

Time stability of catchment model parameters: Implications for climate impact analyses

Ralf Merz,¹ Juraj Parajka,² and Günter Blöschl²

Received 30 April 2010; revised 25 October 2010; accepted 1 November 2010; published 19 February 2011

[1] Climate impact analyses are usually based on driving hydrological models by future climate scenarios, assuming that the model parameters calibrated to past runoff are representative of the future. In this paper we calibrate the parameters of a conceptual rainfall-runoff model to six consecutive 5 year periods between 1976 and 2006 for 273 catchments in Austria and analyze the temporal change of the calibrated parameters. The calibrated parameters representing snow and soil moisture processes show significant trends. For example, the parameter controlling runoff generation doubled, on average, in the 3 decades. Comparisons of different subregions, comparisons with independent data sets, and analyses of the spatial variability of the model parameters indicate that these trends represent hydrological changes rather than calibration artifacts. The trends can be related to changes in the climatic conditions of the catchments such as higher evapotranspiration and drier catchment conditions in the more recent years. The simulations suggest that the impact on simulated runoff of assuming time invariant parameters can be very significant. For example, if using the parameters calibrated to 1976–1981 for simulating runoff for the period 2001–2006, the biases of median flows are, on average, 15% and the biases of high flows are about 35%. The errors increase as the time lag between the simulation and calibration periods increases. The implications for hydrologic prediction in general and climate impact analyses in particular are discussed.

Citation: Merz, R., J. Parajka, and G. Blöschl (2011), Time stability of catchment model parameters: Implications for climate impact analyses, *Water Resour. Res.*, 47, W02531, doi:10.1029/2010WR009505.

1. Introduction

[2] Climate change impacts on water resources are usually analyzed by a scenario approach which consists of four steps: (1) choosing one or more climate change scenarios of global circulation models (GCMs), (2) downscaling the GCM output to the scale of interest, (3) running a hydrological model using the downscaled GCM output, and (4) comparing the model simulations for the current and scenario climates to analyze the impacts [Bergström *et al.*, 2001]. Each of these steps involves various sources of uncertainty. Kay *et al.* [2009] suggested that for their UK catchments, GCM structure was the largest source of uncertainty for simulating floods, followed by emission scenarios and hydrological modeling. Wilby and Harris [2006] found similar results for low flows. This is hardly surprising as GCM results of future climate cannot be validated against data. However, as pointed out by Blöschl *et al.* [2007] and Blöschl and Montanari [2010], the relative magnitudes of the hydrological model errors will depend on the time and space scales involved and may potentially be very significant.

[3] One particular concern with the hydrological modeling in impact analyses is the model parameters. The general idea of parameters in dynamic models is that they represent the stable catchment conditions while the rainfall and other inputs are the time-varying boundary conditions. Ideally, one would like to measure the model parameters in the field, but as has been exhaustively discussed in the literature, a fully reductionist approach of using parameter values observed in the field without any calibration for hydrological predictions has many limitations, so in most cases, some level of model calibration will be useful to reduce bias [e.g., Beven, 2000]. Because of the general modeling idea of parameters representing stable catchment conditions, one would hope that calibrated model parameters are not sensitive to the change in climatic conditions. However, if climate conditions change, there are several reasons that model parameters may potentially change as well. First, calibration parameters tend to compensate model structure problems and data problems and hence may change for different calibration periods. Wagener *et al.* [2003] estimated the parameters of a rainfall-runoff model for different parts of the hydrograph by dynamic identifiability analysis (DYNIA). They found that some of the parameters (such as the maximum soil moisture storage capacity (C_{MAX})) significantly changed for periods with and without rainfall events. There were periods when the estimated C_{MAX} value was around 100, but when there was a significant hydrograph response following a dry period, C_{MAX} was around 500. They attributed these

¹Department for Catchment Hydrology, Helmholtz Centre for Environmental Research (UFZ), Halle, Germany.

²Institute for Hydraulic and Water Resources Engineering, Vienna University of Technology, Vienna, Austria.

changes to problems with the model structure. *Juston et al.* [2009], in contrast, reported that using weekly, monthly, and quarterly subperiods of the data for calibration yielded similar parameters as the full 3 year period.

[4] A second reason for temporal change in calibration parameters may be that some of the model parameters indeed represent transient catchment conditions. For example, changes in the land use may result in changes in the model parameters [Brown et al., 2005; Andréassian et al., 2003]. However, trends in the calibrated model parameters may potentially also be related to climate fluctuations. One example in event-based runoff models is the runoff coefficient. As soil moisture changes in response to climate variations, the runoff coefficients will also change [Merz et al., 2006; Merz and Blöschl, 2009a]. There are similar parameters in continuous runoff models related to runoff generation that may also vary in response to climate fluctuations. However, just how calibrated model parameters change with time is currently not very well understood [Wagner et al., 2010]. In order to account for any changes in the parameters caused by changes in the catchment characteristics, a unique relationship between the two would be needed, but often, there are complex correlations among the parameters and with various catchment characteristics [Wagner, 2007]. It is therefore not straightforward to account for temporal changes in the model parameters, and in most impact analyses it is simply assumed that the model parameters do not change with time. This assumption may produce large errors in simulated runoff if climate and/or catchment conditions change much with time. *Wilby* [2005], for example, found that the uncertainty in projected river flow changes due to the choice of calibration period was similar to the uncertainty due to future greenhouse gas emission scenarios. *Lubès-Niel et al.* [2003] analyzed 17 catchments in Africa where rainfall and runoff significantly decreased over the years. They calibrated a conceptual rainfall-runoff model to different time periods and analyzed the time trends of the calibrated model parameters. In about two thirds of the catchments, the parameters were time stable. In the other catchments they did change with time, but there were no obvious relationships between the parameter trends and the climate trend. The issue of identifying and understanding the time stability of catchment model parameters is therefore far from resolved.

[5] The aim of this paper is to analyze the temporal stability of the parameters of a conceptual rainfall-runoff model and assess the implications for climate impact analyses. Specifically, we address the following research questions: (1) Are there trends in the parameter values when calibrating them to different time periods, and can this trend be interpreted in terms of changes in climate? (2) What errors in the estimates of future runoff will be incurred when assuming that the model parameters are time stable, i.e., when using the parameters from the past for predicting runoff in future decades?

[6] We calibrate a semidistributed conceptual rainfall-runoff model to runoff data from 273 Austrian catchments for different 5 year periods between 1976 and 2006 during which the climatic variables have substantially changed. We then analyze the time trends of the calibrated parameters and estimate what are the effects on low, medium, and high flows of assuming that the model parameters do not change with time.

2. Data and Study Region

2.1. Study Region and Data Set

[7] This study was carried out in Austria using data from the period 1976–2006. Austria is flat or undulating in the east and north and alpine in the west and south. Elevations range from 115 to 3797 m above sea level. 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 Austrian climate is dominated by the Alps as a barrier of continental climate in the north, in which humid westerly winds prevail, and the influence of weather patterns from the Adriatic Sea in the south. Along the rim of the high Alps, mean annual precipitation is almost 3000 mm/yr, which is due to orographic effects. In the inner alpine regions, mean annual precipitation is much lower, and the interannual hydrological variation is dominated by snow accumulation and melt. The lowest precipitation of less than 400 mm/yr occurs in the eastern flatlands, which are in the Pannonian climate zone, with warm and dry summers and cold winters without significant snowfall. The data set used in this study includes measurements of daily precipitation and snow depths at 1091 stations [Weilguni, 2003; Gattermayr et al., 2003] and daily air temperatures at 212 climatic stations [Auer and Böhm, 2003]. The precipitation data were checked for time trends in the measurement errors by comparing them against the Hinstalp data set [Auer et al., 2007], and no significant biases were found in the regional trends. Daily runoff data from 273 gauged catchments were used with areas ranging from 10 to 130,000 km² and a median of 243 km². All discharge data had been screened in previous studies [Merz and Blöschl, 2004; Parajka et al., 2007b; Merz et al., 2008, 2009], and catchments with anthropogenic impacts such as dams were not included in the data set. Water withdrawals do not have any significant effects in the catchments used here.

[8] The climatic data time series for the analyses in this paper 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., 2005] using daily air temperature and potential sunshine duration calculated by the Solei-32 model that incorporates shading by surrounding terrain [Mészáros et al., 2002]. In a second step, all catchments were subdivided into 200 m elevation zones. Time series of daily precipitation, air temperature, potential evapotranspiration, and snow cover were then calculated for each of the elevation zones to be used in the water balance simulations. If the average snow depth within a zone was greater than 5 cm, the zone was considered to be snow covered; otherwise, it was considered to be snow free. The choice of the 5 cm threshold was based on the analyses of Parajka and Blöschl [2008]. All the analyses in this paper were carried out for the set of 273 catchments to span a wide variety of catchment types. To assist in the interpretations, the Austrian catchments were classified by the long-term ratio of

potential evapotranspiration and precipitation used in the Budyko curves [Budyko, 1974]. According to Budyko, catchments with a ratio larger than unity are termed water limited or arid, while catchments with a ratio smaller than unity are termed energy limited or humid. In Austria most catchments are energy limited [see, e.g., Merz and Blöschl, 2009a, Figure 3]. To help make the results of the study transferable to other climatic regions, the analyses in this paper are given for the entire set of catchments and for two subclasses with a ratio of potential evapotranspiration and precipitation larger than 0.6 (termed “drier” catchments) and a ratio less than 0.35 (termed “wetter” catchments). Out of the 273 catchments, 80 were classified as wetter catchments, and 48 were classified as drier catchments. The drier catchments are mainly located in the flat regions in the east of Austria, while the wetter catchments are mainly located in the Alps (Figure 1). The thresholds of 0.35 and 0.6 were chosen as a compromise between balancing the number of catchments in the groups and making the <0.35 and >0.6 groups clearly different with respect to the Budyko curve.

2.2. Evolution of Climate Conditions

[9] To help interpret the temporal stability of model parameters, the climatic variations in the period 1976 to 2006 were analyzed in a first step. Figure 2 shows the annual mean values of catchment precipitation P , air temperature, potential evapotranspiration, specific runoff Q (i.e., runoff divided by catchment area), the ratio of runoff and catchment precipitation Q/P , and the fraction of catchment area covered by snow. The black lines are the averages over the 273 catchments. The grey solid and grey dashed lines are the averages over the wetter catchments and the drier catchments, respectively. The average annual precipitation over all Austrian catchments varies from 1280 to 1360 mm/yr and has slightly increased over the 3 decades.

