Regional calibration of catchment models: Potential for ungauged catchments

Juraj Parajka,[1,2] Günter Blöschl,[1] and Ralf Merz[1]

[1] We present a novel iterative regional calibration (IRC) method in which the model parameters of a number of catchments are calibrated simultaneously. The method exploits the spatial correlations of the parameters to condition their a priori distribution for each catchment. We use a sample of 320 catchments in Austria over a period of 22 years to test the method. The results indicate that the IRC method allows us to reliably calibrate the conceptual hydrologic model used here. The novel method reduces the uncertainty of most parameters as compared to local calibration. This is demonstrated by more consistent model parameters in two independent calibration periods and by an analysis of their spatial variability. Jackknife cross validation indicates that the IRC method tends to improve runoff simulation performance for ungauged catchments as compared to traditional regionalization, although the gain is small in absolute terms.

Citation: Parajka, J., G. Blöschl, and R. Merz (2007), Regional calibration of catchment models: Potential for ungauged catchments, *Water Resour. Res.*, *43*, W06406, doi:10.1029/2006WR005271.

1. Introduction

[2] There are both operational and academic drivers for pursuing rainfall-runoff modeling of ungauged catchments. The former include design applications (of spillways, culverts, and embankments), forecasting applications (flood warning and hydropower operation) and catchment management applications (water allocation, climate impact studies), the latter are geared toward understanding the catchment functioning and how the individual processes combine to produce catchment response. The main challenge with rainfall-runoff modeling in ungauged catchments is the lack of local runoff data that could be used for calibrating model parameters. Parameter calibration is important because of a number of reasons. Calibration can account for the effects of the hydrological setting in a particular catchment; calibration can adjust for biases in the inputs, for example as a result of orographic effects and instrument biases; and calibration can significantly reduce the biases of model predictions by accounting for the lumped effect of the catchment response characteristics. While calibration on runoff data has served hydrology well in the past, this is not an option in ungauged catchments. Alternatives are needed.

[3] For the case of conceptual catchment models the most widely used alternative to calibration is to transfer the parameters from gauged catchments in the region that can be considered hydrologically similar to the target catchment in some way. There are a number of methods of transferring or regionalizing the parameters [*Blöschl*, 2005], but the general pattern is a two step procedure of first calibrating the catchment model to the runoff records of the gauged catchments and then applying a regionalization technique to obtain the model parameters in the ungauged catchment of interest. Typically, the regionalization techniques consist of a regression between model parameters and catchment attributes (such as soil type) that are available in both gauged and ungauged catchments. While widely used for practical purposes, two main difficulties exist: (1) The regionalization relationships are usually poorly defined, and (2) the runoff simulations of the ungauged catchments do not usually perform nearly as well as in the gauged catchments. These difficulties have, *inter alia*, been singled out as the problem of ungauged basins in a recent international initiative [*Sivapalan et al.*, 2003].

[4] *Blöschl* [2005] attributed the poor performance of regionalization relationships to three main reasons.

[5] 1. One explanation is that the measurable catchment attributes that are used in the regionalization relationships may not be very relevant for catchment response. Soil type, e.g., is an important catchment attribute but its generally low predictive power is known from its use in pedotransfer functions [*Wösten et al.*, 2001].

[6] 2. The second explanation is that the structure of the model relating catchment attributes and model parameters may not be suitable. Indeed, the usual choice of a linear regression model is one of convenience rather than one based on known relationships. Nonlinear methods exist, but they tend to increase the number of parameters that need to be estimated.

[7] 3. The third explanation is that there may be significant uncertainty in the calibrated parameter values which may cloud the underlying relationship between calibrated model parameters and catchment attributes [e.g., *Gottschalk*, 2002]. There are methods of accounting for parameter uncertainty in the regionalization of model parameters (see, e.g., *Bates and Campbell* [2001] for the case of an event-scale rainfall-runoff model illustrated on 39 catchments in south-

[1]Institute for Hydraulic and Water Resources Engineering, Vienna University of Technology, Vienna, Austria.
[2]On leave from Institute of Hydrology, Slovak Academy of Sciences, Bratislava, Slovakia.

western Australia) but it is clearly desirable to better constrain model parameters in the calibration process. There are two possibilities. The first is to use additional data on state variables in the calibration procedure. These can be geochemical data [*Mroczkowski et al.*, 1997], groundwater data [*Madsen*, 2003], soil moisture data [*Western and Grayson*, 2000] or snow data [*Parajka et al.*, 2007]. The second possibility is regional calibration in which the model parameters of a number of catchments are calibrated simultaneously. This is the topic of this paper.

[8] The key idea of regional calibration is that more robust model parameters can be obtained if a number of gauged catchments in a region are used simultaneously in parameter estimation. The more robust parameters are then hoped to translate into a reduction in runoff simulation uncertainty when transposing them to ungauged catchments.

[9] There have only been a few attempts at regional calibration in catchment hydrology but they have sparked a heated discussion. Indeed, regional calibration is a controversial issue. There are two main aspects to the debate. The first issue is whether regional calibration per se is a desirable goal. The clash in modeling philosophies is that the raison d'être of parameter calibration is adjustment to the local situation while regional calibration strives for the exact opposite: adjustment to the regional trends. This is a classical no-win situation. In order to get better local estimates in ungauged catchments one compromises for regional trends and hence forfeits some of the local peculiarities in gauged catchments. The second issue is whether regional calibration actually improves hydrological simulations in ungauged catchments over local calibration and classical regionalization. The only test we are aware of is by *Fernandez et al.* [2000]. Their results indicate that the regional calibration indeed improved the relationships between model parameters and catchment attributes but did not improve the runoff simulations at ungauged sites. Given that the amount of information used in regional calibration is significantly larger than that of local calibration for any one catchment one could, however, speculate that there is potential in the regional calibration approach. In different hydrological settings, using different methods and/or more comprehensive data sets regional calibration may well prove more useful than what the results of *Fernandez et al.* [2000] suggest. In this paper we therefore adopt a pragmatic stance by (1) proposing a novel method of regional calibration and (2) testing its potential for improving runoff simulations in ungauged catchments over alternative methods, without engaging in the philosophical debate.

[10] The are two types of methods in regional calibration of catchment hydrological models, methods that concurrently solve the parameter estimation and regionalization problem in a single step; and iterative methods. The first type includes the method of *Fernandez et al.* [2000], who calibrated a monthly water balance model concurrently with regressions between model parameters and catchment attributes in 33 catchments by optimizing a compound objective function involving runoff simulation efficiency and goodness of fit of the regressions. Other examples are those of *Hlavčová et al.* [2000] and *Szolgay et al.* [2003], who first found groups of catchments by cluster analysis of catchment attributes and then calibrated a monthly water balance model to 14 catchments assuming uniform model parame-

ters in each group. In a somewhat similar study, *Drogue et al.* [2002] assumed two parameters of an hourly conceptual catchment model to be uniform in a region of 16 catchments and stratified the two other parameters by lithological groups. *Hundecha and Bárdossy* [2004] assumed preset relationships between model parameters and catchment attributes, and calibrated the coefficients of the relationships rather than the parameters themselves for 30 catchments. The difficulty with most of these methods is that calibration of the relationships instead of the parameters, significantly increases the number of coefficients to be calibrated. For example, in the study of *Fernandez et al.* [2000] the four model parameters for each catchment doubled, as linear regressions involve two coefficients (slope and intercept). *Hundecha and Bárdossy* [2004] used attribute-parameter relationships with four parameters, thereby quadrupling the number of coefficients to be estimated. The benefit, however, is that the number of station years of runoff data also increases, so more information is available to estimate a larger number of coefficients. One could argue that the total number of coefficients to be estimated in the regional calibration should always be smaller than the number of model parameters times the number of gauged catchments (local calibration) as it is this decrease in dimensionality on which regional calibration builds.

[11] Iterative methods have been proposed to reduce the dimensionality of the estimation system as compared to single-step methods. *Lamb et al.* [2000] and *Wagener and Wheater* [2006] sequentially identified model parameters during the calibration. In a first iteration they examined what is the model parameter that can be estimated with the least uncertainty from local calibration in a region. They then regressed the calibration values of this model parameter against catchment attributes, estimated the parameter from the regression for each catchment and froze this parameter value for the remainder of the analysis. In a second step they recalibrated the model to all catchments (without changing the values of the previously identified parameter) and identified the next parameter that could be estimated with least uncertainty. They then proceeded to obtain regressions with catchment attributes for all parameters. While the procedure did involve a number of problems, particularly the inability to define the calibration problem without introducing a bias in the parameter estimates, it is a strategy that may be worth pursuing.

[12] In this paper we propose an iterative regional calibration (IRC) method that builds on the spatial correlations of model parameters and efficiently converges for a large number of catchments. Specifically, the aim of the paper is to test the iterative regional calibration method in terms of the robustness of the estimated parameters, and the potential it has for hydrological simulations in ungauged catchments. We use a sample of 320 catchments over a period of 22 years which will likely allow us to draw more generic inferences than has been possible in previous studies.

2. Method

2.1. Proposed Iterative Regional Calibration Scheme

[13] The basic idea of the method is to combine local and regional information in one objective function to be optimized. The local information involves runoff data and other
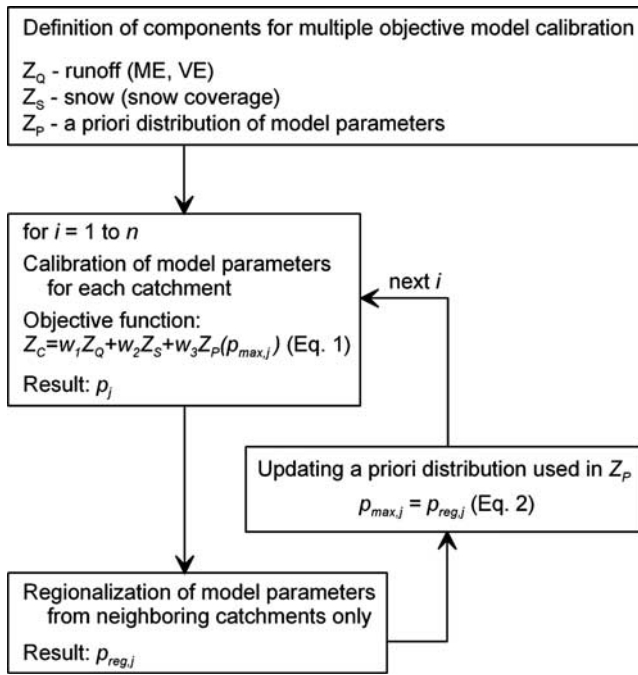
Definition of components for multiple objective model calibration

Z_Q - runoff (ME, VE)
Z_S - snow (snow coverage)
Z_P - a priori distribution of model parameters

↓

for $i = 1$ to n
Calibration of model parameters
for each catchment
Objective function:
$Z_C = w_1 Z_Q + w_2 Z_S + w_3 Z_P(p_{max,j})$ (Eq. 1)
Result: p_j

← next i

Updating a priori distribution used in Z_P

$p_{max,j} = p_{reg,j}$ (Eq. 2)

Regionalization of model parameters
from neighboring catchments only
Result: $p_{reg,j}$

Figure 1. Flowchart of the iterative regional calibration procedure.

hydrological response data such as snow cover. The regional information involves an a priori distribution of each model parameter that is conditioned on parameter values calibrated in gauged catchments within the study region. While there are potentially many different ways of solving this estimation problem we have chosen an iterative approach that is numerically efficient for a large number of catchments.

