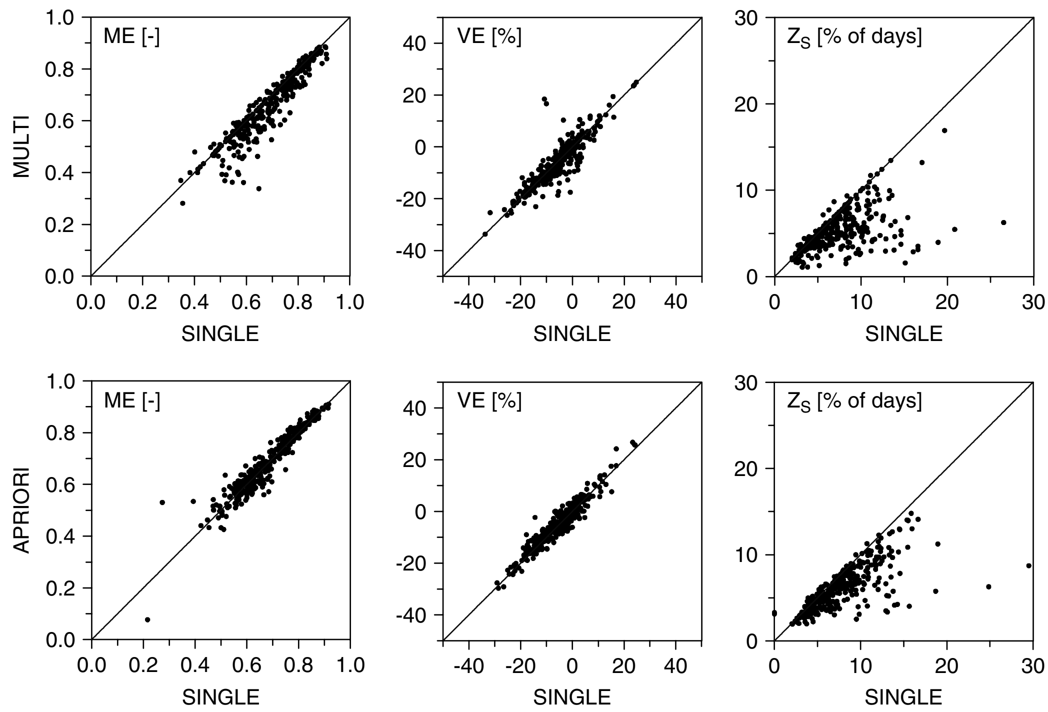

Figure 5. Comparison of the model efficiency of daily runoff ME, runoff volume error VE and snow cover error Z_S estimated by the single objective (SINGLE) and the multiple objective (MULTI and APRIORI) calibration approaches. Each point in a panel relates to one out of 320 catchments for the verification period 1976–1986

Table III. Volume errors VE of runoff for the calibration and verification periods using the SINGLE, MULTI and APRIORI calibration approaches

	VE[a] (%)		
	SINGLE	MULTI	APRIORI
Calibration 1976–1986	−0.4/3.3	−0.8/4.2	−0.1/1.9
Calibration 1987–1997	0.1/3.6	−0.3/4.2	0.3/2.8
Verification 1987–1997	5.7/11.5	4.9/11.5	6.3/12.0
Verification 1976–1986	−5.5/10.7	−6.5/10.5	−5.0/10.1

[a] First value: median of VE over the 320 catchments. Second value: difference of the 75% and 25% quantiles of VE, i.e. a measure of scatter. High model performances are associated with medians close to zero and a small scatter.

Table IV. Snow cover simulation errors Z_S for the calibration and verification periods using the SINGLE, MULTI and APRIORI calibration approaches

	Z_S (percentage of days)		
	SINGLE	MULTI	APRIORI
Calibration 1976–1986	7.7/5.5	5.0/2.9	6.3/3.7
Calibration 1987–1997	6.2/5.2	4.7/2.9	5.9/3.2
Verification 1987–1997	7.5/6.1	5.7/3.4	6.8/4.0
Verification 1976–1986	6.6/4.4	4.8/2.7	5.9/3.5

[a] First value: median over the 320 catchments of the percentage of days with poor snow cover simulations. Second value: difference of the 75% and 25% quantiles, i.e. a measure of scatter. High model performances are associated with small medians and a small scatter.

in the verification periods. The patterns of the VE verification efficiencies (Figure 5) indicate that, for a number of catchments, the water balance is not closed properly for the data. This is the case for all three optimization schemes.