There are large differences between the drier and wetter catchments, but the spatial average tends to increase for all catchment groups. For the drier catchments, the spatial average has increased from 820 to 920 mm/yr, while for the wetter catchments, the spatial average has increased from 1550 to 1650 mm/yr. The air temperature trends are even more pronounced. Between 1976 and 2006, mean annual air temperatures have increased by almost 2°C , on average, for all the catchments. The increase is slightly larger for the wetter catchments (from 2°C to 4.5°C) and smaller for the drier catchments (7.4°C – 8.6°C). The air temperature increase in the European Alps in the past 3 decades has, in fact, been twice the average in the Northern Hemisphere [Auer et al., 2007]. Interestingly, the interannual variability of air temperature has decreased in the last decades [Auer et al., 2007]. Potential evapotranspiration (estimated from air temperature, Figure 2) has increased from about 530 to 570 mm/yr within the period 1976–2006 (averages over all Austrian catchments).

[10] No temporal trend in the mean annual runoff Q is apparent in Figure 2. This applies to the average of all catchments as well as the wetter and drier subregions. The increase in precipitation over the last decades seems to have been compensated by an increase in evapotranspiration, so runoff remained rather constant. This is a very important finding, as impact studies would usually predict an increase in runoff for an increase in precipitation as shown in Figure 2 (top left), but this was apparently not the case. The average runoff of the drier catchments is only about a quarter of that in the wetter catchments, but the time patterns are similar. Q_{95} low flows (the discharge exceeded 95% of the time) and Q_5 high-flow quantiles (the discharge exceeded 5% of the time) do not show time trends either. Because of the trend in precipitation and lack of trend in runoff, the ratio of Q/P shows a decreasing trend (Figure 2).

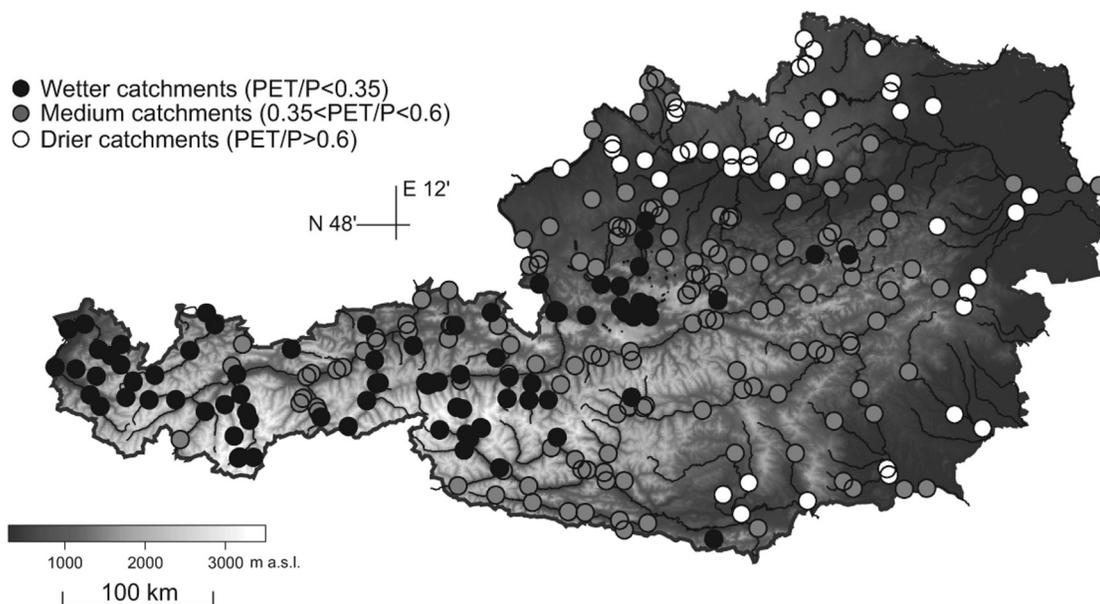


Figure 1. Location of the catchment outlets and classification into drier catchments (white), wetter catchments (black), and medium catchments (grey) depending on the ratio of potential evapotranspiration and precipitation.

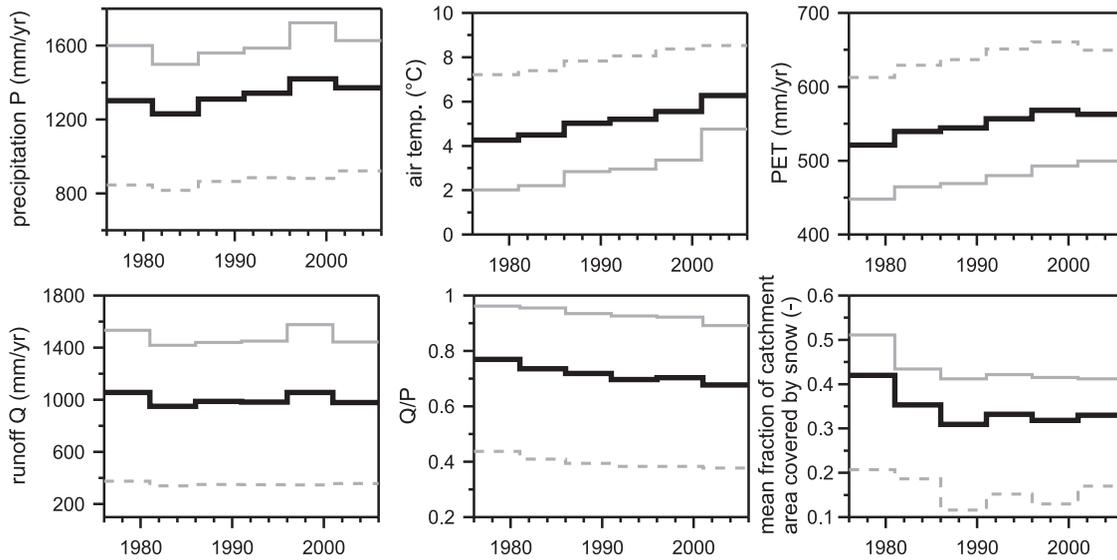


Figure 2. Five year mean annual values of climatic variables averaged over the 273 Austrian catchments (black lines), averaged over the wetter catchments (solid grey lines), and averaged over the drier catchments (dashed grey lines).

The average fraction of the catchment area covered by snow is much lower in the drier catchments than in the wetter catchments as the former are located in the lowlands while the latter are in the Alps. There is a small but consistent decrease in snow cover over the decades that is likely related to the increasing mean air temperatures.

3. Model and Methodology

3.1. Model Structure

[11] The model used in this paper is a semidistributed conceptual rainfall-runoff model with a structure similar to the Hydrologiska Byråns Vattenbalansavdelning (HBV) model [Bergström, 1995]. The model equations are given in the appendix of Parajka *et al.* [2007a]. 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. If air temperature is above a threshold temperature T_R , precipitation is considered to occur as rainfall; below a threshold temperature T_S , it is considered to occur as snowfall, and it is a mix of rain and snow in between. 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 evapotranspiration is at its potential rate, termed the limit for potential evapotranspiration LP, and a parameter in the nonlinear function relating runoff generation to the soil moisture state, termed the nonlinearity parameter B . If B is large, a small amount of rainfall contributes directly to runoff during an event. Conversely, if B is small, direct runoff is large. 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 (which may be interpreted as interflow) based on a fast storage coefficient K_1 ; perco-

lation to the lower zone with a constant percolation rate C_P ; and, if a threshold of the storage state LS_{UZ} is exceeded, an additional outlet (which may be interpreted as overland flow) 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 .

[12] The model was run for all 273 gauged catchments in Austria. Time step is 1 day. Each catchment was subdivided into elevation zones of 200 m vertical range. 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 were 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, Parajka *et al.* [2007b] performed a sensitivity analysis for Austrian catchments (including most of the catchments of this study) which showed that the results were sensitive to many of the model parameters in some catchments but insensitive in others. Three parameters that were among those that generally showed the least sensitivity were preset ($T_R = 2^\circ\text{C}$, $T_S = 0^\circ\text{C}$, $C_R = 25 \text{ d}^2/\text{mm}$) and 11 parameters (Table 1) were estimated by calibration.

3.2. Model Calibration and Verification

[13] To analyze the time stability of the model parameters, we calibrated the model parameters to observed runoff for six consecutive 5 year periods between 1976 and 2006.

[14] The calibration makes use of an objective function that involves the Nash and Sutcliffe efficiency ME of runoff and the Nash and Sutcliffe efficiencies of logarithmic runoff ME_{\ln} . In calibration procedures, the parameter values are usually bounded between two limits [Duan *et al.*, 1992], and otherwise, no a priori assumptions are made

Table 1. A Priori Distribution of Parameter Values^a

Model Parameter j	Definition	Model Component	p_l	p_u	α	β	p_{\max}
SCF	snow correction factor (dimensionless)	snow	1.0	1.5	1.2	4.0	1.03
DDF	degree-day factor (mm/°C day)	snow	0.0	5.0	2.0	4.0	1.25
T_M	melt temperature (°C day)	snow	-1.0	3.0	2	4	0.0
FC	maximum soil moisture storage (mm)	soil	0.0	600	1.1	1.5	100
LP/FC	ratio of limit for potential evapotranspiration and FC (dimensionless)	soil	0.0	1.0	4.0	1.2	0.94
B	nonlinearity parameter of runoff generation (dimensionless)	soil	0.0	20	1.1	1.5	3.4
K_0	storage coefficient of additional outlet (days)	runoff	0.0	2.0	2.0	4.0	0.5
K_1	fast storage coefficient (days)	runoff	2.0	30	2.0	4.0	9.0
K_2	slow storage coefficient (days)	runoff	30	250	1.05	1.05	105
LS _{UZ}	storage capacity threshold (mm)	runoff	1.0	100	3.0	3.0	50
C_P	percolation rate (mm/d)	runoff	0.0	8.0	2.0	4.0	2.0

^aParameters p_l and p_u are the lower and upper bounds of the parameter space used in all iterations, u and v are the initial parameters of the a priori distribution (equation (3)), and p_{\max} is the initial parameter value at which the a priori distribution is at a maximum.

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 ε_P based on a Beta distribution for each parameter:

$$\varepsilon_P = \sum_{j=1}^k \frac{f_{\max,j} - ff\left(\frac{p_j - p_{l,j}}{p_{u,j} - p_{l,j}}\right)}{f_{\max,j}}, \quad (1)$$

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

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

$$f(x|\alpha, \beta) = \frac{1}{\text{Beta}(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad (2)$$

$$0 < x < 1, \alpha > 0, \beta > 0,$$

with

$$\text{Beta}(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

[15] We assumed values of α and β for each parameter k on the basis of prior analyses of the distribution of parameter values from different calibration studies [Merz and Blöschl, 2004; Parajka et al., 2007a] (Table 1). We chose the lower and upper bounds of the parameters on the basis of literature values [Bergström, 1995; Seibert, 1997] and our own assessment. The bounds were the same for all catchments (Table 1).