[14] The iterative scheme is depicted in Figure 1. As a starting point, a compound objective function is defined that linearly combines local and regional information. In this paper we use two types of local information, runoff data and snow cover data. The compound objective function, Z_C, hence is

$$Z_C = w_1 \cdot Z_Q + w_2 \cdot Z_S + w_3 \cdot Z_P \qquad (1)$$

where Z_Q is the runoff related component, Z_S is the snow cover component, Z_P is the regional component representing the conditioned a priori distributions of the parameters, and the w are weights. This objective function is to be minimized to find optimum parameters.

[15] Let p_j be the parameter vector of the hydrological model, m the number of catchments in the study region and n the number of iterations. In a first iterative step, the model parameters are calibrated for each catchment separately. The a priori distributions of the parameters are specified based on a first guess. Different distributions are used for different parameters but, for a given parameter j, the same distribution is used for all m catchments in the region. The result of this calibration is a calibrated parameter vector p_j for each catchment. The calibrated parameters can then be transferred in space by any regionalization procedure. The

regionalization approach may, e.g., utilize the relationship between model parameters and selected catchment attributes, or may be based on the spatial correlation of the model parameters. For a given catchment, the model parameters j are regionalized independently of each other without using the locally calibrated parameters. The result of this regionalization is a regionalized parameter vector $p_{j,reg}$ for each catchment. The a priori distribution for each parameter j of each catchment is now conditioned on these regional parameter values. While there are, again, numerous ways of conditioning distributions to additional information we have chosen a very simple method. We assumed that the regional value is identical with the most likely local value of the parameter but does not affect the upper and lower bounds of the distributions. We hence shifted the mode or maximum $p_{max,j}$ of the distribution to the regional value of the parameter and left the upper and lower bounds unchanged, i.e.,

$$p_{max,j} = p_{reg,j} \qquad (2)$$

[16] The updating of the distribution changes Z_P as Z_P is a function of the distribution (and hence $p_{max,j}$).

[17] The procedure moves to the second iterative step using the updated Z_P instead of the original one. The iteration is performed n times until the parameters converge. The parameters are assumed to have approximately converged if the innovations of the parameters from an iterative step are small.

[18] While, as mentioned above, alternative regional calibration procedures would also be feasible, we consider the iterative approach to have two main strengths. The dimensionality of the optimization system is kept low as, in each iteration, each catchment is calibrated independently. The total number of unknowns in any one step is hence equal to the number of calibration parameters. Second, the iterations allow us to monitor the change of parameters as a function of the iteration step. This provides insight into the robustness of the parameters. Parameters that can be estimated robustly would be expected to not change much between the iteration steps while poorly identifiable parameters may fluctuate strongly between the iterations.

[19] The objective function component Z_P is here defined as a penalty function based on the a priori distribution f_j of the parameters:

$$Z_P = \sum_{j=1}^{k} \frac{f_{max,j} - f_j\left(\frac{p_j - p_{l,j}}{p_{u,j} - p_{l,j}}\right)}{f_{max,j}} \qquad (3)$$

$$f_{max,j} = f_j\left(\frac{p_{max,j} - p_{l,j}}{p_{u,j} - p_{l,j}}\right) \qquad (4)$$

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 corresponding to the mode of the a priori distribution, and k is the number of parameters to be calibrated. The probability density function
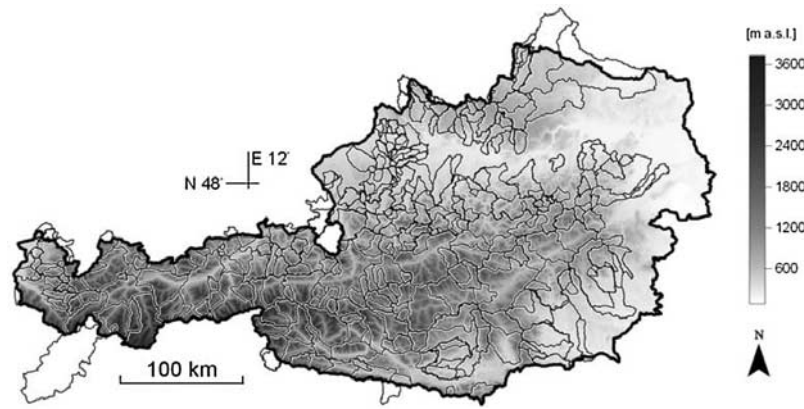
Figure 2. Location of 320 catchments and topography of Austria.

of distribution f was assumed to conform to a Beta distribution with parameters u and v.

$$f(x|u,v) = \frac{1}{B(u,v)} x^{u-1}(1-x)^{v-1} \quad \text{for } 0 < x < 1, u > 0, v > 0 \tag{5}$$

$$\text{with} \quad B(u,v) = \int_0^1 x^{u-1}(1-x)^{v-1}dx = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)} \tag{6}$$

[20] The updating procedure allows the a priori distributions to vary between catchments. The main motivation for choosing this procedure (equation (2)) was its consistency with the a priori component of the objective function, Z_p. If the locally calibrated parameters are identical with the regionally estimated parameters, Z_p is zero which amounts to no penalty from the regional calibration component as the local parameters are consistent with their regional trend. Shifting the mode p_{max} of the distribution to the regional value of the parameter, leaving the upper and lower bounds unchanged, means that the shape of the a priori distribution changes. In the updating step this is accomplished by modifying the parameters u and v in equations (5) and (6). The parameters u and v can be expressed as a function of the mode p_{max} but there is no unique relationship as there are two unknowns with a single equation. An additional assumption is hence needed. We assumed here that the sum of the parameters remains constant, i.e.,

$$u + v = \xi \tag{7}$$

where ξ is a constant to be defined in the modeling set up. This assumption has yielded plausible results judged by comparing the distributions found in the IRC procedure for individual catchments with the frequency distributions of the parameters obtained from all catchments in the region.

2.2. Data

[21] 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 m asl to 3797 m asl. Mean annual precipitation is less than 400 mm/year in the east and almost 3000 mm/year 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 data set used in this study includes measurements of daily precipitation and snow depths at 1091 stations and daily air temperature at 212 climatic stations. Daily runoff data from 320 gauged catchments were used with areas ranging from 10 km^2 to 9770 km^2 and a median of 196 km^2. The boundaries of the gauged catchments are shown in Figure 2.

[22] The inputs to the hydrologic catchment 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 depth, and the least squares trend prediction method was used for air temperature [*Pebesma*, 2001]. The spatial distribution of potential evapotranspiration was estimated by a modified Blaney-Criddle method [*Parajka et al.*, 2003] using daily air temperature and potential sunshine duration calculated by the Solei-32 model [*Mészároš et al.*, 2002; *Mészároš and Miklánek*, 2006] that incorporates shading by surrounding terrain. In a second step, a digital elevation model of 1 × 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.

2.3. Hydrological Model

[23] The model used in this paper to demonstrate the IRC procedure is a semidistributed conceptual rainfall-runoff model, following the structure of the HBV model [*Bergström*, 1976]. The model equations are given in the appendix of *Merz and Blöschl* [2004], although their model version was spatially lumped while a semidistributed version is used here. Each catchment is subdivided into elevation zones of 200 m vertical range. The model runs on a daily time step and consists of a snow routine, a soil moisture routine and a flow routing routine. The snow routine represents snow accumulation and melt by a simple degree-day concept, involving the degree-day factor DDF and melt temperature T_M. Catch deficit of the precipitation gauges during snowfall is corrected by a snow correction factor, SCF. A threshold temperature interval $T_R - T_S$ is

Table 1. Model Parameters to Be Calibrated by the IRC Method[a]

Model Parameter j	Model Component	u	v	p_l	p_u	p_{max}
SCF	snow	1.2	4.0	1.0	1.5	1.03
DDF, mm/°C d	snow	2.0	4.0	0.0	5.0	1.25
LP/FC	soil	4.0	1.2	0.0	1.0	0.94
FC, mm	soil	1.1	1.5	0.0	600	100
B	soil	1.1	1.5	0.0	20	3.4
K_0, days	runoff	2.0	4.0	0.0	2.0	0.5
K_1, days	runoff	2.0	4.0	2.0	30	9.0
K_2, days	runoff	1.05	1.05	30	250	105
C_P, mm/d	runoff	2.0	4.0	0.0	8.0	2.0
C_R, d^2/mm	runoff	1.05	1.05	0.0	50	25
LS_{UZ}, mm	runoff	3.0	3.0	1.0	100	50

[a]Here u and v are the initial parameters of the a priori distribution (equations (5) and (6)), p_l and p_u are the lower and upper bounds of the parameter space used in all iterations, and p_{max} is the initial parameter value at which the a priori distribution is at a maximum (equation (4)). Parameters u, v, and p_{max} are allowed to change between iterations.

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 nonlinear function relating runoff generation to the soil moisture state, termed the nonlinearity parameter B. 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 the base of the transfer function is related to the outflow by a free calibration parameter C_R.

[24] The model was run for all 320 gauged catchments in Austria. Daily inputs (precipitation, air temperature and potential evapotranspiration) were allowed to vary with elevation within a catchment, and the soil moisture accounting and snow accounting was performed independently in each elevation zone. However, the same model parameters were assumed to apply to all elevation zones of a catchment. In order to reduce the number of calibrated model parameters we performed a sensitivity analysis based on the combination of Latin hypercube and one-factor-at-a-time sampling [van Griensven et al., 2006]. Ranking the sensitivity of each model parameter in all 320 Austrian catchments revealed that many of the model parameters are sensitive in some catchments but insensitive in others. Three parameters that were among those that generally showed the least sensitivity were preset ($T_R = 2°C$, $T_S = 0°C$, $T_M = 0°C$) and 11 parameters (Table 1) were estimated by regional calibration. Out of these, the Latin hypercube sampling indicated that the degree-day factor DDF and the snow correction factor SCF were the most sensitive and the slow

storage coefficient K_2 and C_R the most insensitive model parameters in the majority of the 320 catchments.