The snow cover model performances Z_S are presented in Table IV and Figures 4 and 5 (right panels). The poorest snow cover simulations are obtained by the SINGLE optimization approach. The median errors over the 320 catchments are around 7% for the calibration periods and the scatter is more than 5%. The MULTI and APRIORI approaches significantly improve the snow simulations, with medians of less than 5% and around 6% respectively. More importantly, the scatter for the MULTI and APRIORI approaches is only 3% and 3–4% respectively. For the verification periods, the MULTI and APRIORI methods, again, outperform the SINGLE method. The median snow errors are around 5% and 6% respectively in the MULTI and APRIORI cases, with scatters of around 3% and 4% respectively, and the SINGLE case gives median errors of around 7% with a scatter of around 6%. It is clear that the use of snow data in calibration improves the snow simulations significantly in both the calibration and verification periods, but the interesting issue is the extent of such an improvement. The results show that the snow data have reduced significantly not only the median of the snow model performance, but also the regional differences represented by the percentile difference.

PARAMETER UNCERTAINTY

We evaluated the uncertainty of the model parameters obtained from the SINGLE, MULTI and APRIORI

Table V. Parameter uncertainty assessed by the coefficient of determination r^2 as a measure of how similar are the model parameters calibrated for two independent 11-year periods (1976–1986 and 1987–1997)

Parameter	Model part	r^2		
		SINGLE	MULTI	APRIORI
SCF	Snow	0·50	0·40	0·56
DDF	Snow	0·26	0·53	0·36
T_R	Snow	0·15	0·22	0·19
T_S	Snow	0·23	0·41	0·25
T_M	Snow	0·27	0·45	0·34
LP/FC	Soil	0·29	0·18	0·25
FC	Soil	0·40	0·31	0·54
β	Soil	0·44	0·26	0·55
K_0	Runoff	0·26	0·25	0·54
K_1	Runoff	0·49	0·44	0·75
K_2	Runoff	0·48	0·31	0·43
C_P	Runoff	0·49	0·41	0·53
C_R	Runoff	−0·01	0·00	0·01
LS_{UZ}	Runoff	0·14	0·18	0·56

Table VI. Parameter uncertainty assessed by the root-mean-square deviation normalized by the parameter range (RMSD) as a measure of scatter of the model parameters calibrated for two independent 11-year periods (1976–1986 and 1987–1997)

Parameter	Model part	RMSD		
		SINGLE	MULTI	APRIORI
SCF	Snow	0·32	0·34	0·22
DDF	Snow	0·27	0·24	0·11
T_R	Snow	0·42	0·39	0·06
T_S	Snow	0·41	0·37	0·09
T_M	Snow	0·30	0·26	0·13
LP/FC	Soil	0·42	0·43	0·10
FC	Soil	0·35	0·35	0·16
β	Soil	0·38	0·45	0·22
K_0	Runoff	0·41	0·44	0·06
K_1	Runoff	0·29	0·30	0·08
K_2	Runoff	0·34	0·40	0·26
C_P	Runoff	0·27	0·28	0·10
C_R	Runoff	0·40	0·39	0·16
LS_{UZ}	Runoff	0·39	0·41	0·10

optimization approaches by a comparison of their values calibrated for the 1976–1986 period with those calibrated for the 1987–1997 period. As measures of the uncertainty we used the coefficient of determination r^2 and the root-mean-square deviation normalized to the parameter range (RMSD) of each parameter for the two periods. Reliably estimated parameters with small uncertainties are those where r^2 is large, RMSD is small and they should cluster around the 1 : 1 line in scatter plots.