[16] The entire objective function now consists of the following parts:

$$Z = w_1(1 - \text{ME}) + w_2(1 - \text{ME}_{\text{in}}) + w_3\varepsilon_P. \quad (3)$$

[17] The weights w_i represent the trade-off between the goodness of fit to runoff and the a priori parameter distributions. To assist in the choice of weights, sensitivity analyses were performed that showed that the model

results were only moderately sensitive to the choice of weights. On the basis of our prior experience with hydrological modeling in Austria [Merz et al., 2009] we set the weights to $w_1 = 0.4$, $w_2 = 0.4$, and $w_3 = 0.2$, which gives a relative importance of 80% to matching the runoff data and 20% to respecting the a priori distribution of the model parameters. This is because, on average, over the 273 catchments, the sum of the two first terms in equation (3) is similar to the third term. This objective function was minimized using the shuffled complex evolution (SCE-UA) method [Duan et al., 1992]. Model parameters were calibrated to runoff from 1 November to 31 October for each calibration period. Warm-up periods from January to October were used in all calibrations.

[18] Model performance was judged by comparing simulated and observed runoff in terms of the model efficiencies ME for each verification period that was not used for calibration. The second measure of model performance used here is the volume error VE, which is a measure of bias and is defined as

$$\text{VE} = \frac{\sum_{i=1}^n Q_{\text{sim},i} - \sum_{i=1}^n Q_{\text{obs},i}}{\sum_{i=1}^n Q_{\text{obs},i}}. \quad (4)$$

[19] VE = 0 implies no bias, and values larger and smaller than 0 imply an overestimation and an underestimation of the total runoff volume, respectively.

[20] In Figure 3 (top left) the spatial mean of the Nash Sutcliffe efficiencies ME of the 273 catchments calibrated to consecutive 5 year periods are plotted against the period of calibration. The spatial mean of ME of all catchments varies between 0.74 and 0.77 for the 5 year periods, but no trend is apparent. This means that the model parameters can be calibrated equally well to each period. In most calibration periods, the mean Nash Sutcliffe efficiencies of the drier catchments are smaller than those of all catchments. Most of the drier catchments are located in the flatlands of northeastern Austria, where the runoff regime is rainfall dominated. In contrast, most of the wetter catchments are located in the alpine western part of Austria, where the runoff regime is snow dominated. Merz et al. [2009] and T. Nester et al. (Climate and catchment controls on the performance of regional flood simulations, submitted to *Journal of Hydrology*, 2010) showed that catchments with a

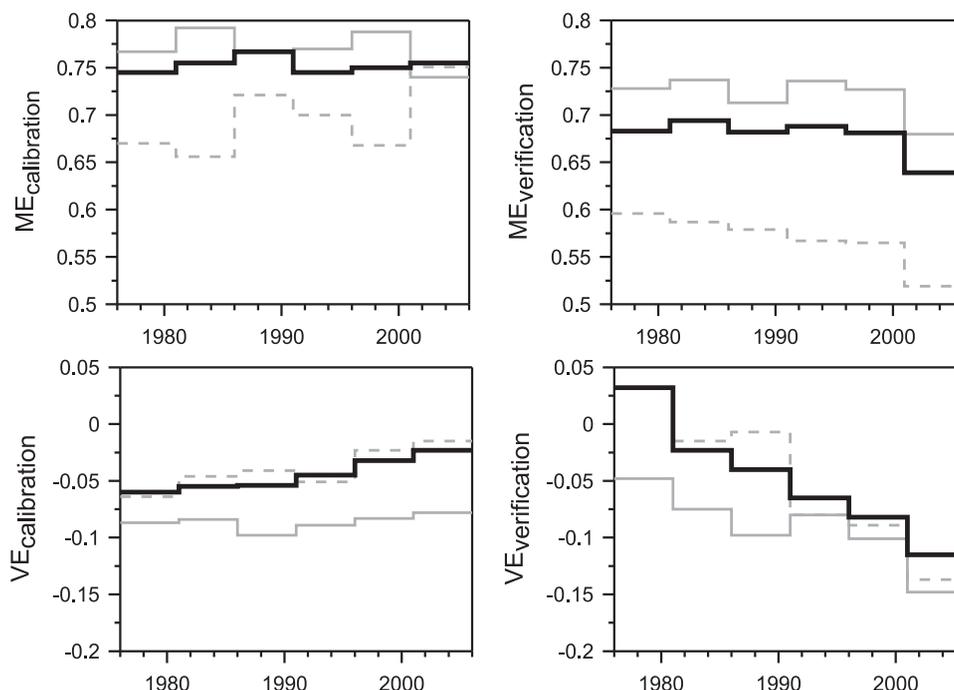


Figure 3. Nash-Sutcliffe model efficiencies (ME) and volume errors (VE) for 5 year calibration and verification periods averaged over the 273 Austrian catchments (black lines), averaged over the wetter catchments (solid grey lines), and averaged over the drier catchments (dashed grey lines). For the verification case, the mean ME and VE of the five remaining 5 year verification periods are plotted on the 5 year period used for calibration.

snow-dominated regime can be much better modeled than catchments with a rainfall-dominated regime because of the stronger seasonality and differences in the space and time scales of the rainfall regime. In Figure 3 (top right), the model efficiencies for the verification periods are plotted. Each parameter set of a 5 year calibration period was verified against the five remaining 5 year periods, and the average of the spatial mean of the remaining periods is shown. For example, the first segment of the thick black line in Figure 3 (top right) ($ME = 0.68$) shows the average spatial mean of the model efficiencies from the parameter set calibrated to the runoff in 1976–1981 and verified against the runoff in 1982–2006. Because of the averaging of five verification periods, the variability between the periods is smaller than for the calibration. The spatial mean of the verification ME varies between 0.64 and 0.69 when averaged over all catchments, and there is a slight trend. For the dry catchments there is a clear trend of decreasing efficiencies with time.

[21] The volume errors of the calibration case (Figure 3, bottom left) are, on average, smaller than 6%. Their magnitudes decrease slightly over the three decades. However, in the verification case, there is a very significant trend. The verification volume errors are about -12% when using model parameters calibrated to the most recent periods. This means that parameters from the recent warm years significantly underestimate runoff in the other years. The converse is true when using the parameters from the late 1970s, although the overestimation is much smaller.

[22] In this paper we are interested in the temporal changes in model parameters and their effects on simulated runoff in typical applications of conceptual hydrological

models. The model used here is among the most widely used soil moisture accounting schemes in the literature. The calibration procedure and performance measures are those used in numerous studies. Also, the model efficiencies found in this paper are similar to studies in the literature that are based on a similar number of catchments [see, e.g., Oudin *et al.*, 2008; Perrin *et al.*, 2001, 2008]. Merz *et al.* [2009] suggested that a minimum calibration period of 5 years is needed for this type of model, so we do not expect any overcalibration. We therefore believe that the model can be used for the time stability analysis in this paper.

4. Results

4.1. Time Stability of Model Parameters

[23] To analyze the general temporal behavior, the spatial averages of the model parameters for all catchments under study were analyzed. Figure 4 shows the spatial average of the snow correction factor SCF, the degree-day factor DDF, the maximum soil moisture storage FC, and the nonlinearity parameter B plotted against the calibration period. There are significant time trends. The snow correction factor SCF decreases with time. In the recent warmer years, less precipitation falls as snow, so one would expect the catch deficit of the precipitation gauges during snowfall to be smaller, which explains the decreasing trend of SCF. The degree-day factor DDF also tends to decrease with time. The average DDF in all catchments, calibrated to runoff data from the late 1970s, is about 1.8, while it is about 1.65 if calibrated to the most recent period. A possible interpretation of this is that the snowpacks, accumulated in

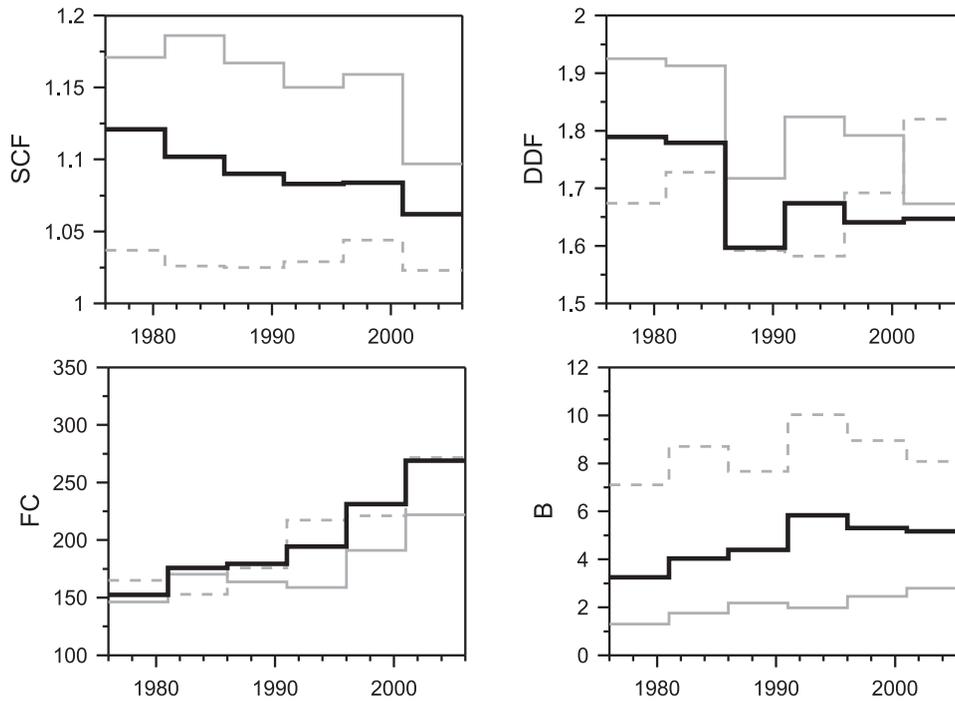


Figure 4. Model parameters (snow correction factor SCF, degree-day factor DDF, maximum soil moisture storage FC, and nonlinearity parameter of runoff generation B) of the 5 year calibration periods averaged over the 273 Austrian catchments (black lines), averaged over the wetter catchments (solid grey lines), and averaged over the drier catchments (dashed grey lines).

the winter, tend to be larger in the colder years and snow melt starts later in spring. Rain-on-snow events are more likely, and more radiation is available for melting snow and ice. The melt rates (relative to air temperature) then tend to be larger, which explains the larger DDF in the late 1970s. No trend in the DDF values is found for the drier catchments, although the mean values vary strongly between the different periods. The runoff regimes in the drier catchments, which are mainly located in the flatter eastern part of Austria, are rainfall dominated, and rain-on-snow events are less frequent. The occurrence of single rain-on-snow events in the different calibration periods therefore influences the calibrated DDF values, which increases the observed temporal variability of DDF. The average values of the maximum soil moisture storage capacity FC strongly increase with time. Evapotranspiration would be expected to increase because of the significant increase in air temperatures. The interpretation of this is that the soils tend to be drier, and on average, more water can be stored in the soils. This is reflected by larger calibration values of FC in the more recent warmer years. Similarly, the nonlinearity parameter B increases with time. B doubles from about 3 to almost 6 in the 3 decades. This increase is apparently related to more linear runoff generation and a lower fraction of rain that becomes runoff in the more recent years (Figure 2). This is plausible as the drier catchments have larger values of B (Figure 4), so as the catchments get slowly drier, B increases. The spatial coefficients of variations of the four model parameters are around 0.15, 0.3, 0.6, and 1.1 (not shown here) and do not change much in time. Also, the trends are consistent between all three catchment groups in Figure 4. This suggests that the trends

of the spatial averages shown in Figure 4 are representing changes in the hydrological conditions, taking place in most catchments, and are not an artifact of parameter uncertainty due to calibration. The other parameters (TM and LS_{UZ}) do not show any significant temporal trend (not shown here), while the parameter LP describing the limit for potential evaporation slightly decreases with time. The latter trend is related to increases in the evapotranspiration over the years.

