



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)



Journal of Hydrology 287 (2004) 95–123

Journal  
of  
**Hydrology**

[www.elsevier.com/locate/jhydrol](http://www.elsevier.com/locate/jhydrol)

## Regionalisation of catchment model parameters

Ralf Merz\*, Günter Blöschl

*Institut für Hydraulik, Gewässerkunde und Wasserwirtschaft, Technische Universität Wien,  
Karlsplatz 13/223, A-1040 Wien, Austria*

Received 10 December 2002; revised 22 September 2003; accepted 26 September 2003

### Abstract

We simulate the water balance dynamics of 308 catchments in Austria using a lumped conceptual model involving 11 calibration parameters. We calibrate and verify the model for two non-overlapping 11-year periods of daily runoff data. A comparison of the calibrated parameter values of the two periods suggests that all parameters are associated with some uncertainty although the degree of uncertainty differs between the parameters. The regional patterns of the calibrated parameters can be interpreted based on hydrological process reasoning indicating that they are able to represent the regional or large-scale differences in the hydrological conditions. Catchment attributes explain some of the spatial parameter variability with coefficients of determination of up to  $R^2 = 0.27$ , but usually the  $R^2$  values are lower. Parameter uncertainty does not seem to cloud the relationship between calibrated parameters and catchment attributes to a significant extent as suggested by an optimised correlation analysis. The median Nash–Sutcliffe efficiencies of simulating streamflow decrease from 0.67 to 0.63 when moving from the calibration to the verification period. This is a small decrease, which suggests that problems with over-parameterisation of the model are unlikely. We then compare regionalisation methods for estimating the model parameters in ungauged catchments, in terms of the model performance. The best regionalisation methods are the use of the average parameters of immediate upstream and downstream (nested) neighbours and regionalisation by kriging. For the calibration period, the average decrease in the Nash–Sutcliffe model efficiency, as a result of the regionalisation, is 0.10 which is about twice the decrease of moving from the calibration to the verification period. The methods based on multiple regressions with catchment attributes perform significantly poorer. Apparently, spatial proximity is a better surrogate of unknown controls on runoff dynamics than catchment attributes.

© 2004 Elsevier B.V. All rights reserved.

**Keywords:** Parameter uncertainty; Patterns of model parameters; Model calibration; Catchment attributes; Regionalisation; Ungauged catchments

### 1. Introduction

Simulations of the water balance dynamics of catchments are needed for addressing a number of

engineering and environmental problems such as assessing anthropogenic effects on water quantity and quality, estimating design values and streamflow forecasting. Conceptual water balance models are widely used in hydrology because the required input data are usually readily available and the models are relatively simple and easy to use. The model

\* Corresponding author. Fax: +43-1-588-1-233-99.  
E-mail address: [merz@hydro.tuwien.ac.at](mailto:merz@hydro.tuwien.ac.at) (R. Merz).

parameters are effective values on the catchment scale and so cannot be measured in the field. Because of this, the model parameters are always calibrated against observed streamflow data, if possible (Klemeš, 1986). For catchments without streamflow observations, parameters have to be estimated from other sources of information, such as neighbouring catchments, or taken from tabulated values from the literature, or assumed based on expert judgement. Because of a lack of calibration, catchment models usually perform significantly poorer in ungauged catchments than they do in gauged catchments, but the ungauged catchment case is important both from practical and theoretical perspectives.

The process of transferring parameters from neighbouring catchments to the catchment of interest is generally referred to as regionalisation (Blöschl and Sivapalan, 1995). The choice of catchments from which information is to be transferred is usually based on some sort of similarity measure, i.e. one tends to choose those catchments that are most similar to the site of interest. One common similarity measure is spatial proximity, based on the rationale that catchments that are close to each other will have a similar runoff regime as climate and catchment conditions will only vary smoothly in space. An example of this type of approach is given by Vandewiele and Elias (1995), who derived the parameters of a monthly water balance model for 75 catchments in Belgium from neighbouring catchments. For a case where they regionalised parameters using kriging, their model performed well for 72% of the catchments while it was only 44% when transferring parameters from the nearest catchment.