2.4. Application of the Iterative Regional Calibration Scheme

[25] For the demonstration of the regional calibration algorithm the local components of the objective function in equation (1) had to be first specified. The runoff objective function Z_Q follows the relationship proposed by Lindström [1997] that combines the Nash-Sutcliffe coefficient (M_E) and the relative volume error (V_E):

$$Z_Q = (1 - M_E) + w_4 \cdot |V_E|, \quad (8)$$

where

$$M_E = 1 - \frac{\sum_{i=1}^{n} (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^{n} (Q_{obs,i} - \overline{Q_{obs}})^2}, \quad (9)$$

$$V_E = \frac{\sum_{i=1}^{n} Q_{sim,i} - \sum_{i=1}^{n} Q_{obs,i}}{\sum_{i=1}^{n} Q_{obs,i}}. \quad (10)$$

[26] $Q_{sim,i}$ is the simulated runoff on day i, $Q_{obs,i}$ is the observed runoff, $\overline{Q_{obs}}$ is the average of the observed runoff over the calibration (or verification) period of n days. The weight w_4 was chosen based on test simulations. These simulations indicated that optimization against ME efficiency alone resulted in simulations that did not close the long-term water balance in a few catchments. We therefore tested different criteria combinations and weighting schemes and found that the weight of $w_4 = 0.1$ gave the most plausible results with respect to both daily variations of runoff and the long-term water balance.

[27] The snow objective function Z_S uses observed and simulated snow coverage. Observed snow coverage was estimated from daily grid maps constructed from the observed snow depth data. If the zone average of snow depth in a catchment was greater than 0.5 mm the zone was considered as snow covered, otherwise as snow free. Simulated snow coverage was derived from the snow water equivalent simulated by the model where an elevation zone was considered as snow covered if the water equivalent was greater than 0.1 mm, otherwise it was considered as snow free. Snow simulations on a particular day were considered to be poor if the absolute difference between simulated and observed snow coverage was greater than 50% of the catchment area. 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:

$$Z_S = \frac{n_{ps}}{n} \quad (11)$$

[28] The a priori distributions of the parameters used in the first iteration of Z_P were set by expert judgment based
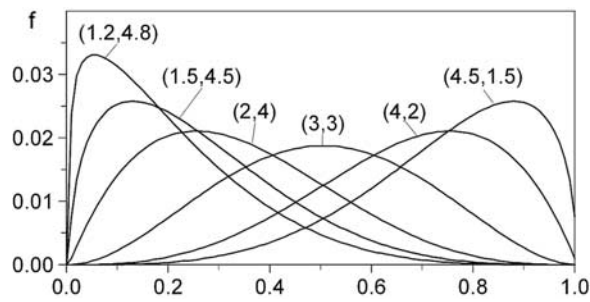
Figure 3. Examples of the shape of the scaled a priori distribution of the model parameters with $\xi = 6$. Numbers in brackets are the u and v parameters of equation (5).

on prior experience with the model, an assessment of hydrological processes in the region and literature values. The coefficients of the distributions used in the first iteration (u and v), and in all iterations (p_l and p_u) are given in Table 1.

[29] The regionalization method chosen in this paper for the spatial transfer of model parameters exploits their spatial correlations. Each model parameter of a given catchment is regionalized by ordinary kriging from the neighboring catchments. This amounts to estimating each parameter in the catchment as a weighted mean of the parameters in catchments within a specified distance of 90 km. The weights are found from the spatial correlations, i.e., the variogram of each parameter. The choice of the kriging regionalization method was motivated by three aspects. First, the analyses of *Merz and Blöschl* [2004] and *Parajka et al.* [2005] indicated that the method is as good or better than alternative regionalization methods in terms of the predictive performance of catchment models in ungauged basins. It should be noted, however, that these studies took advantage of the relatively dense stream gauging network which may not be available in all regions around the world. Second, kriging is a best linear unbiased estimator and hence a logical choice for a regional estimation method. Third, kriging regionalization does not involve additional calibration parameters. In contrast, regionalization by linear regression [*Fernandez et al.*, 2000] in a regional calibration context doubles the number of coefficients to be estimated. A small number of coefficients and hence low dimensionality of the optimization system is considered an advantage.

[30] The selection of weights in equation (1) 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_S) and a priori penalty (Z_P) contributed to the final compound objective function Z_C by one third each. This gave $w_1 = 0.6$, $w_2 = 0.1$ and $w_3 = 0.3$. The objective function Z_C was optimized by the SCE-UA method [*Duan et al.*, 1992]. In each optimization step, a total of 11 variables had to be estimated (i.e., the number of calibration parameters) for each catchment.

[31] In the kriging regionalization step, a variogram needs to be specified for each parameter. We examined the spatial variability of the model parameters calibrated in the first iteration. The distance between two catchments was measured by the geographical distance of the catchment centroids. The experimental variograms found by this analysis suggested that the shape of the variograms is similar for most parameters. The maximum distance over which corre-

lations exist is about 60 km. We therefore assumed a spherical shape of the variogram for all model parameters with a range of 60 km. We calculated the variance (termed the sill in geostatistics) for each parameter and each iteration step separately by setting it to the spatial variance of the calibrated parameters of the same iteration step.

[32] In the current setup of the updating procedure the sum of the u and v parameters, ξ, needs to be specified. On the basis of previous modeling experience in the region and graphical evaluation of various combinations we set $\xi = 6$. This assumption implies that only a subset of the possible shapes of the a priori distribution of the parameters is allowed. Figure 3 shows examples of the shapes of the distribution. The iteration is performed n times until the parameters have converged. Comparisons of the parameters of subsequent iterations in this study suggested that the parameters do not change much after 10 iterations. n was therefore set to $n = 10$ in this paper.

[33] The proposed IRC procedure was evaluated in three steps. First, we examined the calibration and verification efficiencies with respect to measured runoff and snow cover. 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 we performed split sample tests 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. Second, we assessed the parameter uncertainty by comparing the model parameters calibrated for the 1976–1986 period with those calibrated for the 1987–1997 period. A comparison of parameter values obtained from two different periods shows the total amount of uncertainty of these parameters including uncertainty due to input data and model structure. Third, we examined whether the regional calibration actually improves hydrological simulations in ungauged catchments over local calibration and classical regionalization. We emulate the ungauged case for each catchment by jackknife cross validation where we only use regional information for estimating the model parameters (i.e., the regionalization result of the last iteration), simulate runoff for each catchment using these parameters and compare the simulations with the local runoff data as well as the snow data. Strictly speaking, the analysis does not fully represent an ungauged setting as we do use local information of runoff and snow cover in the previous iterations to condition the parameter distributions of the neighboring catchments. However, given the dimensionality of the system the effect is very small. In other words there is a very weak dependence of the cross-validation performance on the local runoff data, so it provides a very good approximation to the ungauged catchment situation. The cross-validation method hence allows us to examine the potential of regional calibration for ungauged sites for all 320 catchments in the study region. The number of catchments involved in the spatial transfer of the model parameters differs regionally from 5 to 66 catchments, with an average of 31 catchments within the 50 km radius.

[34] The results of the first iteration prior to regionalization are the model parameters p_j found by local (at-site) calibration and the associated local calibration and verifica-
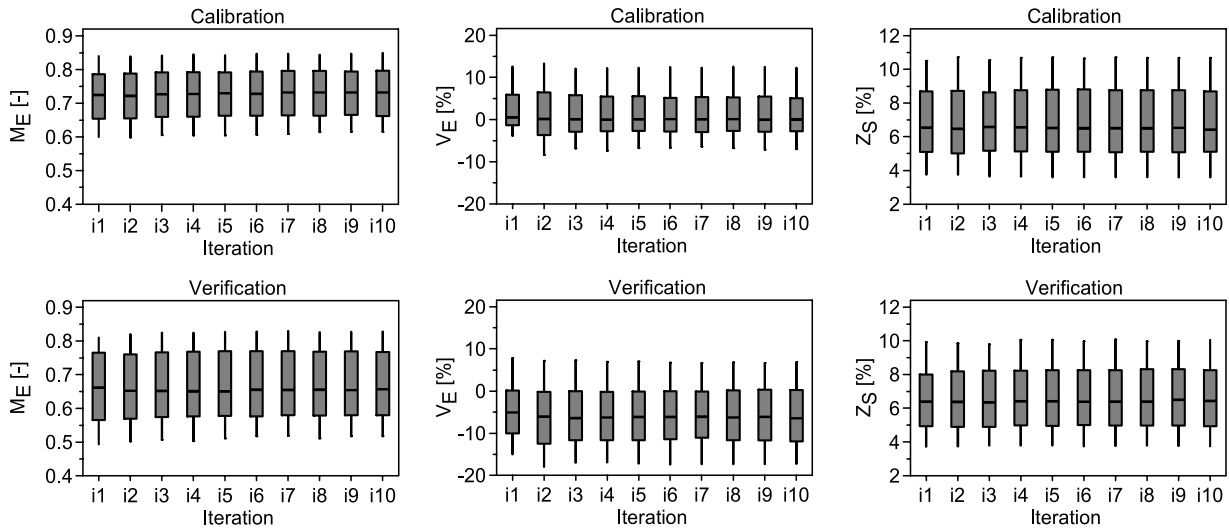
Figure 4. Model efficiencies during the iterative regional calibration for the (top) calibration 1987–1997 and (bottom) verification 1976–1986 periods. M_E is the Nash-Sutcliffe runoff model efficiency, V_E is the runoff volume error, and Z_S is the snow model error expressed as the percentage of days with poor snow cover simulation. The horizontal lines in the boxes represent the median, the box sizes represent the 25th and 75th percentiles, and the vertical lines represent the 10th and 90th percentiles of the model efficiencies over the 320 catchments. The first iteration is equivalent to local calibration without any regionalization.

tion efficiencies. The results of the first and the other iterations prior to regionalization (p_j and associated efficiencies) are shown in Figures 4–11. In addition we examined the potential for ungauged sites in Figure 12 by comparing the results of the first iteration after regionalization ($p_{reg,j}$, corresponding to traditional regionalization) with the results of the last iteration after regionalization ($p_{reg,j}$, corresponding to the ungauged case using the IRC method).