[24] It is now of interest to understand whether these trends can be explained by climatic variability. For each catchment, the temporal correlation between the calibrated model parameters and one of a number of hydroclimatic indicators was estimated, each of them representative of one 5 year period. This means that six data points were used in each regression. As the model parameters and the climate indicators of the six calibration periods are not necessarily normally distributed, the Spearman rank correlation coefficient r_s was used here to measure the dependence of the model parameters on the climate indicators:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \quad d_i = rk(x_i) - rk(y_i), \quad (5)$$

where $rk(x_i)$ is the rank of x_i , where the highest value has rank 1 and the lowest value has rank n . Spearman's r_s varies between -1 and 1 , where -1 represents a completely negative correlation and 1 represents a completely positive correlation. Completely uncorrelated pairs of data have a Spearman's r_s of 0. The spatial variability of these temporal correlation coefficients was then plotted as Box-Whisker

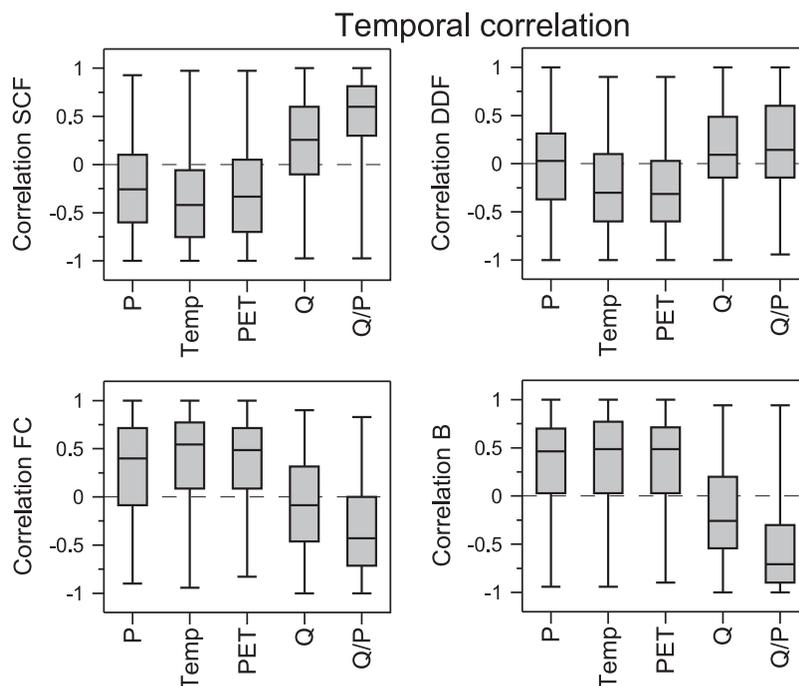


Figure 5. Spearman rank correlation coefficients (temporal correlations) between model parameters and climatic indicators for the six 5 year calibration periods, given separately for each catchment. The Box-Whisker plots show the spatial minimum, maximum, median, lower quartile, and upper quartile of the 273 Austrian catchments.

plots in Figure 5. The hydroclimatic indicators are mean annual precipitation (P), mean annual air temperature (Temp), mean annual potential evapotranspiration (PET), mean annual runoff (Q), and the mean annual ratio of runoff and precipitation (Q/P). On average, the snow correction factor SCF is negatively correlated to the mean annual values of precipitation, temperature, and PET, with correlation coefficients of -0.25 to -0.5 , and positively correlated to Q and Q/P , with correlation coefficients of up to 0.6 . A similar trend of a negative correlation to precipitation, air temperature, and PET and a positive correlation to Q and Q/P occurs for the degree-day factor, while the trend is opposite for the maximum soil moisture storage capacity FC and the nonlinearity parameter B . The spatial median of the correlation coefficients of FC and B to precipitation is around 0.3 , while it is around 0.5 for air temperature and potential evapotranspiration. It is interesting that FC and B are positively correlated to precipitation. This is because of the increasing trend in precipitation (Figure 2), which corresponds to the increasing trends in FC and B .

[25] There is a large variability in the correlations between the catchments for the same parameter–climate indicator combinations. This large variability may be partly related to parameter identifiability issues [Beven and Binley, 1992; Montanari, 2005]. The large variability may also be related to differences in the hydrological processes in the catchments. For example, for most catchments the B parameter is positively correlated with air temperatures because warmer years have less runoff because of increased evapotranspiration. However, there are some catchments where the correlations are negative. These are mostly catchments in the alpine parts of Austria where snow or glaciers play an important

role. In these catchments, years with above-average air temperatures are associated with above-average runoff, which then translates into negative correlations between air temperature and B . Overall, the grey range (25%–75% quantiles) of most of the Box-Whisker plots in Figure 5 is either completely positive or negative, suggesting that most model parameters are indeed meaningfully correlated to the climate indicators.

[26] To provide more insight into the plausibility of the temporal trends of the model parameters, independent data sets were analyzed. In Figure 6 (top) the average calibrated DDF and the average percentage of rain-on-snow days are plotted against the calibration period. A day was considered a rain-on-snow day when more than 30% of the catchment area was covered by snow (as estimated by the snow depth data), air temperature was above 1°C , and precipitation was more than 1 mm/d . The percentage of rain-on-snow days is therefore information independent of the model results. The percentage of rain-on-snow days tends to decrease in the warmer, more recent, periods, which is consistent with the decrease of the DDF, as one would expect larger DDF on rain-on-snow days than on sunny days because of the latent heat and long wave radiation. For the drier catchments (Figure 6, top right) the trends are not consistent, but here the DDF is not well defined as there is little snow in these catchments.

[27] The change in runoff generation over the years, indicated by a trend in the B values of the model, is compared with an independent analysis of event runoff coefficients. Merz *et al.* [2006] back calculated event runoff coefficients from hourly runoff data, hourly precipitation data, and estimates of snowmelt. This means their analysis

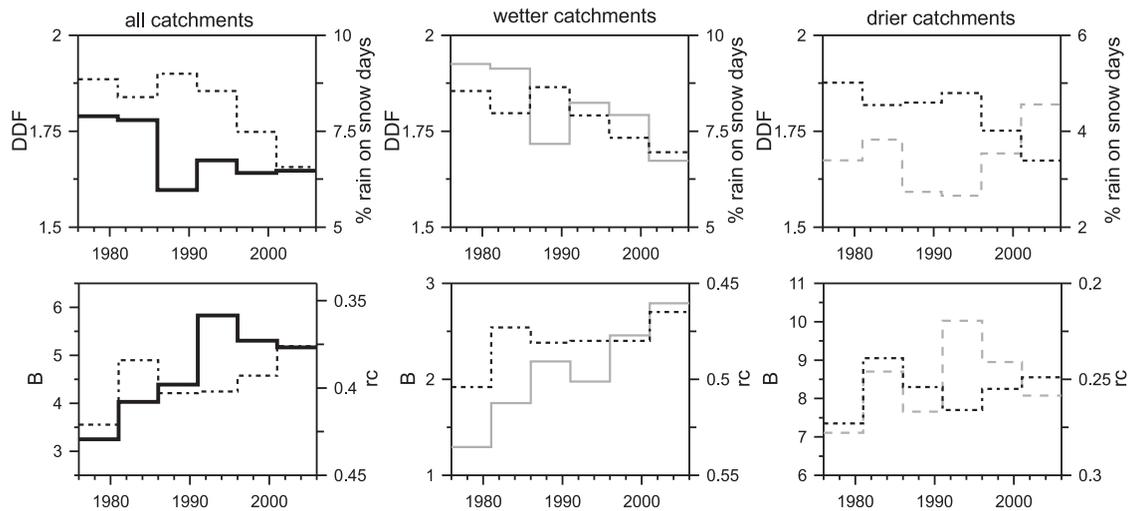


Figure 6. (top left) Degree-day factor (DDF; black solid line) for 5 year calibration periods and observed percent rain-on-snow days (dashed line) of the corresponding period averaged over all catchments. (bottom left) Nonlinearity parameter of runoff generation (B ; black solid line) for 5 year calibration periods and observed mean event runoff coefficients (rc ; dashed line) of the corresponding period averaged over all catchments. (middle and right) Corresponding graphs for the wetter and drier catchments. Note that the rain-on-snow days and the runoff coefficients are information independent of the model simulations.

is different from the one in this paper in terms of the model used (event model versus continuous model) and in terms of the time scale of the data (hourly versus daily). In this paper we adopted the methodology of Merz *et al.* [2006] to estimate event runoff coefficients for a total of 39,700 events in the period 1976 to 2006 and averaged them for the same 5 year periods used here. The mean event runoff coefficient (Figure 6) decreases from 0.42 for the period 1976–1981 to 0.38 for the period 2001–2006 for all catchments under study, while the mean B value increased from 3.2 to 5.2. The interpretation of this is that, due to increasing mean air temperatures, evapotranspiration has increased and catchments have become drier. More rainfall can be stored in the soils, so a smaller portion of rainfall contributes to direct runoff. The model accounts for this change in runoff generation by a change in the calibrated values of the B parameter.

[28] The spatial mean of four routing parameters, the three storage coefficients K_0 , K_1 , and K_2 and the parameter C_P controlling the percolation to the lower zone, are plotted against the period of calibration in Figure 7. K_1 and C_P slightly decrease, while for K_0 and K_2 no trend is apparent. The decrease in K_1 implies that the hillslope runoff response has slightly accelerated in recent years (9.5 as opposed to 11 days). The decrease in C_P in the most recent period implies that the groundwater recharge has decreased (1.5 as opposed to 2 mm/d), which is consistent with the lower catchment soil moisture to be expected in a warmer climate. The spatial coefficient of variation of the routing parameters is around 0.26, 0.3, 0.45, and 0.4 (not shown here) and does not change much in time, which indicates that the trends of the spatial averages in Figure 7 are consistent for most catchments.