An alternative similarity measure is the use of catchment attributes such as land use, soil type and topographic characteristics. In principle, one would assume that the model parameters are closely related to catchment attributes, as the model parameters are designed to represent the functional behaviour of catchment response which, in turn, should be controlled by physical characteristics of catchments such as land use. However, most of the case studies on relating model parameters and catchment attributes published in the literature have found rather low correlations. In a comparative study of 331 catchments in Australia, Peel et al. (2000), for example, found the groundwater recharge parameter of

the SYMHID model to be significantly related to a climate index (coefficient of determination  $R^2 = 0.20$ ). They also found significant correlations for a soil moisture storage parameter, both with a climate index and a relief index ( $R^2 = 0.25$  and  $0.21$ ). For the other parameters and the other catchment attributes, the correlations were lower. Sefton and Howarth (1998) compared calibrated parameters of the IHACRES model with attributes of 60 catchments in England and Wales. The best correlations they obtained were  $R^2 = 0.59$  between a routing parameter and percentage of aquifers, and  $R^2 = 0.69$  between an evaporation parameter and mean annual precipitation. For the storage parameters, no significant correlations were obtained. Seibert (1999) related the model parameters of the HBV model to attributes of 11 Swedish catchments within the NOPEX area. Some relationships between lake percentage and soil parameters found by Seibert (1999) called the process basis of their model into question as they could not be explained by hydrologic reasoning. In contrast, relationships between forest percentage and snow parameters supported the process basis of their model. They found the best correlations between a non-linearity parameter of runoff generation and catchment area with a Spearman rank correlation coefficient of  $R^2 = 0.87$ , but most parameters exhibited hardly significant correlations with catchment attributes.

These typically low correlations are likely to translate into rather low model performances for the ungauged catchment case as indicated by a number of regionalisation studies. Seibert (1999) for the 11 Swedish catchments found a decrease of the median Nash–Sutcliffe model efficiency of 0.81–0.79 when moving from calibrated parameters to regionalised parameters for the same set of catchments, however, the median efficiencies decreased to 0.67 for a separate set of seven catchments. A recent example of a regional application of a conceptual model has been presented by Beldring et al. (2002). They used 141 catchments in Norway for calibrating a version of the HBV model. They then treated 43 additional catchments as ungauged and regionalised the model parameters as a function of land use classes. For both sets of catchments, they found median Nash–Sutcliffe efficiencies of 0.68 and concluded that the regionalisation method represented the main features of

the landscape well. However, for 20% of the second set of stations, the efficiencies were less than 0.3.

There are two alternative explanations of the relatively poor correlations between model parameters and catchment attributes and hence relatively poor performances of catchment models for the ungauged case. One explanation is that the catchment attributes used may not be very relevant for catchment response. For soil type, this is certainly the case as reflected by the usually poor predictive power of pedotransfer functions (Grayson and Blöschl, 2000). The other 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). An analysis of parameter uncertainty should therefore be an integral part of any regionalisation study of catchment models. Parameter uncertainty may result from model over-parameterisation and from data errors (Bergström, 1991; Post and Jakeman, 1999). Most of the analyses of parameter uncertainty in the literature are based on Monte Carlo simulations for the same catchment (Beven and Binley, 1992). Uhlenbrook et al. (1999), for example, analysed the parameter uncertainty of the HBV model for a small mountainous catchment using Monte Carlo simulations. They found some of the parameters such as the maximum soil moisture storage and the lower zone storage coefficient to be poorly defined while other parameters such as the degree day factor (*DDF*) were much better constrained. A similar study was performed by Seibert (1997) for a number of Swedish catchments, but the uncertain parameters were not the same as those in the study of Uhlenbrook et al. (1999). This suggests that parameter uncertainty significantly depends on the catchments studied and data aspects in addition to the model structure. An alternative to Monte Carlo studies is calibrating the model on different subperiods and comparing the calibrated parameters for the respective subperiods. This is in fact a more stringent test of parameter robustness than Monte Carlo analyses as it tests both the identifiability of parameters and the stationarity of the data and their quality. The difference of the parameters of the two subperiods is a measure of the sum of the uncertainties due to poor parameter identifiability and due to data problems. If the calibrated model parameters for the subperiods are

similar, then the uncertainty can be assumed to be small. However, relatively long data series are needed for this type of test to be meaningful.