The coefficients of determination r^2 are presented in Table V. The most uncertain parameters in all three optimization approaches are the routing parameter C_R and the threshold temperature for liquid precipitation T_R. Parameters with smaller uncertainties are the snow correction factor SCF, the fast storage coefficient K_1 and the percolation rate C_P in the SINGLE approach, the degree-day factor DDF and the threshold temperature for snowmelt T_M in MULTI, and practically all soil and runoff parameters (except the routing parameter C_R and the soil LP/FC ratio) estimated by the APRIORI approach. It is now interesting to examine how the uncertainty changes when moving from the SINGLE to the MULTI approach. Strikingly, the coefficients of determination increase significantly for the degree-day factor DDF (from 0·26 to 0·53), for the lower and upper threshold temperatures of the precipitation state T_R and T_S (from 0·15 to 0·22 and from 0·23 to 0·41 respectively) and for the threshold temperature of melt T_M (from 0·27 to 0·45). These are all parameters related to the snow module of the model. In contrast, the other parameters (related to the soil and runoff components) tend to get slightly more uncertain or do not change much in terms of their uncertainty. Clearly, if snow data are used in calibration it is mainly the snow component of the model that can be expected to improve and, interestingly, only the snow component. The APRIORI case, in contrast, tends to reduce the uncertainty of all model parameters compared with the SINGLE case. However, the snow model parameters are more uncertain than in the MULTI case. This is because, in the APRIORI objective function, less weight is given to snow than in the MULTI case.

As an alternative measure of parameter uncertainty, Table VI shows the root-mean-square deviation of the parameters in the two periods normalized by the parameter range (RMSD). The change in uncertainty when moving from the SINGLE to the MULTI approach is similar to that identified by the coefficients of determination. The RMSD values decrease for the snow parameters and tend to increase or remain similar for the other parameters. However, the RMSD values decrease significantly for all parameters when moving to the APRIORI approach. This indicates that the APRIORI approach constrains the parameters much more drastically than the other methods do.

Examples of the visual appearance of the differences between the calibrated parameters for the two periods are plotted in Figure 6 for the degree-day factor DDF, the maximum soil moisture storage FC and the fast storage coefficient K_1. The left panels relate to the SINGLE case, the middle panels to the MULTI case and the right panels to the APRIORI case. The scatter for the degree-day factor DDF decreases when moving from SINGLE to MULTI, but the scatter of the other parameters does not change much. In contrast, the APRIORI case shows significantly less scatter for all parameters. It should be noted that the reduction in parameter variability in the APRIORI case compared with the SINGLE case does not come at the cost of decreased model performance, as the runoff model performances for the two cases are similar.

DISCUSSION AND CONCLUSIONS

In this study we have assessed the model performance and the parameter uncertainty of two multiple objective