[29] The spatial medians of the correlation coefficients between the routing parameters and the climate indicators vary from -0.25 to 0.1 (Figure 8). These are much smaller

values than those for the four soil moisture parameters, which have values of up to 0.6 (Figure 5). While the parameters of the soil moisture routine are obviously linked to changes in soil moisture over the last 30 years, driven by climate, the relationships for the routing parameters are less clear. This would be expected, as runoff routing is mainly controlled by the topography, the river network, geology, and soil type and, to a lesser degree, by the soil moisture state.

4.2. Trading Space for Time

[30] It is interesting to link the temporal relationship between model parameters and climatic forcing to the spatial variability of the model parameters for a given time period. If the increase in, e.g., B values for the last 30 years is caused by higher air temperatures, a similar change of the B values should be found in space if one moves from cold to warm catchments for the same time period. This is the idea of trading space for time. To analyze this, the spatial Spearman rank correlation coefficients of the model parameters and climate indicators were calculated for each of the six calibration periods separately, and the temporal variability over the six periods is given as Box-Whisker plots in Figures 9 and 10. Furthermore, to provide a general idea of the controls on the model parameters, the spatial correlation coefficients between model parameters and catchment attributes that do not change with time are also given. Merz and Blöschl [2009b] term this type of catchment attributes “static” to reflect their temporal stability within the time scale of the analysis. There are, of course, longer-term interactions of climate and these catchment characteristics related to landform and soil evolution [Merz and Blöschl, 2008a, 2008b], but they are considered small for the 30 years considered here. The static catchment attributes are catchment area, mean elevation, river network density (RND), average

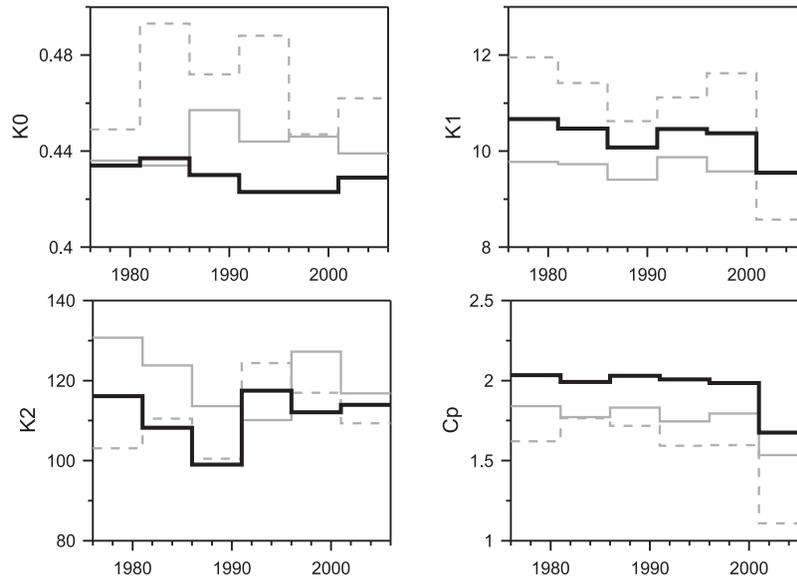


Figure 7. Model parameters (storage coefficients K_0 , K_1 , and K_2 and percolation rate C_p) of the 5 year calibration periods averaged over the 273 Austrian catchments (black lines), averaged over the wetter catchments (solid grey lines), and averaged over the drier catchments (dashed grey lines).

topographic slope, and, as the forested area has only slightly increased in the last decades [Jonas *et al.*, 1998], percent forest cover.

[31] For most model parameters and climate indicators the spatial correlation coefficients are, on average, in a range similar to or slightly higher than the temporal correlation coefficients. For example, the mean temporal

correlation coefficient between B and air temperature is about 0.5 (Figure 5), while the mean spatial correlation coefficient is about 0.7 (Figure 7). This may be attributed to the larger range of climate indicators in space than in time. The mean annual air temperatures have increased by almost 2°C during the time period of this analysis, while the mean annual air temperatures range from about 0°C in

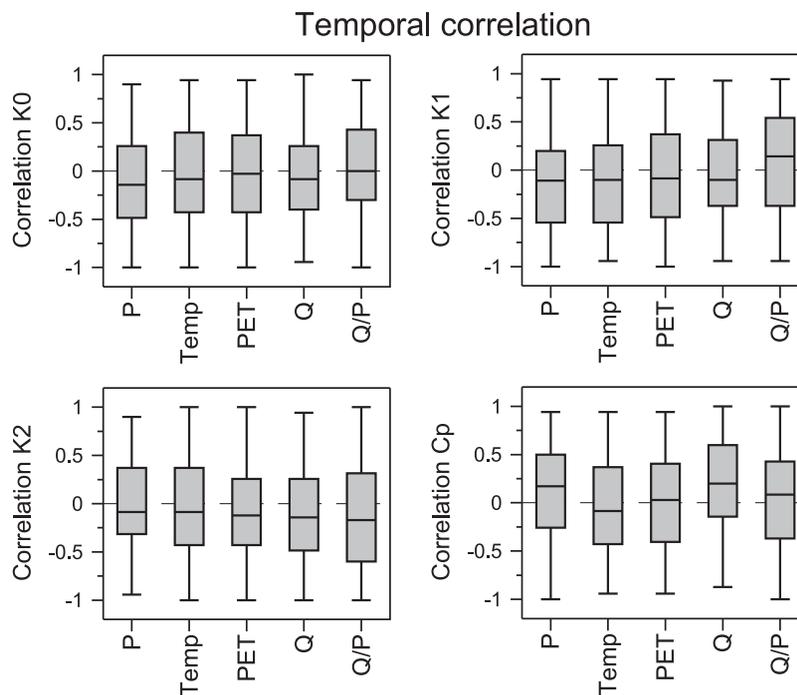


Figure 8. Spearman rank correlation coefficients (temporal correlations) between model parameters and climatic indicators for the six 5 year calibration periods, given separately for each catchment. The Box-Whisker plots show the spatial minimum, maximum, median, lower quartile, and upper quartile of the 273 Austrian catchments.

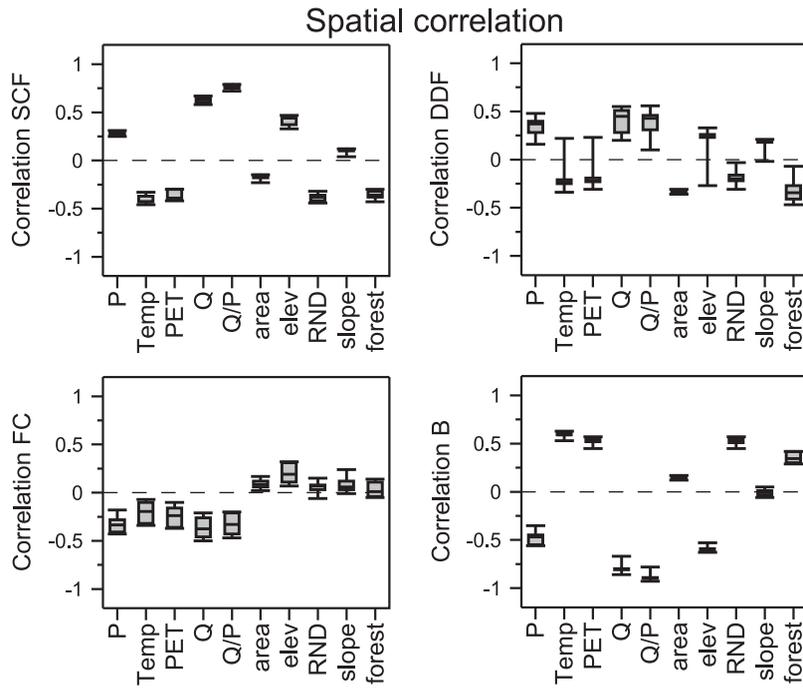


Figure 9. Spearman rank correlation coefficients (spatial correlations) between model parameters and climatic indicators (and static catchment descriptors) for the 273 Austrian catchments, given separately for each of the six 5 year calibration periods. The Box-Whisker plots show the temporal minimum, maximum, median, lower quartile, and upper quartile of the six 5 year calibration periods.

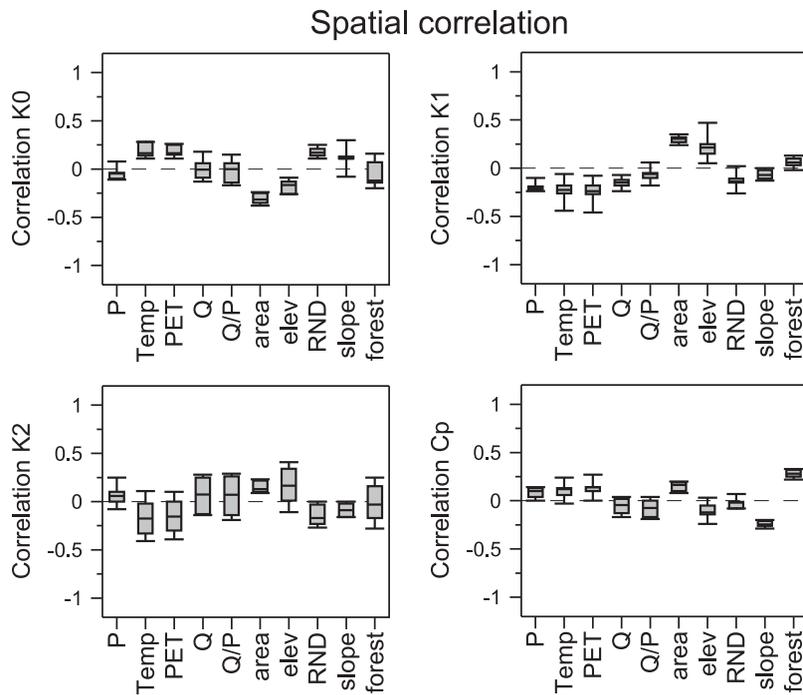


Figure 10. Spearman rank correlation coefficients (spatial correlations) between model parameters and climatic indicators (and static catchment descriptors) for the 273 Austrian catchments, given separately for each of the six 5 year calibration periods. The Box-Whisker plots show the temporal minimum, maximum, median, lower quartile, and upper quartile of the six 5 year calibration periods.

the high Alps to more than 10°C in the eastern flatlands, i.e., a difference of more than 10°C. Similarly, mean annual precipitation has increased, on average, by about 100 mm/yr, while the differences in space are more than 2500 mm/yr (400 mm/yr in the east and 3000 mm/yr in the west). The temporal variability of the spatial correlation coefficients is rather small, i.e., the spatial correlations do not change much between the six calibration periods. The main regional patterns of the catchment characteristics are therefore represented rather well by the parameters. The distinct regional patterns of the parameters in Austria are consistent with the results of *Merz and Blöschl* [2004], who compared the calibrated parameters of two 11 year calibration periods for the same catchments as here (their Figures 4–7). For example, for both calibration periods, B values were high in the eastern flatlands and low in the alpine environment of the west, which is also borne out in the regional calibration of *Parajka et al.* [2007b].