The aim of this paper is to assess the potential of regionalising the parameters of a conceptual daily water balance model for the ungauged catchment case. We use hydrologic data from 308 catchments over a period of 23 years, which will likely allow us to draw more generic inferences on regionalising catchment model parameter than has been possible in most previous studies. Among other things we are able to address the parameter uncertainty issue through a comparison of calibrated parameters for two subperiods. Specifically, we address the following research questions: (a) what are the spatial patterns of calibrated model parameters and can they be interpreted based on process reasoning; (b) how well are they related to catchment attributes; and (c) what is the model performance for the case of ungauged catchments using different regionalisation procedures? We use the same model structure for all catchments. For a regional study as the one presented in this paper, it may not be feasible to compare different model structures, as one would perhaps do if one focused on a single catchment. Also, using different model structures in different catchments would render a regional comparison of model parameters difficult, if not impossible. In the analyses of calibration parameters and the regionalisation comparisons, we then focus only on those catchments with acceptable model performance.

In the next chapter, we present the data, followed by a description of the model. We then analyse the parameter uncertainty and address each of the three research questions in sequence.

## 2. Data

The study region is Austria which is hydrologically quite diverse, ranging from lowlands in the east to high alpine catchments in the west (Fig. 1). Elevations range from less than 200 m a.s.l. to more than 3000 m a.s.l. 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, forested in the medium elevation ranges, while alpine

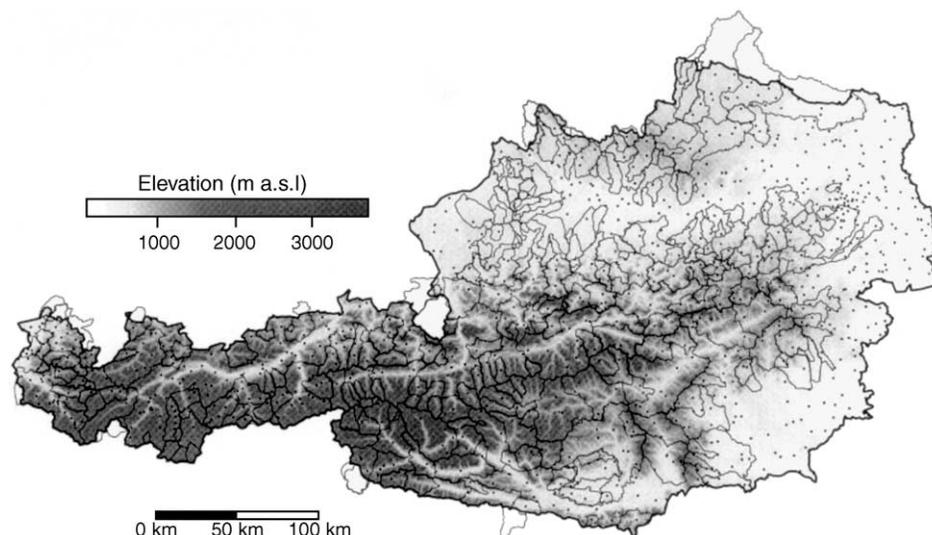


Fig. 1. Topography (m a.s.l.) of Austria and boundaries of the gauged catchments used in this paper. The dots show the rain gauge locations.

vegetation and rocks prevail in the highest catchments.

The study period is 1976–1997. The model input data are daily values of precipitation, air temperature and potential evaporation. Precipitation data from 1029 stations, and air temperature data and potential evaporation from 212 stations have been used. Potential evaporation has been estimated from daily temperature and potential sunshine duration by a modified Blaney–Criddle equation (DVWK, 1996). The estimates were compared to estimates by the Penman equation for a subset of the stations, where radiation data were available. This comparison indicated small biases of the Blaney–Criddle equation for the study area. The daily values of precipitation, air temperature and potential evapotranspiration were spatially interpolated by external drift kriging (Deutsch and Journel, 1997) using elevation as additional information. Fig. 1 indicates that, with a few exceptions, all catchments contain at least one precipitation station, so one would not expect large interpolation errors. These spatial fields were then superimposed on the catchment boundaries to derive catchment average values for each day. Two data sets of catchment boundaries were used (Fig. 1). Most of the boundaries were derived from a digital database digitised from the Austrian 1:50000 scale

map (ÖK 50). The remaining boundaries were derived from a digital elevation model. All catchment boundaries were checked manually using the ÖK 50 map. Catchment centroids were derived from the digital catchment boundaries to measure distance between catchments.