3. Results

3.1. Calibration and Verification Efficiencies

[35] The runoff model performance is presented in terms of its efficiency to simulate runoff (M_E, equation (9), and

V_E, equation (10)) and snow cover (Z_S, equation (11)) in Figure 4 and Tables 2 and 3. Percentiles of the efficiencies over all 320 catchments in each of the ten iterations of the IRC procedure are given. For a favorable model performance, the M_E runoff efficiencies should be large, the V_E volume errors should be close to 0 with a small scatter and the Z_S snow cover errors should be small. Figure 4 presents the results from the calibration period 1987–1997 and the verification period 1976–1986 which is based on the 1987–1997 parameters. Tables 2 and 3 include the cases where the two periods have been interchanged.

[36] Iteration 1 is equivalent to local calibration without regionalization (i.e., using p_j). For the 1976–1986 calibra-
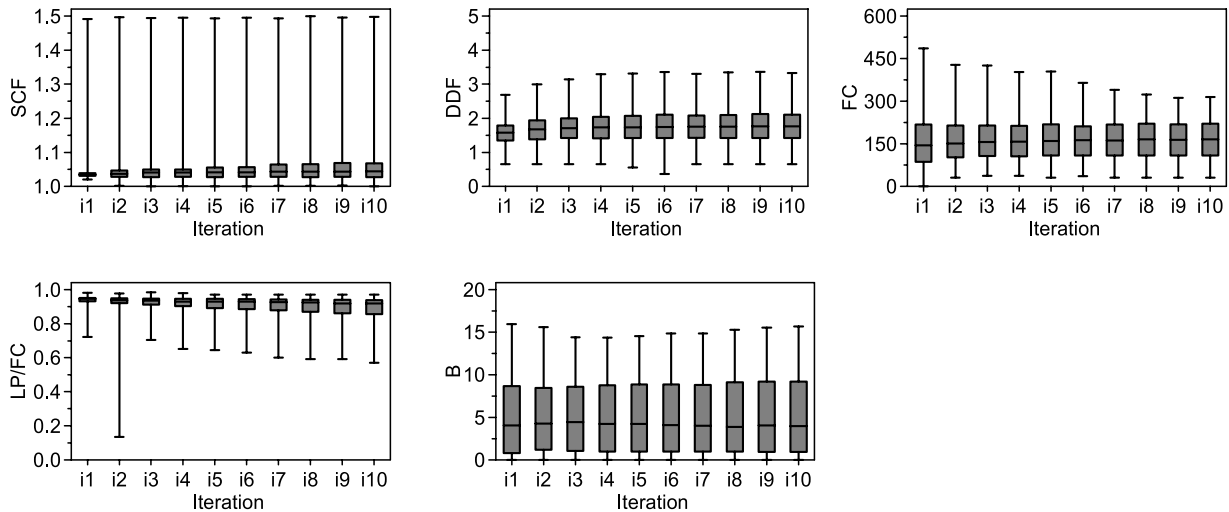

Figure 5. Snow and soil model parameters during the iterative regional calibration. Box-whisker plots show the minimum, maximum, median, and 25th and 75th percentiles of model parameters over the 320 catchments calibrated for the period 1987–1997. The first iteration is equivalent to local calibration without any regionalization.
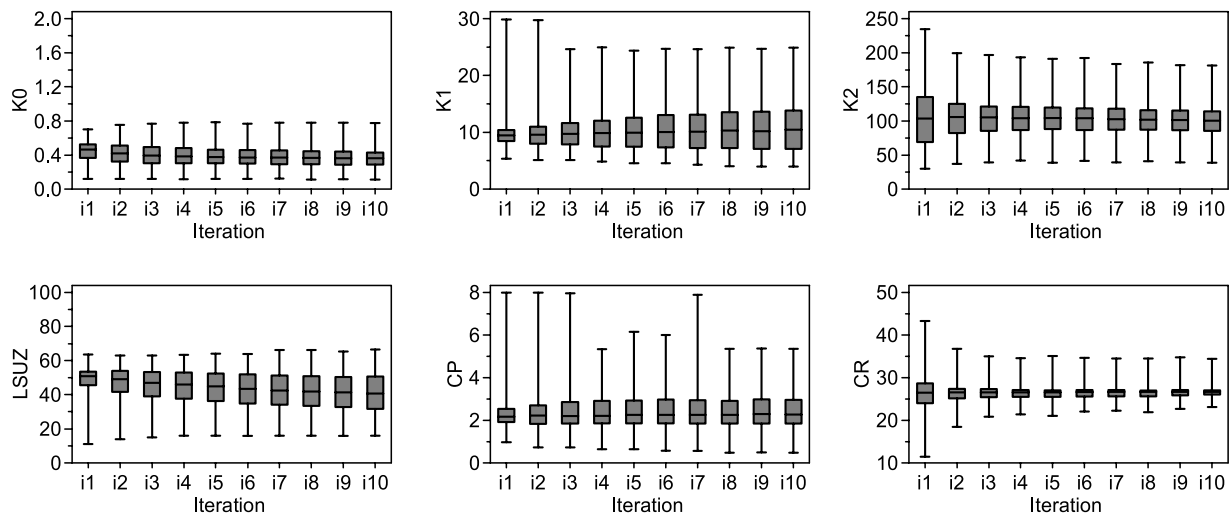
Figure 6. As in Figure 5 but for the runoff routing model parameters.

tion period the median and scatter of the M_E efficiencies are 0.71 and 0.19, respectively. The corresponding values for the 1987–1997 calibration period are 0.72 and 0.13. The median of V_E is close to zero and the scatter is small which indicates that the runoff volumes are essentially unbiased. The median of the snow performance measure Z_S is 6.8% and 6.5% in the two calibration periods. It is now of interest to see how these efficiency measures change during the iteration steps of the iterative regional calibration procedure. Figure 4 and Table 2 indicate that there is in fact very little change. Both the median and the scatter remain close to the values obtained from local calibration (iteration 1).

[37] There is a slight decrease in model efficiency when moving from the calibration to the verification period (Figure 4 and Table 3). Iteration 1 is equivalent to model verification based on locally calibrated parameters p_j. The medians of M_E of the two periods are 0.66 and 0.69 which is lower than their calibration counterparts (0.71 and 0.72,

respectively) but the scatter of the efficiencies practically does not change. The runoff volume errors V_E in the verification periods are larger than those in the calibration period (medians of around 6% and −6% for the two verification periods). These differences are probably related to generally drier climatic conditions in the 1976–1986 period which cannot be fully accounted for by the calibration parameters. The snow cover errors in the verification period are very similar to those of the calibration period. Again, a comparison of the first iteration with the subsequent ones shows that the verification efficiencies change very little during the iteration steps.

3.2. Variability and Uncertainty of Model Parameters

[38] In this section we analyze the variability of model parameters in each iteration step of the IRC procedure. Figure 5 shows the minimum, maximum and percentiles over the 320 catchments of the snow and soil model parameters, Figure 6 shows analogous results for the runoff
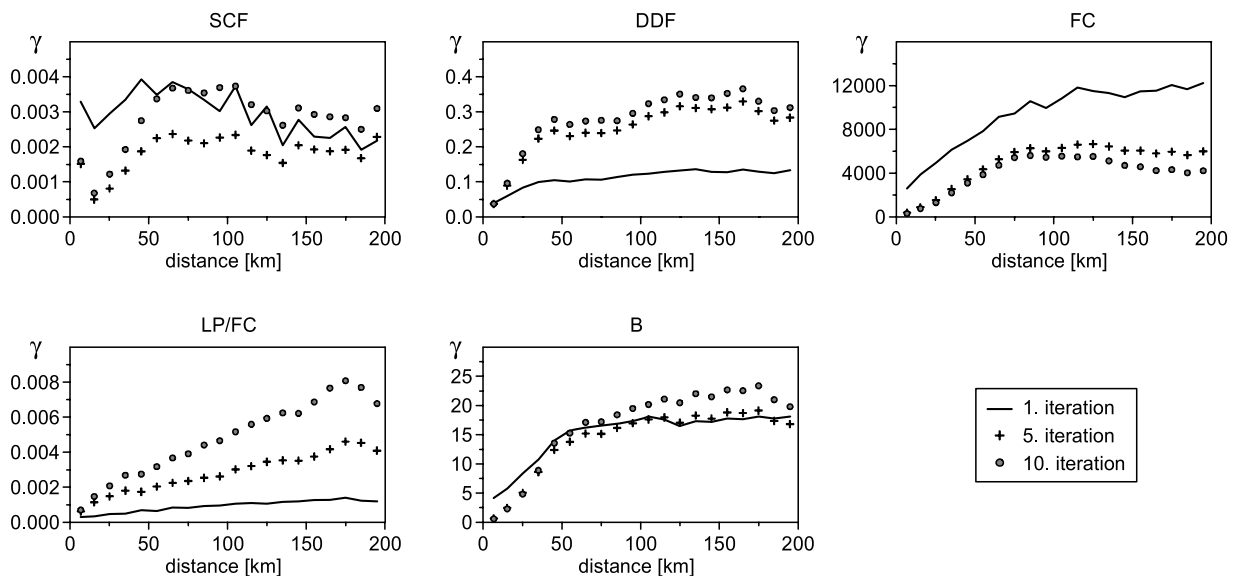