[32] The spatial correlations of SCF, DDF, and B with most climate-related variables (air temperature, PET, runoff, and runoff-rainfall ratio) are consistent with the corresponding temporal correlations. SCF and DDF are negatively correlated with temperature and PET and positively correlated with runoff and the runoff-rainfall ratio, both in the temporal (Figure 5) and spatial (Figure 9) correlations. Similarly, B is positively correlated with temperature and PET and negatively correlated with runoff and the runoff-rainfall ratio, both in the temporal (Figure 5) and spatial (Figure 9) correlations. This lends additional credence to the interpretations above. However, for some parameters and climate indicators the temporal and spatial correlations are inconsistent, but at least part of this inconsistency can be explained on hydrological grounds. For example, the temporal correlation of FC and air temperature is positive; that is, with increasing mean air temperature in the more recent years, the maximum storage capacity tends to increase. In warmer years, more soil moisture evaporates, the soils are drier, and more water can be stored, resulting in a higher maximum storage capacity. In the spatial correlation analysis, FC and air temperature are negatively correlated, meaning that FC tends to increase as one moves to colder catchments. This negative correlation can be explained by the regional pattern of FC [see, e.g., *Merz and Blöschl*, 2004, Figure 5]. In Austria, FC tends to be high in the warm and dry flatlands of the east and in the dry and cold inner-alpine catchments in the west. The lowest FC values are found in the wet catchments at the northern rim of the high Alps, where orographic effects enhance precipitation. The mean annual temperature of this region is between those of the high Alps and the flatlands of the east. As more inner-alpine catchments than warm flatland catchments are included in the spatial analysis, the correlation of FC and temperature is negative. A similar reasoning may explain the differences in the correlations of other parameters and other climate indicators.

[33] For the snow routine (SCF and DDF) and runoff generation (FC and B) parameters, the correlation coefficients with climate indicators are higher than those with the static catchment attributes such as RND and catchment area. This suggests that the variability in snow and runoff generation processes at the scale of Austria is more strongly controlled by climatic variability than by the differences in static catchment attributes. This is in line with the analyses of event run-

off coefficients in Austria by *Merz and Blöschl* [2009a]. They concluded that event runoff coefficients were most strongly correlated to indicators representing climate, such as mean annual precipitation, through controlling the seasonal soil moisture variability. Land use, soil types, and geology did not seem to exert a major control on the runoff coefficients with the data they had available and at the regional scale of the analysis. Clearly, if more detailed soils and geology data were available for smaller catchment scales, one would expect very strong controls.

[34] For the four parameters (K_0 , K_1 , K_2 , and C_P) of the routing component, the correlation coefficients to climate indicators are always small and on the same order or slightly lower than those of the static catchment attributes (Figure 10). This is consistent with the results of *Merz and Blöschl* [2004], who found the highest correlations of K_0 , K_1 , K_2 , and C_P with the mean slope, the river network density, the percentage of tertiary and quaternary geological units, and catchment area, respectively. For a similar model type and 11 catchments in Sweden, *Seibert* [1999] found the best correlation of the flow routing parameters with catchment area, lake percentage, and forest percentage. Interestingly, in Figure 10 there is a consistent correlation between the percolation rate C_P and percent forest cover, which would be expected because of more permeable soils in the forest than for other land uses.

4.3. Implications for Climate Impact Analysis

[35] The previous analyses have shown that some of the calibrated parameters consistently vary with time and can be related to climate fluctuations. Given these parameter changes, an obvious question now is whether they matter for hydrologic prediction. The effects of the parameter changes on three flow indicators are analyzed here: the Q_{95} low-flow quantile (the discharge exceeded 95% of the time), the Q_{50} median flow, and the Q_5 high-flow quantile that is exceeded 5% of the time. Q_{95} is widely used in Europe and is relevant for numerous problems in water resources management [e.g., *Smakhtin*, 2001; *Laaha and Blöschl*, 2007]. The median is a measure of the water balance. Q_5 was used instead of peak discharges as it can be more robustly estimated from the 5 year periods used here.

[36] The error of runoff predictions consists of two parts. The first stems from the imperfect fit of the model to the runoff data during the calibration period and is mainly related to data and model structure errors [*Di Baldassarre and Montanari*, 2009]. The second arises when moving from the calibration to the verification (or any other) period and is mainly related to less than optimum parameters. Verification performance therefore tends to be lower than the calibration performance [see, e.g., *Merz et al.* 2009]. Assuming that the model parameters remain constant with time is expected to increase the second part of the error if nonstationarities are present. Both errors are shown in Figure 11. The dotted black lines in Figure 11 show the average of the errors $\Delta Q_{i,i}$ in the calibration period (i.e., the first part of the error due to an imperfect calibration):

$$\Delta Q_{i,i} = \frac{Q_{\text{sim},i,i} - Q_{\text{obs},i}}{Q_{\text{obs},i}}, \quad (6)$$

where $Q_{\text{sim},i,i}$ is the simulated flow characteristic in period i using the parameters calibrated for the same period, and

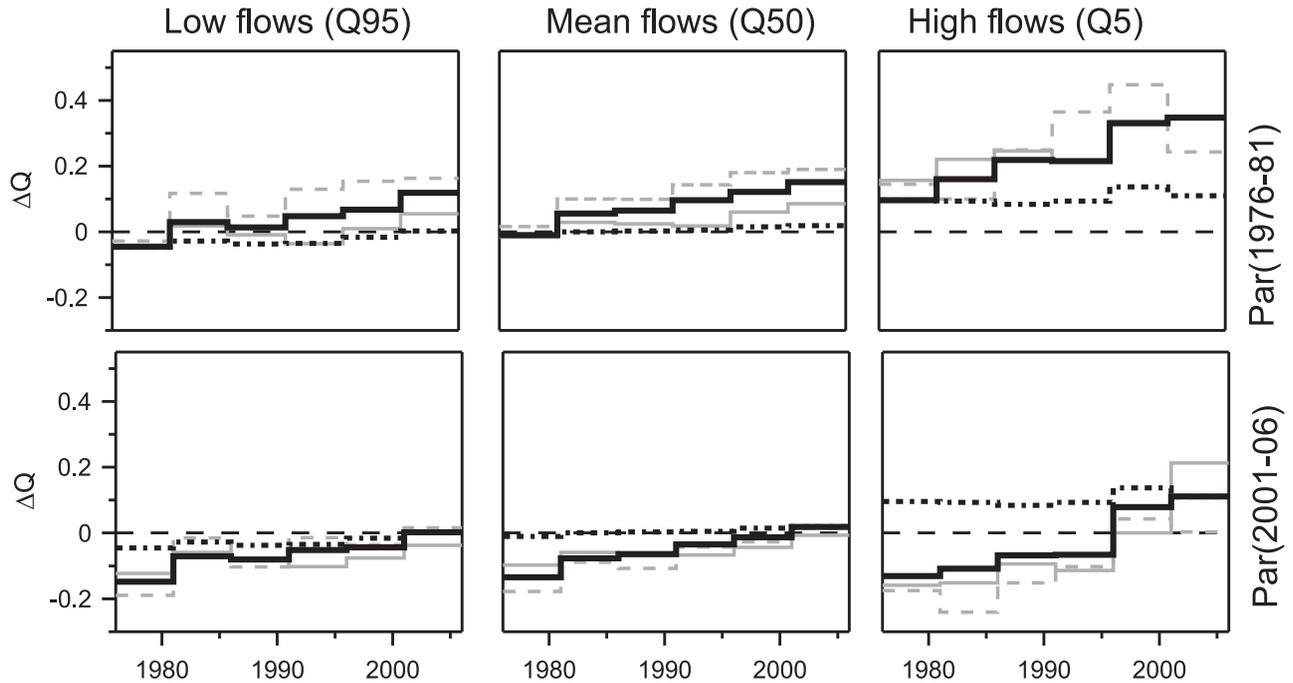


Figure 11. Relative validation errors of observed and simulated low (Q_{95}), median (Q_{50}), and high (Q_5) flow quantiles for different 5 year periods, averaged over the 273 Austrian catchments (thick black lines) when time stability of the parameters is assumed. The spatial averages of the wetter and drier catchments are shown as solid and dashed grey lines, respectively. The dotted line shows the relative error using the calibration parameter set from the period on the horizontal axis.

$Q_{obs,i}$ is the observed flow characteristic in period i . For example, the last segment of the dotted black line in Figure 11 (top right) shows 0.08, which is the relative difference of the simulated and observed Q_5 for the period 2001–2006 using the parameters calibrated for 2001–2006. The thick solid black lines in Figure 11 show the average of the errors $\Delta Q_{i,j}$ in verification period i (i.e., both parts):

$$\Delta Q_{i,j} = \frac{Q_{sim,i,j} - Q_{obs,i}}{Q_{obs,i}},$$

where $Q_{sim,i,j}$ is the simulated flow characteristic in period i using the parameters calibrated for a different period j . The horizontal axes of the panels relate to the periods i , while the different rows relate to different periods j . For example, the last segment of the thick solid black line in Figure 11 (top right) shows 0.30, which is the relative difference of the simulated and observed Q_5 for the period 2001–2006 using the parameters calibrated for 1976–1981. The spatial averages of the wetter and drier catchments are shown as solid and dashed grey lines, respectively, as in Figure 2.

[37] Figure 11 indicates that there are significant errors in the simulated flow quantiles if one assumes that the model parameters are time stable. When using the parameters calibrated to the period 1976–1981 for predicting the flows in 2001–2006, the Q_{95} low flows are overestimated by about 12%, Q_{50} is overestimated by about 15%, and the Q_5 high flows are overestimated by about 35% (Figure 11, top, thick solid black lines). The parameters from the period 1976–1981 represent colder periods with less evapotranspiration

and relatively higher runoff generation rates (lower B values; see Figure 4) and smaller soil moisture storages (lower FC values; see Figure 4). The model hence produces too much runoff when applying the parameters to the drier period 2001–2006. Similarly, using parameters calibrated to recent periods for simulating flow in the earlier periods tends to underestimate the flow characteristics (Figure 11, bottom). As expected, the differences between simulated and observed flows increase with increasing time lag between the period used to calibrate the parameters and the period for which flows are simulated.

[38] Interestingly, the differences between simulated and observed flows tend to be smaller for the Q_{95} low flows and the Q_{50} median flows than for the Q_5 high flows. High-flow situations may differ substantially between the different calibration periods. Apparently, the model only represents high flows well if it is calibrated to a period in which similar high-flow conditions were observed, thus reducing the model performance as one moves away from the calibration period. Mean flows and low flows are more stable in time, so mean and low-flow conditions of one period may be more representative of other periods [Laaha and Blöschl, 2005]. Low flows in the Austrian lowlands are caused by long periods of no or little rainfall in summer, while low flows in the Alps are caused by snow and freezing processes in winter. These processes are associated with longer time scales than flood processes, and aquifer storage additionally increases time scales [Skøien et al., 2003]. This explains the lower time dependence. Also, errors in the input rainfall data will be less important in periods of no or little rainfall.