To calibrate and verify a catchment model, daily runoff data were used. In a first step, we carefully screened the data for errors and, in a second step, we removed all stations with significant anthropogenic effects from the data set (Piöck-Ellena and Blöschl, 1998). Anthropogenic effects were assessed in terms of the presence of significant reservoirs in the catchment (ratio of volume and catchment area larger than 0.2 m) and the presence of significant water transfers (effective catchment area larger than 150% or less than 50% of the topographic catchment area). In a third step, we performed some initial analyses to examine whether it was possible to close the long term water balance for the remaining catchments. In a number of catchments, this was indeed not possible because of significant subsurface flows across the topographic catchment boundaries due to karstic conditions and porous aquifers. Those catchments were not further used in this paper. After the screening procedures, a set of 459 gauged catchments with reliable runoff data remained. The areas of these

catchments range from 3 to 5000 km<sup>2</sup> with a median of 162 km<sup>2</sup>.

A number of catchment attributes were used. Average catchment elevation and average topographic slope were derived from the digital elevation model. Mean annual precipitation and mean maximum annual daily precipitation (i.e. the long term mean of the series) were spatially interpolated using observed precipitation from the raingauges in Fig. 1. The record lengths ranged between 45 and 97 years. Catchment average values were then found by integration within each catchment boundary. River network density was calculated from the digital river network map at the 1:50000 scale for each catchment. The boundaries of porous aquifers were taken from the Hydrographic Yearbook (HZB, 2000), and by combining them with the catchment boundaries, the areal portion of porous aquifers in each catchment was estimated. The FARL (flood attenuation by reservoirs and lakes) lake index was calculated according to IH (1999, pp. 5/19-27). Digital maps of land use (Ecker et al., 1995), regional soil types (based on the FAO map, see ÖBG, 2001) and the main geological formations (Geologische Bundesanstalt, 1998) were also used. These digital maps were combined with the catchment boundaries to derive areal portions of each land use type, soil type, and geological unit.

### 3. Model structure and model calibration

The model used in this paper is a lumped conceptual rainfall-runoff model, following the structure of the HBV model (Bergström, 1976). The model runs on a daily time step and consists of a snow routine, a soil moisture routine and a routing routine. The snow routine represents snow accumulation and melt by a simple degree day concept involving *DDF*. Catch deficit of the precipitation gauges during snowfall is corrected by a snow correction factor, *SCF*. The soil moisture routine represents runoff generation and changes in the soil moisture state of the catchment and involves three parameters, the maximum soil moisture storage, *FC*, a parameter representing the soil moisture state above which evaporation is at its potential rate, termed the limit for potential evaporation, *LP*, and a parameter in the non-linear function relating runoff generation to the soil

moisture state, termed the non-linearity parameter, *beta*. The response function represents runoff routing on the hillslopes, and consists of 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 with a fast storage coefficient of *k*<sub>1</sub>, percolation to the lower zone with a constant percolation rate *c*<sub>perc</sub>, and, if a threshold *LS*<sub>uz</sub> of the storage state is exceeded, through an additional outlet with a storage coefficient of *k*<sub>0</sub>. Water leaves the lower zone with a slow storage coefficient of *k*<sub>2</sub>. The outflow from both reservoirs is then routed by a triangular transfer function representing runoff routing in the streams, where *c*<sub>route</sub> is a free parameter. This model involves a total of 11 calibration parameters. A more detailed description of the model is given in Appendix A.

We calibrated the model parameters to observed runoff making use of an automated procedure that involves an objective function consisting of five terms. The first term involves the Nash and Sutcliffe (1970) coefficient of efficiency, *ME*, of the match of simulated and observed daily runoff (Eq. (B1)). The second term involves the volume error of runoff, *VE* (Eq. (B2)). The third and fourth terms are penalty functions to avoid snow and moisture to accumulate without bounds over the years (Eq. (B3)). The fifth term is a penalty function that allows to include an informed guess about the a priori distribution of each parameter (Table A1). The weights associated with the five terms were determined by test computations. More details on the objective function are given in Appendix B. We optimised the objective function using the shuffled complex evolution (SCE-UA) scheme (Duan et al., 1992).

In test simulations, not shown here, we used the model efficiency as the sole objective function. These simulations resulted in higher calibration efficiencies than those from the compound objective function but the verification efficiencies were lower. This indicates that the compound objective function used here results in a more robust parameter estimation. As the main focus of this paper was on the estimation and use of model parameters rather than on optimising at-site streamflow simulations, the use of a compound objective function is preferable.