Figure 7. Variograms of the snow and soil model parameters regionally calibrated for the period 1987–1997. The first iteration is equivalent to local calibration without any regionalization.
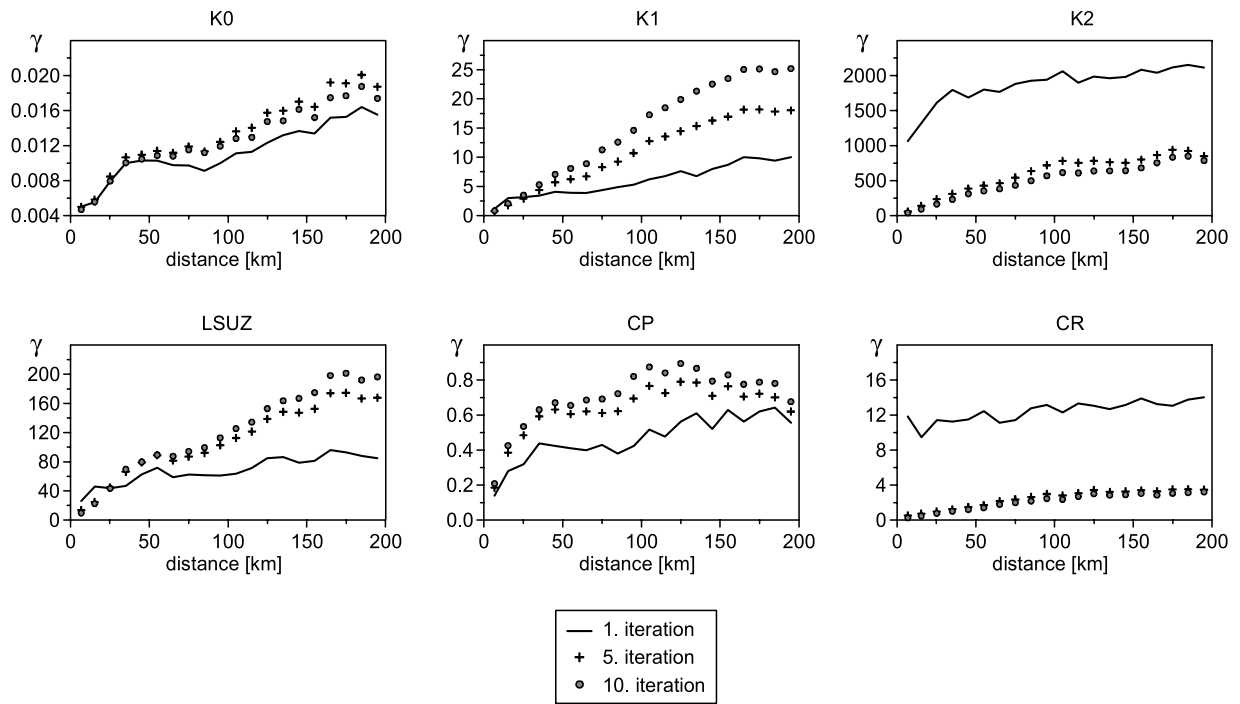
Figure 8.   As in Figure 7 but for the runoff routing model parameters.
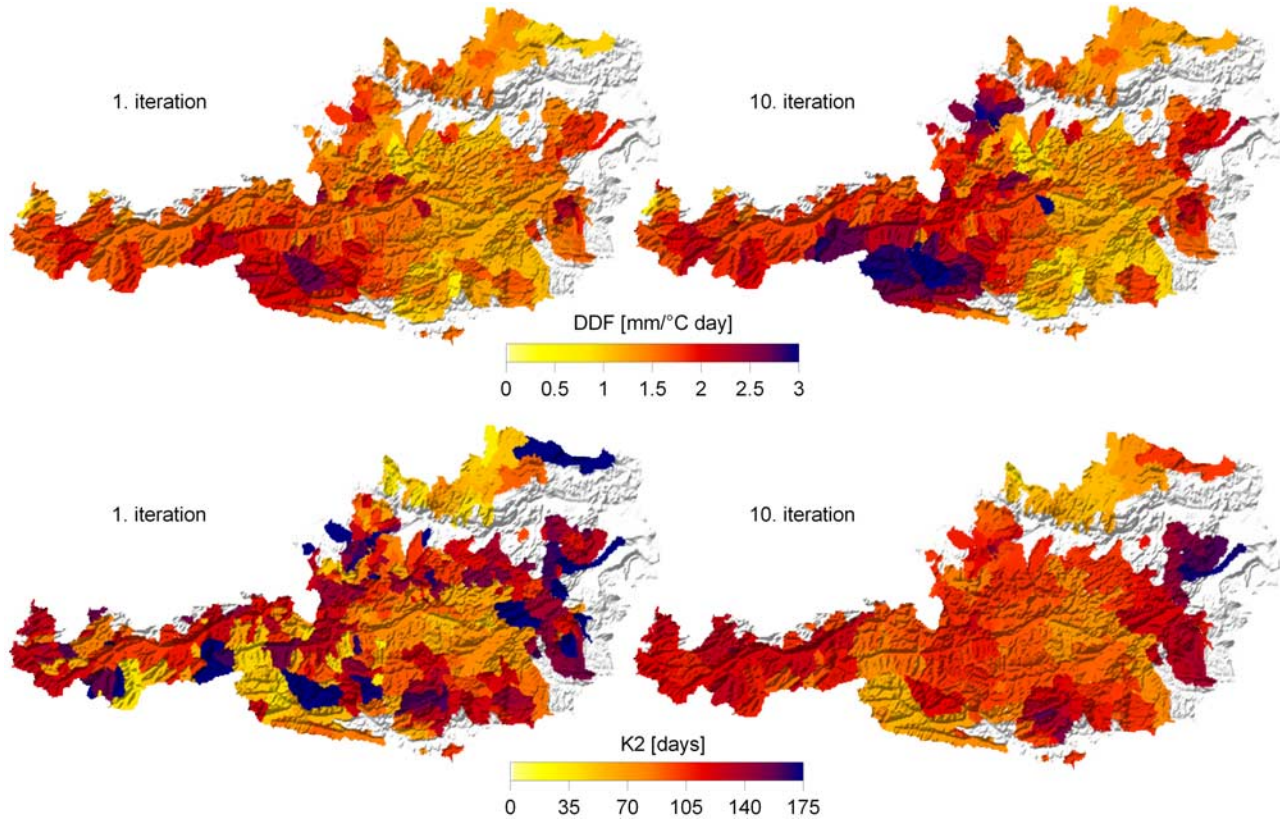

Figure 9.   Spatial patterns of the degree-day factor (DDF) and the slow storage coefficient (K_2). The first iteration is equivalent to local calibration without any regionalization; the tenth iteration is a result of the IRC procedure.
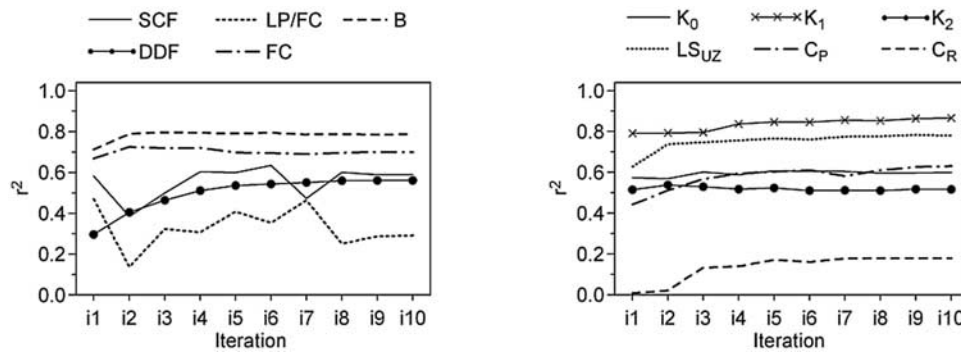
Figure 10. Change in parameter uncertainty during the iterations of regional calibration. Parameter uncertainty is represented by the coefficient of determination (r^2) between the model parameters calibrated from the two 11 year periods (1976–1986 and 1987–1997). Large r^2 indicate little uncertainty. The first iteration is equivalent to local calibration without any regionalization. (left) Snow and soil model parameters and (right) runoff routing model parameters.

routing parameters. The box size represents the difference of the 75th and 25th percentiles which is a measure of the spatial variability of the parameters. This spatial variability measure tends to increase with the number of iterations for the snow correction factor (SCF), the degree-day factor (DDF) and the LP to FC ratio (LP/FC) while it does not change much for the case of the field capacity (FC) and the nonlinearity (B) model parameters (Figure 5). The spatial variability tends to increase with the number of iterations for the fast storage coefficient (K_1), the storage state threshold (LS_{UZ}) and the percolation rate (C_P), and tends to decrease for the slow storage coefficient (K_2) and the outflow scaling parameter (C_R) (Figure 6). For all model parameters the median remains practically unchanged during the regional calibration, except the LS_{UZ} model parameter which decreases from 51 mm in the first iteration to 41 mm in the tenth iteration. The other calibration period gives similar results in terms of the dependence of the spatial variability of the parameters on the number of iterations (Table 4). However, the medians of the B and FC parameters differ markedly. The medians of both parameters tend to be smaller in the 1976–1986 calibration period than in the 1987–1997 calibration period. These differences are likely related to generally different climatic conditions in these two periods.

[39] Figures 7 and 8 analyze the spatial variability of the parameters in terms of the variograms. The sill or overall level of the variogram is equivalent to the spatial variance of the parameters and hence a similar measure as the difference of the 75th and 25th percentiles of Figures 5 and 6. The increase in the sill and hence spatial variance of the DDF and LP/FC model parameters with the number of iterations in Figure 7 is consistent with Figure 5. The snow correction factor (SCF) shows a somewhat erratic behavior in Figure 7 which is due to a small number of catchments that exhibit SCF values much larger than the median. The sill of the FC parameter tends to decrease with the number of iterations pointing to a decrease in its spatial variability. The spatial variability of C_R and K_2 decreases with the number of iterations which is consistent with Figure 6. The variability of the other runoff routing parameters increases which is, again, consistent with Figure 6 and Table 4. In addition to the overall spatial variability the variogram shows the spatial coherence of the parameters which allows some interpretation of parameter uncertainty (see section 4.2).