DOI: 10.1002/hyp

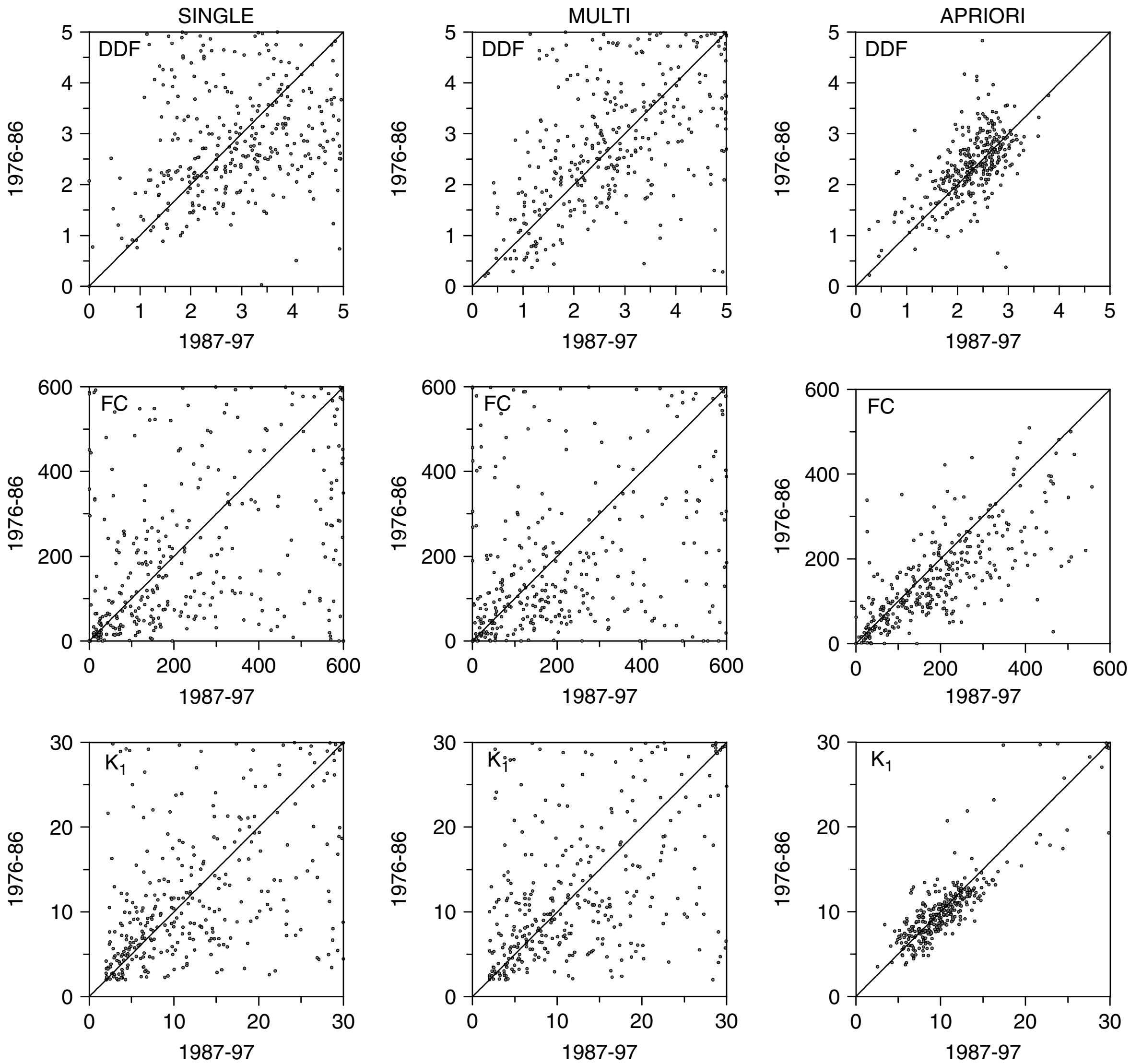

Figure 6. Selected model parameters estimated by the SINGLE (left), the MULTI (centre) and the APRIORI (right) calibration approaches. Horizontal axes show the parameters calibrated for the period 1987–1997; vertical axes show the parameters calibrated for the period 1976–1986. The top panels show the degree-day factor DDF, the middle panels the maximum soil moisture storage FC, and the bottom panels show the fast storage coefficient K_1

calibration approaches and compared their effectiveness with a single objective calibration procedure that involved only daily runoff data in the calibration. The verification approach used for the validation of the three calibration procedures corresponds to the undisturbed-catchment multi-response split-sample and independent-sample strategies proposed by Mroczkowski *et al.* (1997). The first multiple objective approach (termed MULTI) uses snow cover data in addition to runoff. The results indicate that the use of snow cover data significantly improves the model performance in terms of simulating snow cover, but slightly decreases the performance in terms of simulating runoff. This is true of both the calibration and the verification periods. It is interesting that this finding applies to a large number of catchments in diverse hydroclimatic regions of Austria. For a much smaller number of catchments (only two) and different processes (groundwater instead of snow), Seibert (2000) concluded that, when calibrating a model against two objectives, 'the values of the objective functions were about 5 per cent below their values from the single-criterion calibration for both criteria'. Similar results have been reported by Madsen (2003), who showed that the calibration based only on groundwater levels provided poor simulation of catchment runoff. However, a minor relaxation of the performance of the groundwater level simulations led to a significant improvement in runoff simulation. Along similar lines, Seibert and McDonnell (2002) reported that the decrease in model performance was mainly caused by a conflict between the criteria used in parameter optimization for their case of incorporating soft data in multiple objective calibration.

We also used expert judgement to constrain the parameter distribution in addition to using runoff and snow cover data (termed APRIORI case). For the calibration periods, the runoff model efficiencies of this case were similar to single objective calibration on runoff only and for the verification periods the efficiencies were even

DOI: 10.1002/hyp

slightly better. This suggests that it is useful to make *a priori* assumptions on the distribution of parameters even if these are approximate estimates.

In the APRIORI case, the snow cover model efficiency was somewhat lower than in the MULTI case where runoff and snow cover data were used; this is because less weight is given to snow in the objective function. However, the MULTI case yielded significantly better snow simulations than the SINGLE case where no snow data were used. It appears that similar values of the runoff objective function do not necessarily imply a similar hydrological response of the catchment. As pointed out by Gupta *et al*. (2003), 'regardless of the objective function, the response surface contains numerous very similar solutions (in terms of objective function value) at widely differing locations in the parameter space'. The reduction of parameter uncertainty, hence, is clearly of paramount importance.