The period from 1976 to 1997 was split into two 11-year periods. In a first step, the parameters were

calibrated to the period from January 1, 1987 to December 31, 1997 and verified for the period from January 1, 1976 to December 31, 1986. In a second step, the two periods were swapped, i.e. the model was calibrated to the period from 1976 to 1986 and verified for the period from 1987 to 1997. One year prior to the beginning of each period was used as a spin up period. Catchments for which less than 1825 days (= 5 years) of observed runoff data were available in any of the periods were not used in the further analysis. For some of the catchments,

the calibration efficiency was so poor that we concluded there may still be data problems and/or problems with the model structure. Catchments with calibration efficiencies  $ME < 0.5$  or volume errors  $|VE| > 0.25$  were, therefore, not used in the further analysis. The remaining number of catchments was 308. These were used for all analyses in this paper.

We judged the model performance by a split sample test in the terminology of Klemeš (1986). We compared simulated and observed runoff in terms of model efficiencies  $ME$  (Eq. (B1)) and volume errors

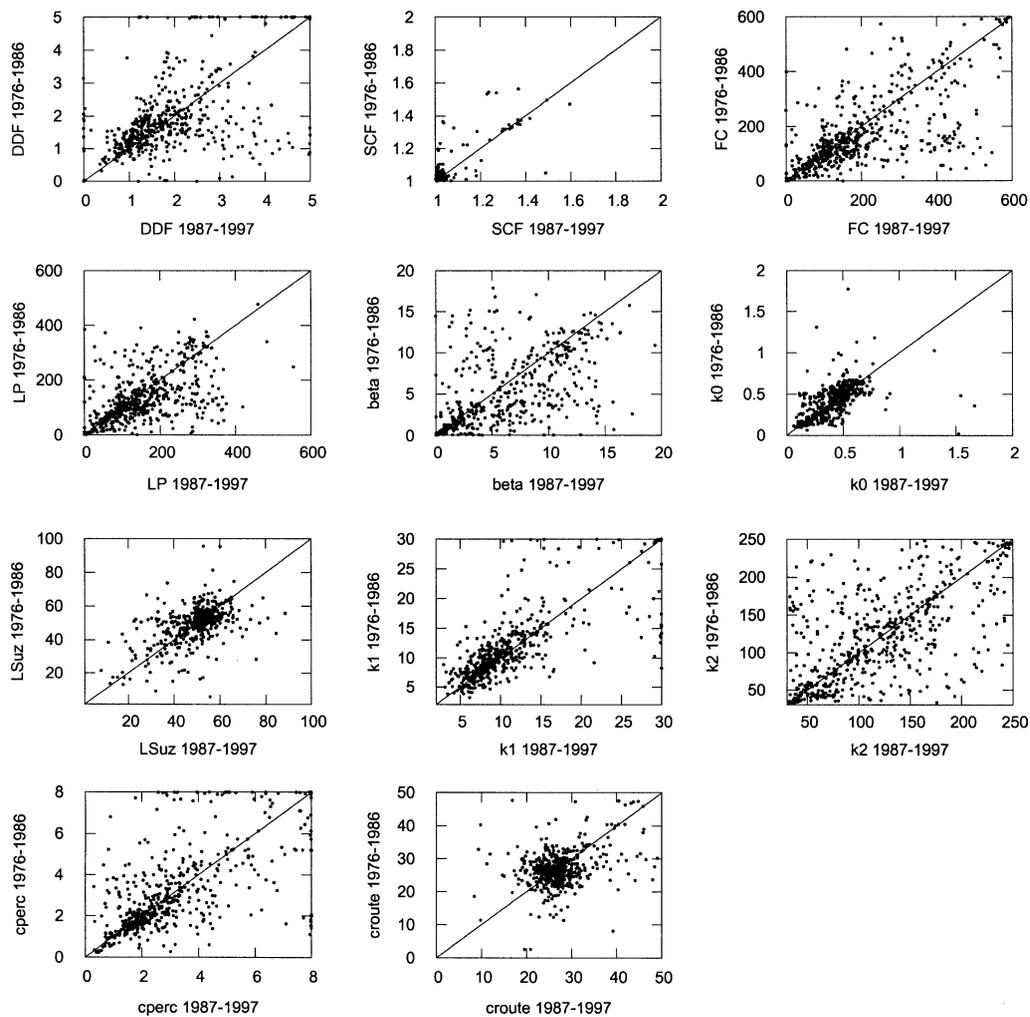


Fig. 2. Assessment of parameter uncertainty. Model parameters calibrated on the period 1976–1986 plotted against those calibrated on the period 1986–1997.