[40] Examples of the change in the spatial patterns of the parameters as a result of the IRC are presented in Figure 9. For the case of the DDF, the spatial patterns become more diverse as the iterative calibration proceeds. In the south-
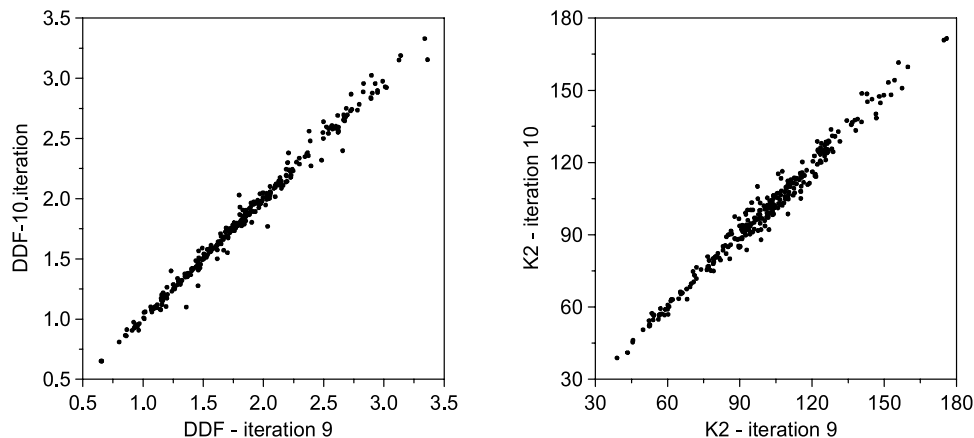

Figure 11. Convergence of the model parameters. Comparison of iterations 9 and 10 of the DDF and the slow storage coefficient (K_2) is shown. Each point relates to one catchment.
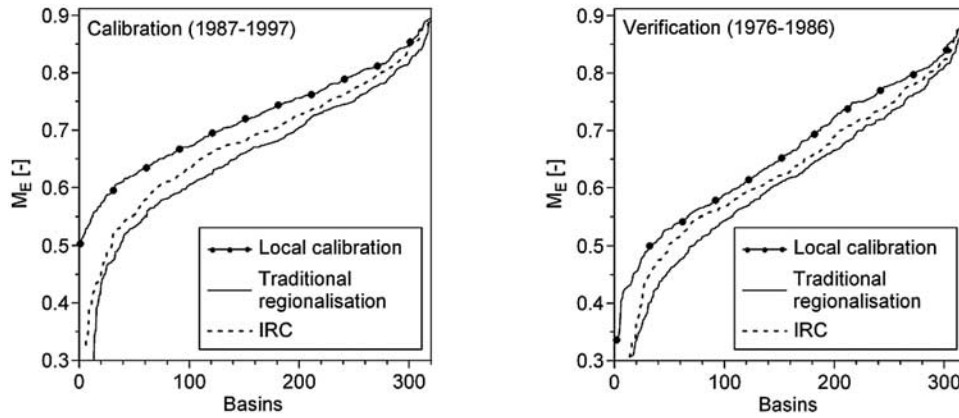
Figure 12. Cumulative distribution functions of the model efficiencies M_E of daily runoff for the 320 catchments using model parameters estimated by various methods. Local calibration, parameters p_j from the first iteration; traditional regionalization, parameters $p_{reg,j}$ from the first iteration; IRC, parameters $p_{reg,j}$ from the tenth iteration. Traditional regionalization and IRC both relate to the ungauged catchment case. (left) Calibration period and (right) verification period. The same parameters are used in the two periods.

west of Austria and in the northwest, consistent regions of large degree-day factors form. Both are plausible. The former is plausible as this is a high mountain region where large snow densities and hence high melt rates can be expected, the latter as it is a region with frequent rain-on-snow events where advection melt can give rise to enhanced melting [*Merz and Blöschl*, 2003]. The patterns of the low storage coefficient (K_2) (Figure 9, bottom) become more uniform. The large-scale patterns are those that can be identified by IRC, the more local-scale variability may be due to noise that is reduced during IRC. The two parameters are typical examples, the other parameters show either of the two aspects.

[41] The uncertainty of the model parameters is evaluated by a comparison of parameter values calibrated for the 1976–1986 period with those calibrated for the 1987–1997 period. The coefficient of determination (r^2) between the parameters from the two periods was calculated separately for each model parameter and iteration. One would expect reliably calibrated parameters to be similar in the two periods and hence exhibit large r^2 values. The coefficients of determination in Figure 10 and Table 5 suggest that the most uncertain parameters are the routing parameter C_R and the soil LP/FC ratio. The parameters associated with the least uncertainty are the soil parameters B and FC and the runoff parameters K_1 and LS_{UZ}. It is interesting that the soil parameters B and FC are those with noticeable differences between their median values obtained in the two periods (Table 4). This indicates that the climatic difference of the two calibration periods (the period 1976–1986 is noticeably drier than the period 1987–1997) resulted in different soil parameter values, but these values are very similar in their tendency. One of the motivations of the IRC procedure is to reduce parameter uncertainty. One would therefore hope that the coefficient of determination increases with the number of iterations. Figure 10 and Table 5 indicate that this is indeed the case for most of the parameters. The most significant increase in r^2 occurs for the snow DDF model parameter, from 0.30 in the first iteration (local calibration only) to 0.56 in the tenth iteration. Significant increases also

occur for C_P and C_R. The latter is a poorly defined parameter for which the r^2 from local calibration is essentially zero which means that C_R cannot be identified from local runoff data. For the tenth iteration r^2 is 0.18 which is still very small, but the increase indicates that the IRC procedure does provide information beyond the local runoff data. The LP/FC ratio is the only parameter that does not increase but varies erratically during the regional calibration iterations. The SCF parameter varies erratically as well. The SCF is designed to account for local catch deficit of the precipitation gauge. It is therefore likely that a regional estimation of this parameter may not be very robust.

[42] A total of ten iteration steps has been chosen in this study. The statistics of the parameters have indicated convergence within the ten iterations as the statistics of the final iterations are very similar. However, it is also of interest whether the individual parameters converge. As an example, Figure 11 shows a comparison of the degree-day factors and the slow storage coefficients (K_2) obtained in iterations 9 and 10 (parameters p_j in Figure 1). For a full convergence

Table 2. Model Efficiency of Runoff According to Nash-Sutcliffe (M_E), Runoff Volume Error (V_E), and Snow Model Error (Z_S) Obtained in Two Calibration Periods (1976–1986 and 1987–1997)[a]

Model Efficiency in Calibration Periods	Iteration 1 (Local Calibration)	Iteration 5	Iteration 10
M_E 1976–1986	0.71/0.19	0.71/0.19	0.71/0.19
M_E 1987–1997	0.72/0.13	0.73/0.13	0.73/0.13
V_E 1976–1986	−0.3/4.7	−0.0/6.4	0.0/6.4
V_E 1987–1997	0.3/7.4	0.0/8.2	0.0/7.8
Z_S 1976–1986	6.8/3.3	6.8/3.6	6.8/3.6
Z_S 1987–1997	6.5/3.7	6.5/3.7	6.4/3.6

[a]First value is median of model efficiency over the 320 catchments. Second value is difference of the 75th and 25th percentiles of model efficiencies, i.e., a measure of scatter. Iteration 1 is equivalent to local calibration without regionalization (i.e., using p_j). Iterations 5 and 10 are a result of the IRC procedure based on the p_j parameters.

Table 3. As in Table 2 but for the Verification Periods[a]

Model Efficiency in Verification Periods	Iteration 1 (Local Calibration)	Iteration 5	Iteration 10
M_E 1976–1986	0.66/0.20	0.65/0.19	0.66/0.19
M_E 1987–1997	0.69/0.14	0.70/0.14	0.70/0.14
V_E 1976–1986	−5.1/10.2	−6.2/11.6	−6.5/12.2
V_E 1987–1997	6.2/13.0	6.6/13.4	6.5/13.5
Z_S 1976–1986	6.4/3.6	6.4/3.9	6.4/3.9
Z_S 1987–1997	6.8/4.0	7.0/4.1	7.0/4.2

[a]The 1976–1986 verification period is based on the parameters of the 1987–1997 calibration period and vice versa.

one would expect the parameters to fall on the 1:1 line. This is approximately the case, indicating that these two parameters converge to good approximation within ten iterations. The other model parameters converge in a similar way.

3.3. Model Performance in Ungauged Catchments

[43] We examined the predictive accuracy of the proposed IRC method by jackknife cross validation. In this approach, we treated one gauged catchment as ungauged and simulated the water balance dynamics using model parameters $p_{reg,j}$ estimated from regional information only (Figure 1). In a second step, we estimated the model performance measures M_E, V_E, and Z_S by comparing the simulated and observed hydrographs as well as the simulated and observed snow cover. We repeated the analysis for each catchment in turn and calculated the statistics of these performance measures for the entire study region. We examined two cases by jackknife cross validation: Using the $p_{reg,j}$ parameters from iteration 1 which is equivalent to traditional regionalization; and using the $p_{reg,j}$ parameters from iteration 10 which is the result of the IRC procedure. The comparison of these performance measures with those for the locally calibrated case indicates what decrease of model performance one would have to expect when moving from gauged to ungauged catchments.

[44] The quantile statistics of model performance of traditional regionalization (iteration 1) and IRC (iteration 10) for the calibration and verification periods are presented in

Tables 6 and 7, respectively. The corresponding quantile statistics of the local calibration case are given in the first columns of Tables 2 and 3. The distribution functions of the Nash-Sutcliffe runoff model efficiencies M_E are shown in Figure 12 as an example. The median model efficiencies M_E for the local calibration, traditional regionalization and IRC are 0.71, 0.66, and 0.69 (1976–1986 calibration period) and 0.72, 0.67, and 0.69 (1987–1997 calibration period), respectively. This means that when moving from locally calibrated parameters to traditionally regionalized parameters one loses 0.05 in terms of median M_E. If one uses IRC instead of traditional regionalization one only loses 0.02–0.03 in terms of the median M_E. In other words, iterative regional calibration halves the loss in accuracy encountered when moving from gauged to ungauged catchments as compared to the traditional approach. This is also borne out in Figure 12 (left) as the cumulative distribution function for IRC (dotted line) lies halfway in between the local calibration and traditional regionalization curves. For the verification periods, the M_E runoff efficiencies are lower (Tables 3 and 7 and Figure 12, right), but there is a similar increase in the efficiency when using IRC as compared to traditional calibration.

[45] The scatter of the runoff volume errors V_E for the local calibration, traditional regionalization and IRC are 5, 16 and 13% (1976–1986 calibration period) and 7, 17 and 14% (1987–1997 calibration period), respectively. This means that, when moving from locally calibrated parameters to traditionally regionalized parameters, one loses 10% in terms of the scatter of the V_E but only 8% if one uses IRC instead. This is a small gain but it is consistent over both calibration and verification periods.

[46] Tables 6 and 7 also indicate an improvement in the snow simulation performance when using IRC instead of traditional regionalization but it is very small and hardly significant.