We used a method for assessing parameter uncertainty that is based on a split sample comparison of model parameters for two periods following the suggestion of Merz and Blöschl (2004). The advantage of the method over, say, Monte Carlo methods is that it accounts for a range of uncertainties, including data uncertainties and non-stationarity, but it may only be used in a meaningful way if a large set of catchments is available for testing, as is the case here. Quantification of parameter uncertainty by the coefficient of determination and the root-mean-square deviation showed that the use of additional snow cover data significantly reduced the parameter uncertainty for the parameters representing the snow cover dynamics, but tended to increase the parameter uncertainty for other model parameters. This is in line with Seibert (2000), who, for the case of using additional groundwater level data, noted that 'the parameter uncertainty decreases for five parameters of the response routine, which is the part of the model representing groundwater dynamics'. Our results show that incorporating an expert guess about the parameter distribution, additionally to runoff and snow cover data, reduced the scatter significantly between parameters optimized for different periods. It is important that this finding applies to an ensemble of 320 catchments. For individual catchments, expert knowledge may or may not reduce parameter uncertainty. An individual example where this has been the case was reported by Seibert and Mc Donnell (2002), who found a reduction of the parameter uncertainty when soft data were added to the multi-criteria model calibration.

The use of expert knowledge about the distribution of model parameters in the optimization procedure appears a promising avenue for further reducing parameter uncertainty. In this study we have used the same parameter distribution for all catchments, but it may be of advantage to use different distributions for different catchments, depending on catchment characteristics for example. Alternatively, neighbouring catchments could be used, as is the case in regional calibration (e.g. Szolgay *et al*., 2003). We believe that constraining the parameter distribution in a regional calibration procedure will improve the effectiveness of multiple objective calibration and provide even more robust regional patterns of model parameters.

ACKNOWLEDGEMENTS

We would like to thank the Marie Curie Fellowship of the European Community programme HUMAN POTENTIAL under contract number HPMF-CT-2002-01872, and the Austrian Academy of Sciences, project HÖ, for financial support. We would also like to thank the Austrian Hydrographic Service (HZB) for providing the hydrographic data. We are grateful also to the three anonymous reviewers for their thoughtful comments and suggestions.

REFERENCES

Beldring S 2002. Multi-criteria validation of a precipitation-runoff model. *Journal of Hydrology* **257**: 189–211.

Bergström S 1976. *Development and application of a conceptual runoff model for Scandinavian catchments*. Department of Water Resources Engineering, Lund Institute of Technology/University of Lund, Bulletin Series A, no. 52.

Bergström S 1992. *The HBV model—its structure and applications*. Report No. 4, Swedish Meteorological and Hydrological Institute.

Beven KJ, Binley AM. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* **6**: 279–298.

Beven KJ, Freer J 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems. *Journal of Hydrology* **249**: 11–29. DOI: 10·1016/S0022-1694(01)00421-8.

Blöschl G 2005. Rainfall-runoff modeling of ungauged catchments. In *Encyclopedia of Hydrological Sciences*, volume 3, Anderson MG, McDonnell JJ (eds). Wiley: Chichester; 2061–2080.

Blöschl G, Kirnbauer R, Gutknecht D 1991. Distributed snowmelt simulations in an Alpine catchment. 1. Model evaluation on the basis of snow cover patterns. *Water Resources Research* **27**(12): 3171–3179.

Cheng CT, Ou CP, Chau KW 2002. Combining a fuzzy optimal model with genetic algorithm to solve multi-objective rainfall-runoff model calibration. *Journal of Hydrology* **268**: 72–86. DOI: 10·1016/S0022-1694(02)00122-1.

Duan Q, Sorooshian S, Gupta VK. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* **28**: 1015–1031. DOI: 10·1029/91WR02985.

Franks SW, Gineste P, Beven KJ, Merot P. 1998. On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into calibration process. *Water Resources Research* **34**: 787–797. DOI: 10·1029/97WR03041.

Grayson RB, Blöschl G 2000. *Spatial Patterns in Catchment Hydrology: Observations and Modelling*. Cambridge University Press: Cambridge, UK.

Grayson RB, Blöschl G, Western A, McMahon T 2002. Advances in the use of observed spatial patterns of catchment hydrological response. *Advances in Water Resources* **25**: 1313–1334.