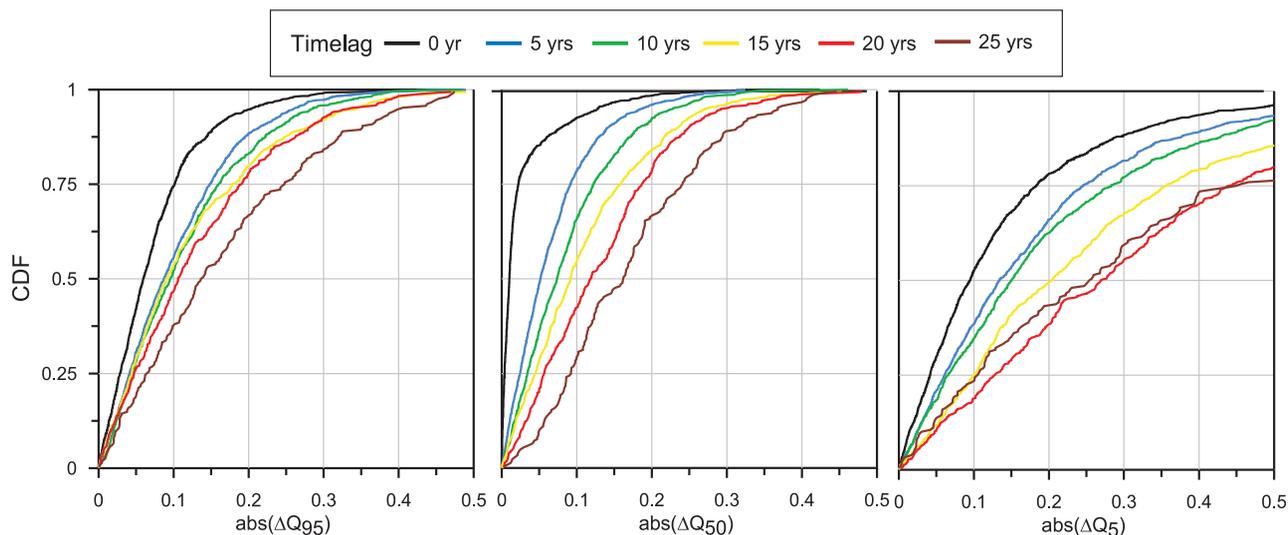


Figure 12. Cumulative distribution of the absolute validation errors of simulated low (Q_{95}), median (Q_{50}), and high flows (Q_5) for different 5 year periods, when time stability of the parameters is assumed. Curve parameter is the time lag between the calibration and the verification periods. A time lag of 0 corresponds to the calibration errors.

[39] To analyze the predictive errors due to the time lag between the calibration and prediction periods in more detail, the cumulative distribution function (CDF) of the absolute differences between simulated and observed flows have been plotted in Figure 12. The absolute errors of the Q_{50} median flows for time lags of 0, 5, 15, and 25 years are at least 1%, 5%, 9%, and 16%, respectively, for half the catchments (CDF = 0.5). The errors of low and high flows are larger as it is more difficult to represent the extremes well. As expected, the absolute errors are smallest for a time lag of zero, as this reflects the calibration case where the model performance is better than in the validation case. For all three flow indices, Q_{95} , Q_{50} , and Q_5 , the errors increase with increasing time lag. This means that the longer one extrapolates to the future (or the past), the more the model errors will increase. Only for the Q_5 high flows are the errors of 20 and 25 years similar. In the case of the low and high flows, the contribution of the calibration error (6% and 9%, respectively) to the total error is much larger

than in the case of the median flows (1%) (CDF = 0.5). This is partly related to the choice of the objective function that gives more weight to median flow than the high flows and partly to the time scales. However, in all instances the error due to the time trends is very important. The differences in the large errors are even larger, i.e., if one examines the errors exceeded in one quarter of the catchments (CDF = 0.75), as indicated in Table 2.

Table 2. Absolute Values of the Relative Validation Errors (%) of Simulated Low (Q_{95}), Median (Q_{50}), and High Flows (Q_5) for Different 5 Year Periods, When Time Stability of the Parameters Is Assumed^a

T (years)	CDF = 0.50			CDF = 0.75		
	Q_{95}	Q_{50}	Q_5	Q_{95}	Q_{50}	Q_5
0	6	1	9	10	2	18
5	9	5	13	15	9	24
10	10	7	15	16	12	28
15	10	9	20	18	16	35
20	11	12	27	19	19	44
25	14	16	25	25	24	47

^aT is the time lag between the calibration and the verification periods. A time lag of 0 corresponds to the calibration errors. Numbers given are the errors exceeded by half of the 273 catchments (CDF = 0.50) and a quarter of the catchments (CDF = 0.75). CDF, Cumulative distribution function.

5. Conclusions and Implications

[40] The main findings of this paper are as follows. The parameters of the snow and soil moisture accounting schemes of the model (the snow correction factor, the degree-day factor, the maximum soil moisture storage, and the nonlinearity parameter of runoff generation) show clear time trends if calibrated to different periods. The trends are similar for different subregions of the study area, which suggests they are real rather than calibration artifacts. The temporal changes in the calibrated parameters can be related to climate indicators such as the air temperature and potential evapotranspiration. The mean annual air temperatures in the catchments have increased, on average, by almost 2°C between 1976 and 2006, resulting in higher evapotranspiration and drier catchment conditions in the more recent years, which explains much of the time trends of the parameters. In most instances, the temporal correlations between the model parameters and the climate indicators for each catchment are consistent with spatial correlations between the model parameters and the climate indicators for each time period, lending additional credence to the results. Also, the relationships can be interpreted on hydrological grounds and are consistent with independent data on snow depths and runoff coefficients. The parameters related to the routing scheme show less clear time trends, and the correlations to the climate indices are weaker. To examine the effect of assuming that the model parameters do not change with time, simulations were

performed using parameters calibrated from periods that are different from the simulation periods. The simulation errors clearly increase with the time lag between the calibration period and the simulation period. They are largest for high-flow simulations followed by low flows and median flows.

[41] These results have important implications for hydrologic prediction in general and climate impact analyses in particular. The perhaps most obvious implication is that care needs to be taken when using calibrated parameters for predictions of the future. The predictive errors can be very large, and a large portion of the errors may be due to a transient climate. Using calibrated parameters for predictions of a transient future will increase model uncertainty beyond the uncertainties usually dealt with in simulation studies. These uncertainties relate to climate impact studies as discussed by *Blöschl and Montanari* [2010] but, in fact, any hydrological predictions will be affected by parameter errors due to a transient climate to a larger or lesser degree. Assume we were in the year 1981 and climate models predicted a 100 mm increase in annual precipitation and an almost 2°C increase in air temperature. Most hydrological models would probably predict an increase in runoff, similar to the model used here. We now know that such predictions were wrong, as the average observed runoff has not increased and the increase in precipitation has been offset by increases in actual evapotranspiration. The model estimates, obviously, depend on how evapotranspiration is exactly parameterized, but the model used here is a very common one in hydrology. Also, the model used here does predict increased evapotranspiration rates related to the increases in air temperature, but they are much smaller than what can be inferred from the runoff data. Runoff simulations may therefore be much less reliable than what is usually thought in impact studies [*Blöschl and Montanari*, 2010]. This applies even more so to low flows and floods.

[42] There are a number of potential options for dealing with this nonstationarity issue and to reduce the errors below those shown in Table 2. The first is to explicitly account for nonstationary model parameters. This paper shows strong evidence of correlations with climate variables, so one could exploit these correlations to predict the parameters to be used in the catchment model as part of impact studies. This option is not an elegant one as the usual philosophy of dynamic models is to use time-dependent boundary conditions (such as rainfall) and non-time-dependent model parameters to make the model directly applicable to predictions using future (or different) boundary conditions. Also, this may be complicated by complex correlations among the parameters and with various catchment characteristics [*Andréassian et al.*, 2003; *Wagener*, 2007]. An alternative would be to change the model structure. What is needed are models that if used in an analysis similar to the one in this paper, will give calibration parameters that are time stable. When using the parameters calibrated for the late 1970s to simulate runoff at the beginning of the twentieth century, runoff is overestimated (Figure 11). The main reason for this is that evapotranspiration at the beginning of the twentieth century was larger than in the late 1970s and only part of the difference can be explained by the evapotranspiration component of the model. Apparently, the vegetation transpired more

efficiently than would be expected from air temperature alone. This is likely related to a longer growing season. The model structure used here, however, assumes that the processes determining the length of the growing season are outside the model domain and therefore external to the model. This assumption ignores any feedbacks between the hydrological cycle and the processes determining the length of the growing season. Incorporating them into the model may help reduce the biases. However, it would not be expected that all the errors shown in Figure 12 can be avoided as they consist of both biases and random error.

[43] Some people may argue that relying less on calibration would help improve the model predictions. It is true that if one avoids calibration, one also avoids the issues discussed here, but then one does not take advantage of the bias reduction calibration can provide. The errors may perhaps not change with time and may be large all the time. This is reminiscent of the prediction in ungauged basins (PUB) problem where the issue is to predict runoff for catchments where no runoff data are available for calibration [*Sivapalan*, 2003]. In fact, this is very similar to the problem discussed here (see, e.g., the discussion of model testing by *Klemeš* [1986]), with the difference that PUB relates to extrapolation in space while here one is interested in the extrapolation in time. It is suggested that the term “prediction in ungauged climates” would be fitting for the transient prediction problem discussed in this paper.

[44] In the literature related to the PUB problem [e.g., *Sivapalan*, 2003, 2009], there has been an ongoing discussion of what the prospects of models that do not need calibration are, but as stated, in typical hydrological studies, calibration will help reduce bias [*Blöschl*, 2005; *Montanari and Toth*, 2007]. It will be interesting to develop models that are more robust to extrapolation in time than the current generation of conceptual models and still have comparable model biases. Whether these models will be more complex than the current generation of conceptual models remains to be seen. It seems clear, however, that a new generation of models will profit from identifying and testing them to hydrological data that go beyond runoff. It will likely be necessary to use soil moisture and snow data [*Blöschl and Kirnbauer*, 1992; *Parajka et al.*, 2006, 2007a] and other hydrological response data [*Fenicia et al.*, 2007, 2008; *Winsemius et al.*, 2009] to identify model structures that are able to reliably represent hydrological processes in a changing world.

[45] **Acknowledgments.** We would like to thank the Austrian Academy of Sciences (project: Predictability of runoff in a changing environment) and the Austrian Science Funds (FWF), project P18993-N10, for financial support. We would also like to thank the Austrian Hydrographic Services for providing the hydrographic data and Alberto Viglione for useful comments on the manuscript.