VE (Eq. (B2)) for the calibration period as well as for the verification period, which was not used for calibration.

#### 4. Parameter uncertainty

We judged the reliability of the model parameters by comparing the parameters calibrated for the 1987–1997 period with those calibrated for the 1976–1986 period. The calibrated parameters for the two periods have been plotted against each other in Fig. 2. If the parameters for the two periods are similar, i.e. cluster around the 1:1 lines, then the uncertainty can be assumed to be small while a large scatter indicates large uncertainties. The correlation coefficients and the fraction of catchments for which the differences between the two calibrated parameter sets is smaller than 5, 10 and 50% of the possible parameter range are given in Table 1. The parameters show significant differences in their uncertainty. When judging the uncertainty jointly by the figures in Table 1 and the visual appearance in Fig. 2, the most uncertain parameters are the slow storage coefficient,  $k_2$  and the routing parameter,  $c_{\text{route}}$ . The parameters with the smallest uncertainties are the fast storage coefficient,  $k_1$  and the threshold storage coefficient,  $k_0$ . While *SCF* gives large correlation coefficients, they are due to *SCF* values close to one in most catchments with

a small number of outliers. We will be able to attribute more credibility to relationships between the latter parameters and catchment attributes than to relationships involving the former parameters. The non-linearity parameter, beta, exhibits a more complex pattern. There is little scatter for small beta values but significant scatter for large beta values. This implies that for catchments that behave linearly (small beta values), beta can be better identified than for catchments that behave more non-linearly.

It is interesting to compare these results with those of other authors that have examined a similar model. In Uhlenbrook et al. (1999), the most uncertain parameters were *FC* and  $k_2$ , while the most certain parameters were *DDF* and *LP*. In Seibert (1997), the uncertain and certain parameters were  $c_{\text{perc}}$  and *LP*, and *DDF*,  $k_1$  and *SCF*, respectively. It seems that the relative parameter uncertainty significantly depends on the catchments studied.

To examine whether the model is over-parameterised, we analysed the interdependence of the calibrated model parameters. If obvious interdependences were present, we would have to reconsider the model structure with a view to reduce the number of calibration parameters. In Fig. 3, we plotted all calibrated parameters for the two periods against each other. In the lower left half and the upper right half of the matrix, the parameters calibrated on the 1987–1997 and 1976–1986 periods, respectively, are shown. The ranges of the axes are the possible parameter ranges as of Table A1. Overall, the interdependences are weak, if at all present. The exception is the relationship between the maximum soil moisture storage, *FC*, and the limit for potential evaporation, *LP*. *LP* is always equal or smaller than *FC*, as defined in the model, so one would expect the kind of dependence shown in Fig. 3. The interdependences of other parameters are much weaker. Those that can be discerned can be interpreted on hydrological grounds or interpreted based on the model structure. There is a tendency for  $k_0$  and  $k_1$  to be negatively related. These two parameters are time constants and, although the permissible parameter ranges of the three parameters are quite different (Table A1), one can take over the role of the other to some extent, so one would expect a negative correlation. The  $k_0$  and  $LS_{\text{uz}}$  values exhibit a weak negative dependence. As  $LS_{\text{uz}}$  is the threshold beyond

Table 1  
Parameter uncertainty: coefficient of determination,  $R^2$ , as a measure of how similar the model parameters calibrated for the 1987–1997 and 1976–1986 periods are, and fraction of catchments exhibiting differences  $\Delta P$  in calibrated parameters for the two periods less than 5, 10 and 50% of the possible parameter range

Parameter	$R^2$	$\Delta P < 5\%$	$\Delta P < 10\%$	$\Delta P < 50\%$
DDF (mm/(day °C))	0.45	0.36	0.58	0.90
SCF (–)	0.63	0.92	0.93	0.95
FC (mm)	0.41	0.41	0.59	0.92
LP (mm)	0.41	0.43	0.62	0.94
beta (–)	0.52	0.44	0.59	0.91
$k_0$ (days)	0.50	0.63	0.83	0.94
$LS_{\text{uz}}$ (mm)	0.42	0.53	0.74	0.95
$k_1$ (days)	0.64	0.48	0.71	0.93
$k_2$ (days)	0.35	0.33	0.50	0.89
$c_{\text{perc}}$ (mm/day)	0.51	0.45	0.60	0.91
$c_{\text{route}}$ (days <sup>2</sup> /mm)	0.09	0.38	0.64	0.94