Gupta HV, Sorooshian S, Hogue TS, Boyle DP. 2003. Advances in automatic calibration of watershed models. In *Calibration of Watershed Models*, Duan Q, Gupta HV, Sorooshian S, Rousseau AN, Turcotte R (eds). *Water Science and Application Series*, Volume 6. AGU: Washington, DC; 9–28.

Holko L, Lepistö A 1997. Modelling the hydrological behaviour of a mountain catchment using TOPMODEL. *Journal of Hydrology* **196**: 361–377. DOI: 10·1016/S0022-1694(96)03237-4.

Klemeš V 1986. Operational testing of hydrological simulation models. *Hydrological Sciences Journal* **31**: 13–24.

Kuczera G, Mroczkowski M 1998. Assessment of hydrological parameter uncertainty and the worth of multiresponse data. *Water Resources Research* **34**: 1481–1489. DOI: 10·1029/98WR00496.

Kuczera G, Parent E. 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis

algorithm. *Journal of Hydrology* **211**: 69–85. DOI: 10·1016/S0022-1694(98)00198-X.
Lindström G 1997. A simple automatic calibration routine for the HBV model. *Nordic Hydrology* **28**(3): 153–168.
Lindström G, Johansson B, Persson M, Gardelin M, Bergström S 1997. Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology* **201**: 272–288. DOI: 10·1016/S0022-1694(97)00041-3.
Madsen H 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Advances in Water Resources* **26**: 205–216.
Merz R 2002. *Understanding and Estimating Flood Probabilities at the Regional Scale*. *Wiener Mitteilungen, Wasser–Abwasser–Gewässer*, Band 181. Technische Universität Wien: Vienna.
Merz R, Blöschl G 2004. Regionalisation of catchment model parameters. *Journal of Hydrology* **287**: 95–123. DOI: 10·1016/j.jhydrol.2003·09·028.
Mészároš I, Miklánek P, Parajka J. 2002. Solar energy income modelling in mountainous areas. In *ERB and NEFRIEND Proj.5 Conference, Interdisciplinary Approaches in Small Catchment Hydrology: Monitoring and Research*, Holko L, Miklanek P (eds). Slovak NC IHP UNESCO/UH SAV: Bratislava, Slovakia; 127–135.
Miettinen K 1999. *Nonlinear Multiobjective Optimization*. *Kluwer's International Series in Operations Research & Management Science*, Volume 12. Kluwer: Boston, MA.
Mroczkowski M, Raper GP, Kuczera G 1997. The quest for more powerful validation of conceptual catchment models. *Water Resources Research* **33**: 2325–2336. DOI: 10·1029/97WR01922.
Parajka J, Merz R, Blöschl G 2003. Estimation of daily potential evapotranspiration for regional water balance modeling in Austria. In *11th International Poster Day and Institute of Hydrology Open Day "Transport of Water, Chemicals and Energy in the Soil–Crop Canopy–Atmosphere System"*. Slovak Academy of Sciences: Bratislava; 299–306.
Pebesma EJ. 2001. *Gstat user's manual*. Department of Physical Geography, Utrecht University, The Netherlands.
Peck A, Gorelick S, de Marsily G, Foster S, Kovalevsky V (eds). 1988. *Consequences of Spatial Variability in Aquifer Properties and Data Limitations for Groundwater Modelling Practice*. *IAHS Publication* No. 175. IAHS Press: Wallingford.
Schrödter H 1985. *Verdunstung—Anwendungsorientierte Meßverfahren und Bestimmungsmethoden*. Springer Verlag: Berlin.
Seibert J 1997. Estimation of parameter uncertainty in the HBV model. *Nordic Hydrology* **28**: 247–262.
Seibert J 2000. Multi-criteria calibration of a conceptual runoff model using genetic algorithm. *Hydrology and Earth System Sciences* **4**: 215–224.
Seibert J, McDonnell J 2002. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. *Water Resources Research* **38**: 1241. DOI: 10·1029/2001WR000978.
Szolgay J, Hlavčová K, Kohnová S, Danihlík R 2003. Estimating regional parameters of a monthly balance model. *Journal of Hydrology and Hydromechanics* **51**: 256–273.
Vrugt JA, Gupta HV, Bastidas LA, Bouten V, Sorooshian S. 2003a. Effective and efficient algorithm for multi-objective optimization of hydrological models. *Water Resources Research* **39**: 1214. DOI: 10·1029/2002WR001746.
Vrugt JA, Gupta HV, Bouten W, Sorooshian S. 2003b. A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* **39**: 1201. DOI: 10·1029/2002WR001642.
Yu P, Yang T. 2000. Fuzzy multi-objective function for rainfall-runoff model calibration. *Journal of Hydrology* **238**: 1–14. DOI: 10·1016/S0022-1694(00)00317-6.