References

- Andréassian, V., E. Parent, and C. Michel (2003), A distribution-free test to detect gradual changes in watershed behavior, *Water Resour. Res.*, 39(9), 1252, doi:10.1029/2003WR002081.
- Auer, I., and R. Böhm (2003), Jahresmittel der Lufttemperatur, in *Hydrologischer Atlas Österreich*, Karte 1.6, Österreichischer Kunst und Kulturverlag, Vienna.
- Auer, I., et al. (2007), HISTALP: Historical instrumental climatological surface time series of the Greater Alpine Region, *Int. J. Climatol.*, 27, 17–46.

- Bergström, S. (1995), The HBV model, in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, pp. 443–476, Water Resour. Publ., Highlands Ranch, Colo.
- Bergström, S., B. Carlsson, M. Gardelin, G. Lindström, A. Pettersson, and M. Rummukainen (2001), Climate change impacts on runoff in Sweden: Assessments by global climate models, dynamical downscaling and hydrological modelling, *Clim. Res.*, 16, 101–112.
- Beven, K. J. (2000), Uniqueness of place and process representations in hydrological modelling, *Hydrol. Earth Syst. Sci.*, 4, 203–213.
- Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, 6, 279–298, doi:10.1002/hyp.3360060305.
- Blöschl, G. (2005), Rainfall-runoff modelling of ungauged catchments, in *Encyclopedia of Hydrological Sciences*, edited by M. G. Anderson, pp. 2061–2080, John Wiley, Chichester, New York.
- Blöschl, G., and R. Kirnbauer (1992), An analysis of snow cover patterns in a small Alpine catchment, *Hydrol. Processes*, 6, 99–109.
- Blöschl, G., and A. Montanari (2010), Climate change impacts—Throwing the dice?, *Hydrol. Processes*, 24, 374–381.
- Blöschl, G., S. Ardoïn-Bardin, M. Bonell, M. Dorninger, D. Goodrich, D. Gutknecht, D. Matamoros, B. Merz, P. Shand, and J. Szolgay (2007), At what scales do climate variability and land cover change impact on flooding and low flows?, *Hydrol. Processes*, 21, 1241–1247.
- Brown, A. E., L. Zhang, T. A. McMahon, A. W. Western, and R. A. Vertessy (2005), A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation, *J. Hydrol.*, 310, 28–61, doi:10.1016/j.jhydrol.2004.12.010.
- Budyko, M. I. (1974), *Climate and Life*, Academic, New York.
- Di Baldassarre, G., and A. Montanari (2009), Uncertainty in river discharge observations: A quantitative analysis, *Hydrol. Earth Syst. Sci.*, 13, 913–921.
- 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.
- Fenicia, F. F., H. H. G. Savenije, P. Matgen, and L. Pfister (2007), A comparison of alternative multiobjective calibration strategies for hydrological modelling, *Water Resour. Res.*, 43, W03434, doi:10.1029/2006WR005098.
- Fenicia, F., H. H. G. Savenije, and H. C. Winsemius (2008), Moving from model calibration towards process understanding, *Phys. Chem. Earth*, 33, 1057–1060.
- Gattermayr, W., M. Kuhn, and V. Weilguni (2003), Schneemessstellen und beobachtete Gletscher, in *Hydrologischer Atlas Österreich*, Karte 4.1, Österreichischer Kunst und Kulturverlag, Vienna.
- Jonas, M., B. Mayr, S. Schidler, M. Sotoudeh, and H. M. Knoflacher (1998), Land-use change and forestry in Austria, *Interim Rep. IR-98_028/May*, Int. Inst. for Appl. Syst. Anal., Laxenburg, Austria.
- Juston, J., J. Seibert, and P. O. Johansson (2009), Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment, *Hydrol. Processes*, 23, 3093–3109.
- Kay, A. L., H. N. Davies, V. A. Bell, and R. G. Jones (2009), Comparison of uncertainty sources for climate change impacts: Flood frequency in England, *Clim. Change*, 92, 41–63.
- Klemesš, V. (1986), Dilettantism in hydrology: Transition or destiny?, *Water Resour. Res.* 22(9), 177–188.
- Laaha, G., and G. Blöschl (2005), Low flow estimates from short stream flow records—A comparison of methods, *J. Hydrol.*, 306, 264–286.
- Laaha, G., and G. Blöschl (2007), A national low flow estimation procedure for Austria, *Hydrol. Sci. J.*, 52, 625–644.
- Lubès-Niel, H., J. E. Paturol, and E. Servat (2003), Study of parameter stability of a lumped hydrologic model in a context of climatic variability, *J. Hydrol.*, 278, 213–230.
- Merz, R., and G. Blöschl (2004), Regionalisation of catchment model parameters, *J. Hydrol.*, 287, 95–123.
- Merz, R., and G. Blöschl (2008a), Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, *Water Resour. Res.*, 44, W08432, doi:10.1029/2007WR006744.
- Merz, R., and G. Blöschl (2008b), Flood frequency hydrology: 2. Combining data evidence, *Water Resour. Res.*, 44, W08433, doi:10.1029/2007WR006745.
- Merz, R., and G. Blöschl (2009a), A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria, *Water Resour. Res.*, 45, W01405, doi:10.1029/2008WR007163.
- Merz, R. and G. Blöschl (2009b), Process controls on the statistical flood moments—A data based analysis, *Hydrol. Processes*, 23(5), 675–696.
- Merz, R., G. Blöschl, and J. Parajka (2006), Spatio-temporal variability of event runoff coefficients, *J. Hydrol.*, 331, 591–604, doi:10.1016/j.jhydrol.2006.06.008.
- Merz, R., G. Blöschl, and G. Humer (2008), National flood discharge mapping in Austria, *Nat. Hazards*, 46(1), 53–72, doi:10.1007/s11069-007-9181-7.
- Merz, R., J. Parajka, and G. Blöschl (2009), Scale effects in conceptual hydrological modeling, *Water Resour. Res.*, 45, W09405, doi:10.1029/2009WR007872.
- Mészáros, I., P. Miklánek, and J. Parajka (2002), Solar energy income modelling in mountainous areas, In *Interdisciplinary Approaches in Small Catchment Hydrology: Monitoring and Research—Proceedings of the 9th Conference of the European Network of Experimental and Representative Basins*, edited by L. Holko and P. Miklánek, pp. 127–135, U. N. Educ., Sci., and Cult. Organ., Paris.
- Montanari, A. (2005), Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, 41, W08406, doi:10.1029/2004WR003826.
- Montanari, A., and E. Toth (2007), Calibration of hydrological models in the spectral domain: An opportunity for ungauged basins?, *Water Resour. Res.*, 43, W05434, doi:10.1029/2006WR005184.
- Oudin, L., V. Andréassian, C. Perrin, C. Michel, and N. Le Moine (2008), Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, 44, W03413, doi:10.1029/2007WR006240.
- Parajka, J., and G. Blöschl (2008), The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models, *J. Hydrol.*, 358, 240–258.
- 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., V. Naeimi, G. Blöschl, W. Wagner, R. Merz, and K. Scipal (2006), Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale, *Hydrol. Earth Syst. Sci.*, 10, 353–368.
- Parajka, J., R. Merz, and G. Blöschl (2007a), Uncertainty and multiple objective calibration in regional water balance modelling, *Hydrol. Processes*, 21, 435–446, doi:10.1002/hyp.6253.
- Parajka, J., G. Blöschl, and R. Merz (2007b), Regional calibration of catchment models: Potential for ungauged catchments, *Water Resour. Res.*, 43, W06406, doi:10.1029/2006WR005271.
- Pebesma, E. J. (2001), Gstat user's manual, Dep. of Phys. Geogr., Utrecht Univ., Utrecht, Netherlands.
- Perrin, C., C. Michel, and V. Andréassian (2001), Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, *J. Hydrol.*, 242, 275–301.
- Perrin, C., V. Andréassian, C. Rojas Serna, T. Mathevet, and N. Le Moine (2008), Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments, *Water Resour. Res.*, 44, W08447, doi:10.1029/2007WR006579.
- Seibert, J. (1997), Estimation of parameter uncertainty in the HBV model, *Nord. Hydrol.*, 28, 247–262.
- Seibert, J. (1999), Regionalisation of parameters for a conceptual rainfall-runoff model, *Agric. For. Meteorol.*, 98–99, 279–293.
- Sivapalan, M. (2003), Prediction of ungauged basins: A grand challenge for theoretical hydrology, *Hydrol. Processes*, 17, 3163–3170.
- Sivapalan, M. (2009), The secret to “doing better hydrological science”: Change the question!, *Hydrol. Processes*, 23, 1391–1396, doi:10.1002/hyp.7242.
- Sköien, J. O., G. Blöschl, and A. W. Western (2003), Characteristic space scales and timescales in hydrology, *Water Resour. Res.*, 39(10), 1304, doi:10.1029/2002WR001736.
- Smakhtin, V. U. (2001), Low flow hydrology: A review, *J. Hydrol.*, 240, 147–186.
- Wagner, T. (2007), Can we model the hydrological impacts of environmental change?, *Hydrol. Processes*, 21, 3233–3236, doi:10.1002/hyp.6873.
- Wagner, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modeling: Dynamic identifiability analysis, *Hydrol. Processes*, 17, 455–476.
- Wagner, T., M. Sivapalan, P. A. Troch, B. L. McGlynn, C. J. Harman, H. V. Gupta, P. S. C. Rao, N. B. Basu, and J. S. Wilson

- (2010), The future of hydrology: An evolving science for a changing world, *Water Resour. Res.*, *46*, W05301, doi:10.1029/2009WR008906.
- Weilguni, V. (2003), Niederschlagsmessstellen, in *Hydrologischer Atlas Österreich*, Karte 2.1, Österreichischer Kunst und Kulturverlag, Vienna.
- Wilby, R. L. (2005), Uncertainty in water resource model parameters used for climate change impact assessment, *Hydrol. Processes*, *19*, 3201–32219, doi:10.1002/hyp.5819.
- Wilby, R. L., and I. Harris (2006), A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, *Water Resour. Res.*, *42*, W02419, doi:10.1029/2005WR004065.
- Winsemius, H. C., B. Schaefli, A. Montanari, and H. H. G. Savenije (2009), On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, *Water Resour. Res.*, *45*, W12422, doi:10.1029/2009WR007706.

G. Blöschl and J. Parajka, Institute for Hydraulic and Water Resources Engineering, Vienna University of Technology, Karlsplatz 13/223, A-1040 Vienna, Austria. (merz@hydro.tuwien.ac.at)

R. Merz, Department for Catchment Hydrology, Helmholtz Centre for Environmental Research (UFZ), Theodor-Lieser-Straße 4, D-06120 Halle, Germany. (ralf.merz@ufz.de)