4. Discussion and Conclusions

4.1. Calibration and Verification Efficiencies

[47] Local calibration to gauged catchments yields median M_E runoff simulation efficiencies of 0.71 and 0.72 for the two calibration periods, and 0.66 and 0.69 for the two

Table 4. Median and 75th Minus 25th Percentile Difference of Model Parameters Calibrated for Two Independent 11 Year Periods (1976–1986 and 1987–1997) Over 320 Catchments[a]

Model Parameter	Calibration (1976–1986)			Calibration (1987–1997)		
	Iteration 1 (Local Calibration)	Iteration 5	Iteration 10	Iteration 1 (Local Calibration)	Iteration 5	Iteration 10
SCF	1.04/0.01	1.04/0.04	1.04/0.06	1.03/0.01	1.04/0.03	1.04/0.04
DDF, mm/°C d	1.7/0.6	1.9/0.9	2.0/0.8	1.6/0.4	1.7/0.6	1.8/0.7
LP/FC	0.95/0.02	0.93/0.03	0.92/0.07	0.94/0.02	0.93/0.05	0.92/0.08
FC, mm	114/102	129/95	132/87	145/130	160/110	166/111
B	2.7/5.2	2.0/4.9	1.8/4.5	4.1/7.8	4.2/7.8	4.0/8.2
K_0, days	0.5/0.1	0.4/0.2	0.4/0.2	0.5/0.1	0.4/0.2	0.4/0.1
K_1, days	9.3/2.0	9.7/4.8	10.4/5.8	9.4/1.9	10.0/5.0	10.5/6.6
K_2, days	104/77	105/34	100/34	104/66	104/31	100/28
C_P, mm/d	2.1/0.5	2.2/0.9	2.2/0.9	2.2/0.6	2.3/1.0	2.3/1.1
C_R, d²/mm	26.1/4.8	26.2/1.7	26.4/1.6	26.5/4.6	26.6/1.5	26.7/1.0
LS_{UZ}, mm	50/7	43/17	40/20	51/8	45/16	41/18

[a]The first value is the median, and the second value is the difference. Results are given for the first, fifth, and tenth iteration of the regional calibration procedure.

Table 5. Parameter Uncertainty Assessed by the Coefficient of Determination (r^2) Between the Model Parameters Calibrated for Two Independent 11 Year Periods (1976–1986 and 1987–1997)[a]

Model Parameter	Iteration 1 (Local Calibration)	Iteration 5	Iteration 10
SCF	0.58	0.60	0.59
DDF, mm/°C d	0.30	0.54	0.56
LP/FC	0.47	0.41	0.29
FC, mm	0.67	0.70	0.70
B	0.71	0.79	0.79
K_0, days	0.57	0.61	0.60
K_1, days	0.79	0.85	0.87
K_2, days	0.51	0.52	0.52
C_P, mm/d	0.44	0.60	0.63
C_R, d^2/mm	0.01	0.17	0.18
LS_{UZ}, mm	0.63	0.77	0.78

[a]Large r^2 indicate little uncertainty. Results are given for the first, fifth, and tenth iteration of the regional calibration procedure.

verification periods. These efficiencies are within the range of what one can expect at the regional scale. *Merz et al.* [2006] summarized studies of regional catchment modeling on a daily time step and noted that the median M_E efficiencies typically range between 0.6 and 0.7. There is a decrease in model performance when moving from the calibration to the verification periods, but the decrease is small and consistent with the results of other regional studies [e.g., *Young*, 2006]. The other performance measures (runoff volume error, snow cover simulation error) similarly yield statistics that are within the range of other studies [e.g., *Parajka et al.*, 2005]. In dedicated catchment studies one would expect better efficiencies. Dedicated studies are typically based on a higher density of the hydrographic network than what is available here and the model structure is often adjusted to the local particularities of the catchment [e.g., *Reszler et al.*, 2006]. For a regional study the model performance obtained here is considered very reasonable.

[48] The use of regional information in the iterative regional calibration (IRC) procedure does not decrease the model performance for the case of gauged catchments. One would potentially expect a decrease in model performance as, in addition to the local runoff data, regional information

Table 6. Model Efficiencies for the Ungauged Catchment Case[a]

Model Efficiency in Calibration Periods	Iteration 1 (Regionalization)	Iteration 10
M_E 1976–1986	0.66/0.20	0.69/0.19
M_E 1987–1997	0.67/0.16	0.69/0.15
V_E 1976–1986	−3.5/16.0	−2.0/13.4
V_E 1987–1997	0.1/16.9	−0.8/14.3
Z_S 1976–1986	7.0/3.6	6.8/3.9
Z_S 1987–1997	6.7/4.2	6.6/4.0

[a]Nash-Sutcliffe runoff model efficiency (M_E), runoff volume error (V_E), and snow model error (Z_S) obtained in two calibration periods (1976–1986 and 1987–1997). First value is median of model efficiency over the 320 catchments. Second value is difference of the 75th and 25th percentiles of model efficiencies, i.e., a measure of scatter. Iteration 1 is equivalent to traditional regionalization (i.e., using $p_{reg,j}$). Iteration 10 is a result of the IRC procedure based on the $p_{reg,j}$ parameters.

Table 7. As in Table 6 but for the Verification Periods[a]

Model Efficiency in Verification Periods	Iteration 1 (Regionalization)	Iteration 10
M_E 1976–1986	0.62/0.17	0.63/0.18
M_E 1987–1997	0.66/0.20	0.67/0.16
V_E 1976–1986	−8.1/16.1	−8.2/14.0
V_E 1987–1997	4.6/20.0	5.6/17.8
Z_S 1976–1986	6.6/3.3	6.7/3.2
Z_S 1987–1997	6.9/4.5	6.9/4.4

[a]The 1976–1986 verification period is based on the parameters of the 1987–1997 calibration period and vice versa.

is incorporated into the objective function and the combined objective function then represents a trade off between regional and local information. If the regional information is not fully consistent with the local runoff data, the ability of representing local runoff may decrease. *Fernandez et al.* [2000, pp. 698–699] reported that "the at site calibration approach is nearly always an improvement over the regional calibration approach," and *Szolgay et al.* [2003, p. 267] noted "the regional calibration methods slightly degraded the model performance in individual catchments when compared to the at site calibration, but on the average it proved to be comparable to the manual calibration method." The main motivation of regional calibration is to find more reliable parameters to be used for ungauged catchments, so a slight decrease of model performance for gauged catchments may be acceptable. In the light of these studies, the consistency of the efficiencies obtained here suggests that the proposed IRC approach allows to robustly calibrate the hydrologic model and provides a framework that accounts for local and regional information in a consistent way.

4.2. Variability and Uncertainty of Model Parameters

[49] Although the overall model efficiencies remain practically unchanged during the iterative calibration, the model parameters do change significantly. The analysis of the spatial variability both through the percentiles of the parameters (Table 4 and Figures 5 and 6) and through the variogram analyses (Figures 7 and 8) sheds light on the parameter uncertainty.

[50] The first indicator to parameter uncertainty is the overall spatial variability of the parameters. For some model parameters the spatial variability increases with the number of iterations, i.e., it increases as one moves from local calibration (iteration 1) to regional calibration (iteration >1). For example, the variabilities of the degree-day factor and the fast storage coefficient (K_1) increase. These are also the parameters where the split sample comparison (Table 5 and Figure 10) suggests that they are rather well defined. Apparently, for well defined parameters, the IRC approach exploits the ability to account for the spatially variable shape of the a priori parameter distributions. Also, the spatial patterns of these parameters obtained by the IRC method are more plausible. This has been illustrated for the case of the DDF for which consistently large values in the southwest and in the northwest of Austria could be interpreted by large snow densities in the high mountain region and frequent rain-on-snow events [*Merz and Blöschl*, 2003]. For other model parameters the spatial variability decreases as one moves from local calibration (iteration 1) to regional

calibration (iteration >1). Examples are the C_R parameter and the slow storage coefficient K_2. These are the most poorly defined parameters that cannot be estimated reliably from local runoff data. For the case of the C_R parameter, the 75th minus 25th percentile difference decreases from around 5 to less than 2 d²/mm as one moves from local calibration to regional calibration (tenth iteration), and at the same time the correlation increases from r^2 of 0.01 to 0.18. It appears that the reduction in the spatial variability of these parameters is related to the ability of the IRC method to reduce the noise in the estimates of poorly identifiable model parameters. This is important as it is one of the motivations for using a regional calibration procedure. There may also exist a compensation effect as not all the parameters are statistically independent [see, e.g., *Merz and Blöschl*, 2004, Figure 3]. For example, the fast storage coefficient K_1 and the slow storage coefficient K_2 appear to be interrelated, so a decrease in the spatial variability of one model parameter may be related to an increase in the spatial variability of another parameter.

[51] The second indicator of parameter uncertainty is the nugget of the variogram, i.e., the gamma value for very short distances. The nugget indicates how similar the parameters of adjacent catchments are, with small nuggets representing similar parameters. Noise that is purely due to artifacts in the parameter estimation process can be assumed to be spatially independent. In the variogram this would appear as a nugget effect. Figures 7 and 8 show indeed large nuggets for the poorly defined parameters and the nuggets decrease significantly with the number of iterations. The most obvious example is the C_R parameter which exhibits a very large nugget for the local calibration case that is dramatically reduced through the IRC. This is because the regional calibration reduces the noise that results from poor identifiability. To a lesser extent this reduction in the nugget also occurs with the field capacity (FC) and the nonlinearity (B) model parameters and with the storage state threshold (LS$_{UZ}$). The snow correction factor (SCF) is an interesting case as it is designed to account for the local catch deficit of the precipitation gauge. It is therefore likely that the regional estimation of this parameters may not add information. Indeed, the nugget of the variogram stays relatively large for all iterations (Figure 7, top left).

[52] It is also of interest to interpret the distance over which spatial correlations are present, i.e., the distance where the variogram reaches the sill. It appears that the shapes of the snow and soil variograms (Figure 7) are different from the shapes of the runoff routing variograms (Figure 8). The former tend to approach a sill at distances of around 50 km with a flat sill at larger distances. This means that very little correlation exists for distances larger than 50 km. In contrast, most of the runoff parameter variograms (Figure 8) do not reach a sill but increase consistently with distance. These differences in the shape can be interpreted in terms of the role of the parameters in the hydrological model. The snow and soil parameters tend to be local parameters, i.e., mainly represent vertical fluxes while the runoff routing parameters represent lateral flow processes associated with larger space scales. It is hence not surprising that the local parameters are only correlated over a limited distance while the nonlocal parameters exhibit correlations over much larger distances.

[53] Application of an iterative method also raises the issue of convergence. The comparison of consecutive iterations (Figure 11) showed that the parameters did not change much between iterations 9 and 10, so for the data used here the method seems to converge after ten iterations. Also, the change in the spatial variability of the model parameters suggested that noise in the poorly defined parameters can be reduced within ten iterations. In general, the number of iterations needed will depend on the regionalization method used. In this study, kriging has been used which is a linear method. For nonlinear regionalization methods the convergence may be slower.