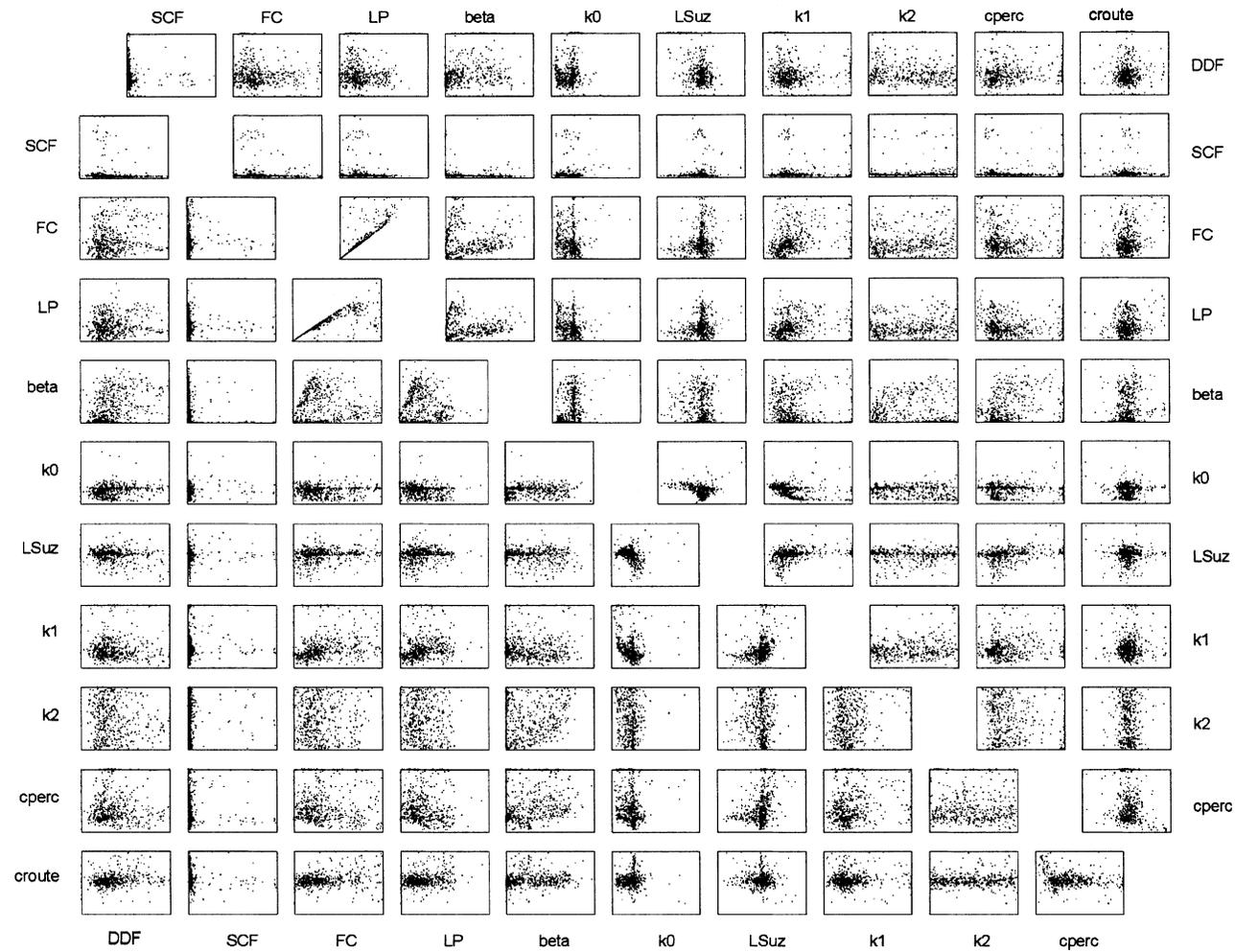


Fig. 3. Matrix of calibrated model parameters. Lower left half: calibration on period 1987–1997. Upper right half: calibration on period 1976–1986. Parameter ranges on the axes are as of Table A1. For the SCF, the range is 1.0–1.5.

which the  $k_0$  reservoir becomes operative, this type of dependence would be expected. The  $k_1$  and  $LS_{uz}$  values exhibit a weak positive dependence which is likely related to the dependences of  $k_0$  and  $LS_{uz}$ , and  $k_0$  and  $k_1$  discussed above. It is possible that there exist more complex relationships between three or more parameters, or regionally different relationships which do not appear in the global scatter plots but, given that the simple dependences are very weak, we do not expect that the more complex relationships are significant. We, therefore, believe that the number of model parameters cannot be reduced easily by, say, introducing functional relationships between the parameters. An additional assessment of the potential for over-parameterisation is given later in the paper in the context of comparing calibration and verification efficiencies of the model.