APPENDIX A

Snow model

The snow routine represents snow accumulation and melt by a simple degree-day concept. Mean daily precipitation P in an elevation zone is partitioned into rain P_R and snow P_S based on the mean daily air temperature T_A:

$$\begin{array}{ll} P_R = P & \text{if } T_A \geq T_R \\ P_R = P\dfrac{T_A - T_S}{T_R - T_S} & \text{if } T_S < T_A < T_R \\ P_R = 0 & \text{if } T_A < T_S \\ P_S = P - P_R & \end{array} \tag{A.1}$$

where T_S and T_R are the lower and upper threshold temperatures respectively. Melt starts at air temperatures above a threshold T_M:

$$M = (T_A - T_M)\text{DDF} \qquad \text{if } T_A > T_M \text{ and SWE} > 0 \tag{A.2}$$

where M is the amount of melt water per time step, DDF is the degree-day factor and SWE is the snow water equivalent. The catch deficit of the precipitation gauges during snowfall is corrected by a snow correction factor SCF. Changes in the snow water equivalent from days $i-1$ to i are accounted for by

$$\text{SWE}_i = \text{SWE}_{i-1} + (\text{SCF}P_S - M)\Delta t \tag{A.3}$$

where Δt is the time step of 1 day.

Soil moisture accounting

The soil moisture routine represents runoff generation and changes in the soil moisture state of the catchment:

$$S_{SM,i} = S_{SM,1-i} + P_R + M - E_A \tag{A.4}$$

where S_{SM} is the soil moisture of a top soil layer controlling runoff generation and actual evaporation E_A. The contribution ΔS_{UZ} of rain and snowmelt to runoff is calculated by an explicit scheme as a function of the soil moisture of the top layer S_{SM} using a non-linear relationship with two free parameters, FC and β:

$$\Delta S_{UZ} = \left(\frac{S_{SM}}{\text{FC}}\right)^{\beta}(P_R + M) \tag{A.5}$$

FC is the maximum soil moisture storage. The parameter β controls the characteristics of runoff generation and is a non-linearity parameter. If the top soil layer is saturated, i.e. $S_{SM} = \text{FC}$, then all rainfall and snowmelt contributes to runoff. The actual evaporation E_A is calculated from potential evaporation E_P by a piecewise linear function of the soil moisture of the top layer:

$$\begin{array}{ll} E_A = E_P\dfrac{S_{SM}}{\text{LP}} & \text{if } S_{SM} < \text{LP} \\ E_A = E_P & \text{if } S_{SM} \geq \text{LP} \end{array} \tag{A.6}$$

where LP is a parameter termed the limit for potential evaporation E_P.

Response and transfer functions

The response function represents runoff routing on the hillslopes and consists of two reservoirs, representing two soil zones. The storage states of the upper and lower zones are S_{UZ} and S_{LZ} respectively. ΔS_{UZ} enters the upper zone reservoir and leaves this reservoir through three paths: outflow from the reservoir with a fast storage coefficient of K_1, percolation to the lower zone with a constant percolation rate C_P, and, if a threshold LS_{UZ} of the storage state is exceeded, through an additional outlet with a storage coefficient of K_0. Water leaves the lower zone with a slow storage coefficient of K_2. The outflow from both reservoirs Q_G is then routed by a triangular transfer function, which represents the runoff routing in the streams:

$$\begin{array}{ll} B_Q = B_{MAX} - C_R Q_G & \text{if } (B_{MAX} - C_R Q_G) \geq 1 \\ B_Q = 1 & \text{otherwise} \end{array} \tag{A.7}$$

where B_Q is the base of the transfer (triangular) function, B_{MAX} is the maximum base at low flows and C_R is a free scaling parameter. The B_{MAX} model parameter was set to a constant value of $B_{MAX} = 10$ days on the basis of sensitivity analyses, whereas the remaining 14 parameters were found by the calibration.

DOI: 10.1002/hyp