[54] The uncertainty of model parameters has been quantified more directly by a split sample comparison for two independent 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 nonstationarity. However, this type of evaluation may only be used in a meaningful way if a large set of catchments is available for testing as is the case here. This comparison has indicated that the IRC procedure increases the similarity of the model parameters estimated from the two periods as compared to local calibration. This is consistent with the spatial analyses of the parameters. The similarity of the parameters in the two periods does not imply spatially more uniform parameters, rather it implies that the differences between the two parameter values relative to their spatial variability are smaller. A reduction in parameter uncertainty may also give rise to the hope that regionalization relationships between calibrated model parameters and physiographic catchment attributes could be better defined if the IRC method is used as compared to local calibration.

4.3. Model Performance in Ungauged Catchments and Outlook

[55] The evaluation of the predictive accuracy by jackknife cross validation indicated that the proposed iterative regional calibration method has the potential for improving model performance in ungauged catchments for both the calibration and verification periods. The main mechanism seems to be the constraint on the shape of the parameter distribution through regional calibration, which tends to translate into more accurate runoff simulations in catchments considered as ungauged. Apparently, for well defined parameters, the IRC approach exploits the ability of accounting for the spatially variable shape of the a priori parameter distributions, and for the poorly identifiable model parameters the IRC approach reduces the noise in the parameter estimates. When moving from locally calibrated parameters to traditionally regionalized parameters one loses 0.05 in terms of the median runoff simulation efficiency M_E. If one uses IRC instead of traditional regionalization one only loses 0.02–0.03 in terms of the median M_E. In other words, iterative regional calibration halves the loss in accuracy encountered when moving from gauged to ungauged catchments as compared to the traditional approach. The loss in accuracy is smaller than that obtained by various parameter regionalization methods of *Parajka et al.* [2005]. It is also smaller than the loss in mean runoff efficiency of 0.07 found by the regional calibration study of *Drogue et al.* [2002].

[56] Although the reduction in the loss of accuracy is noticeable in relative terms and is consistent across all calibration and verification periods, it is small in absolute terms. Typically, one can expect an improvement in the median M_E of 0.03 when using the IRC method. This is only ten percent of the difference of a perfect simulation ($M_E = 1.0$) and the typical model efficiencies of about 0.7. Striving for more accurate data sets and hydrological models that better account for local effects is hence probably as important as are improvements in the regionalization methods. While comparisons of the performance statistics as in this paper are useful for testing methods it should also be noted that for practical applications one is not interested in the median model efficiency over many catchments but in the model efficiency in one particular catchment. The results of this study suggest that the likelihood of obtaining more reliable model parameters increases when using IRC as compared to traditional calibration but it does not necessarily entail more reliable parameters in any one catchment.

[57] For reducing parameter uncertainty and improving model simulations in ungauged catchments 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 calibration to runoff alone. As an alternative to direct measurements of additional data, the use of "soft data" or qualitative information from field surveys has been suggested in the literature to constrain model parameters [e.g., *Blöschl*, 2005]. "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 and more formal methods of incorporating soft information have been proposed [e.g., *Seibert and McDonnell*, 2002]. In this study, snow data have been used in the calibration procedure as well as regional parameter information. The use of regional information to condition the a priori parameter distribution can be considered as a type of soft information in the calibration process. More local types of soft information that could be combined with the present approach are maps of saturation areas, for example. Combining local data with the regional calibration approach may perhaps help resolve some of the conflicts in the current philosophical debate related to regional calibration, of adjusting parameters to the local situation as opposed to adjusting parameters to the regional trend as is the case in regional calibration.

[58] There are a number of logical extensions of the method proposed here. The most obvious extension would be a test of alternative regionalization methods such as top kriging [*Skøien et al.*, 2006], the use of physiographic catchment characteristics in the regionalization procedure [*Parajka et al.*, 2005] or the use of pooling parameters from hydrologically similar catchments [*McIntyre et al.*, 2005]. Performance measures, such as peak flow statistics and time to peak, may help to better discriminate between alternative methods. Finally, it would be worth improving the model efficiency for the local calibration case, perhaps, by varying the model structure between catchments depending on regional runoff processes.

References

Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, *37*, 937–947.

Bergström, S. (1976), Development and application of a conceptual runoff model for Scandinavian catchments, *Univ. Lund Bull., Ser. A*, *52*, 134 pp.

Blöschl, G. (2005), Rainfall-runoff modelling of ungauged catchments, in *Encyclopedia of Hydrological Sciences*, pp. 2061–2080, John Wiley, Chichester, U. K.

Drogue, G., T. Leviandier, L. Pfister, A. EL Idrissi, J. F. Iffly, L. Hoffmann, F. Guex, B. Hingray, and J. Humbert (2002), The applicability of a parsimonious model for local and regional prediction of runoff, *Hydrol. Sci. J.*, *47*, 905–920.

Duan, Q., S. Sorooshian, and V. K. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*, 1015–1031.

Fernandez, W., R. M. Vogel, and A. Sankarasubramanian (2000), Regional calibration of a watershed model, *Hydrol. Sci. J.*, *45*, 689–707.

Gottschalk, L. (2002), Advances in observational hydrology—Field experiments and modelling, in *Proceedings of Workshop on the Prediction of Ungaged Basins (PUBs)*, edited by K. Takeuchi, pp. 26–34, Yamanashi Univ., Kofu, Japan.

Hlavčová, K., J. Szolgay, M. Čistý, S. Kóhnová, and M. Kalaš (2000), Estimation of mean monthly flows in small ungauged catchments, *Slovak J. Civ. Eng.*, *8*, 21–29.

Hundecha, Y., and A. Bárdossy (2004), Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalisation of a watershed model, *J. Hydrol.*, *292*, 281–295, doi:10.1016/j.jhydrol.2004.01.2002.

Klemeš, V. (1986), Operational testing of hydrological simulation models, *Hydrol. Sci. J.*, *31*, 13–24.

Lamb, R., J. Crewett, and A. Calver (2000), Relating hydrological model parameters and catchment properties to estimate flood frequencies from simulated river flows, paper presented at 7th National Hydrology Symposium, Br. Hydrol. Serv., Newcastle-upon-Tyne, U. K.

Lindström, G. (1997), A simple automatic calibration routine for the HBV model, *Nord. Hydrol.*, *28*(3), 153–168.

Madsen, H. (2003), Parameter estimation in distributed hydrological catchment modeling using automatic calibration with multiple objectives, *Adv. Water Resour.*, *26*, 205–216.

McIntyre, N., H. Lee, H. Wheater, A. Young, and T. Wagener (2005), Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, *41*, W12434, doi:10.1029/2005WR004289.

Merz, R., and G. Blöschl (2003), A process typology of regional floods, *Water Resour. Res.*, *39*(12), 1340, doi:10.1029/2002WR001952.

Merz, R., and G. Blöschl (2004), Regionalisation of catchment model parameters, *J. Hydrol.*, *287*, 95–123, doi:10.1016/j.jhydrol.2003.09.028.

Merz, R., G. Blöschl, and J. Parajka (2006), Regionalization methods in rainfall-runoff modelling using large catchment samples, in *Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment—MOPEX*, edited by V. Andréassian et al., *IAHS Publ.*, *307*, 117–125.

Mészáros, I., and P. Miklánek (2006), Calculation of potential evapotranspiration based on solar radiation income modeling in mountainous areas, *Biologia*, *61*, suppl. 19, 284–288.

Mészáros, I., P. Miklánek, and J. Parajka (2002), Solar energy income modelling in mountainous areas, in *ERB 2002 and NE FRIEND 5 International Conference—Interdisciplinary Approaches in Small Catchment Hydrology: Monitoring and Research*, edited by L. Holko and P. Miklanek, pp. 127–135, U. N. Educ. Sci. and Cultural Organ., Paris.

Mroczkowski, M., G. P. Raper, and G. Kuczera (1997), The quest for more powerful validation of conceptual catchment models, *Water Resour. Res.*, *33*, 2325–2336.

Parajka, J., R. Merz, and G. Blöschl (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,"* pp. 299–306, Slovak Acad. of Sci., Bratislava.

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

Parajka, J., R. Merz, and G. Blöschl (2007), Uncertainty and multiple objective calibration in regional water balance modeling—Case study in 320 Austrian catchments, *Hydrol. Processes*, *21*, 435–446.

Pebesma, E. J. (2001), Gstat user's manual, 100 pp., Dep. of Phys. Geogr., Utrecht Univ., Utrecht, Netherlands.

Reszler, C., J. Komma, G. Blöschl, and D. Gutknecht (2006), Ein Ansatz zur Identifikation flächendetaillierter Abflussmodelle für die Hochwasservorhersage, *Hydrol. Wasserbewirtschaftung*, *50*(5), 220–232.

Seibert, J., and J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, *38*(11), 1241, doi:10.1029/2001WR000978.

Sivapalan, M., et al. (2003), IAHS decade on predictions in ungauged basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, *48*, 857–880.

Skøien, J., R. Merz, and G. Blöschl (2006), Top-kriging—Geostatistics on stream networks, *Hydrol. Earth Syst. Sci.*, *10*, 277–287.

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

van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan (2006), A global sensitivity analysis tool for the parameters of multi-variable catchment models, *J. Hydrol.*, *324*, 10–23, doi:10.1016/j.jhydrol.2005.09.008.

Wagener, T., and H. S. Wheater (2006), Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, *J. Hydrol.*, *320*, 132–154, doi:10.1016/j.jhydrol.2005.07.015.

Western, A., and R. Grayson (2000), Soil moisture and runoff processes at Tarrawarra, in *Spatial Patterns in Catchment Hydrology: Observations and Modelling*, edited by R. Grayson and G. Blöschl, chap. 9, pp. 209–246, Cambridge Univ. Press, Cambridge, U. K.

Wösten, J. H. M., Y. A. Pachepsky, and W. J. Rawls (2001), Pedotransfer functions: Bridging the gap between available basic soil data and missing soil hydraulic characteristics, *J. Hydrol.*, *251*, 123–150, doi:10.1016/S0022-1694(01)00464-4.

Young, A. R. (2006), Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model, *J. Hydrol.*, *320*, 155–172, doi:10.1016/j.jhydrol.2005.07.017.

––––––––––––––––––––

G. Blöschl, R. Merz, and J. Parajka, Institute for Hydraulic and Water Resources Engineering, Vienna University of Technology, Karlsplatz 13, A-1040 Vienna, Austria. (bloeschl@hydro.tuwien.ac.at; merz@hydro.tuwien.ac.at; parajka@hydro.tuwien.ac.at)