### 5. Spatial patterns of model parameters

As the parameters of the catchment model are designed to represent the peculiarities of the runoff

dynamics of each catchment, there should exist spatial patterns of the parameters that are co-located with the physiographic patterns in the study region. The similarities of the parameter patterns of the two calibration periods are an indication of the parameter reliability and identifiability. Figs. 4–7 show the spatial patterns of the calibrated model parameter for the two calibration periods.

In Fig. 4, the parameters of the snow module,  $DDF$ , and  $SCF$ , for the 1987–1997 calibration period are shown left, the parameters for the 1976–1986 calibration period are shown on the right hand side. For the 1987–1997 period, the  $DDF$  values are large in the prealpine regions in the north of the country as well as in the hilly regions in the southeast of the country. In early winter and spring, rainfall on an existing snow pack is an important contribution to runoff from these catchments. During these runoff situations, air humidity is usually high and cloud covers prevail which may induce large latent heat fluxes and large long wave radiation fluxes into the snow pack, hence the  $DDF$  for these catchments is quite high. Values of the  $DDF$  of

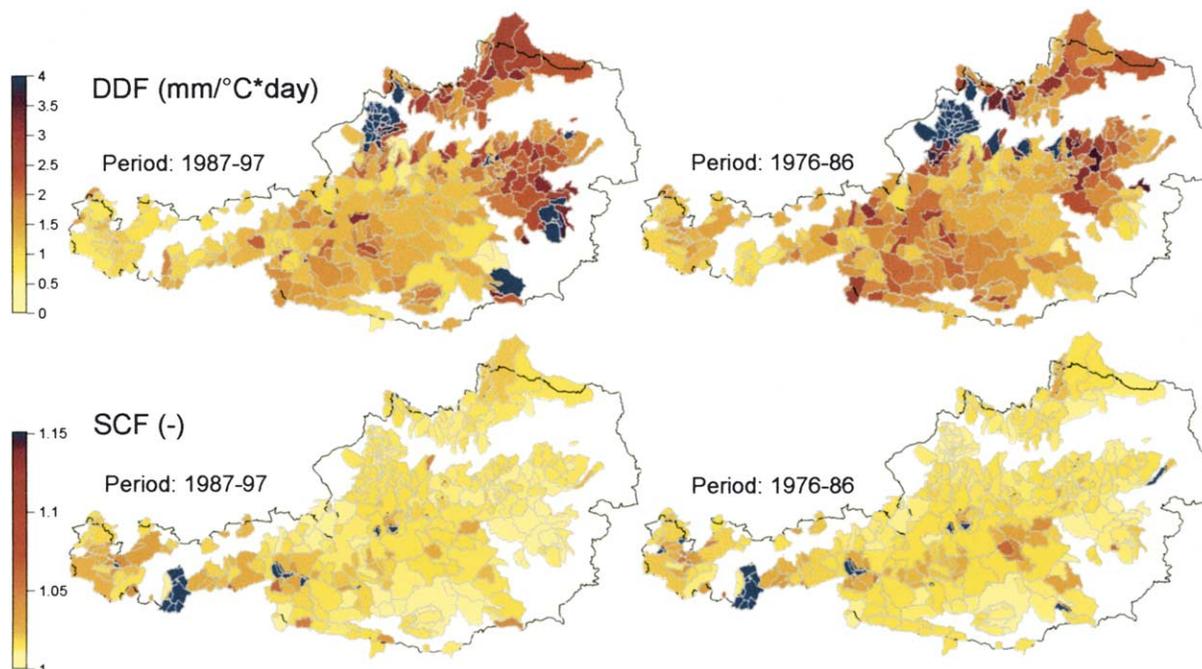


Fig. 4. Patterns of calibrated snow model parameters (left: calibration period 1987–1997, right: calibration period 1976–1986). Top: degree day factor,  $DDF$ , ( $\text{mm}/(\text{day } ^\circ\text{C})$ ); bottom: snow correction factor,  $SCF$  (-).

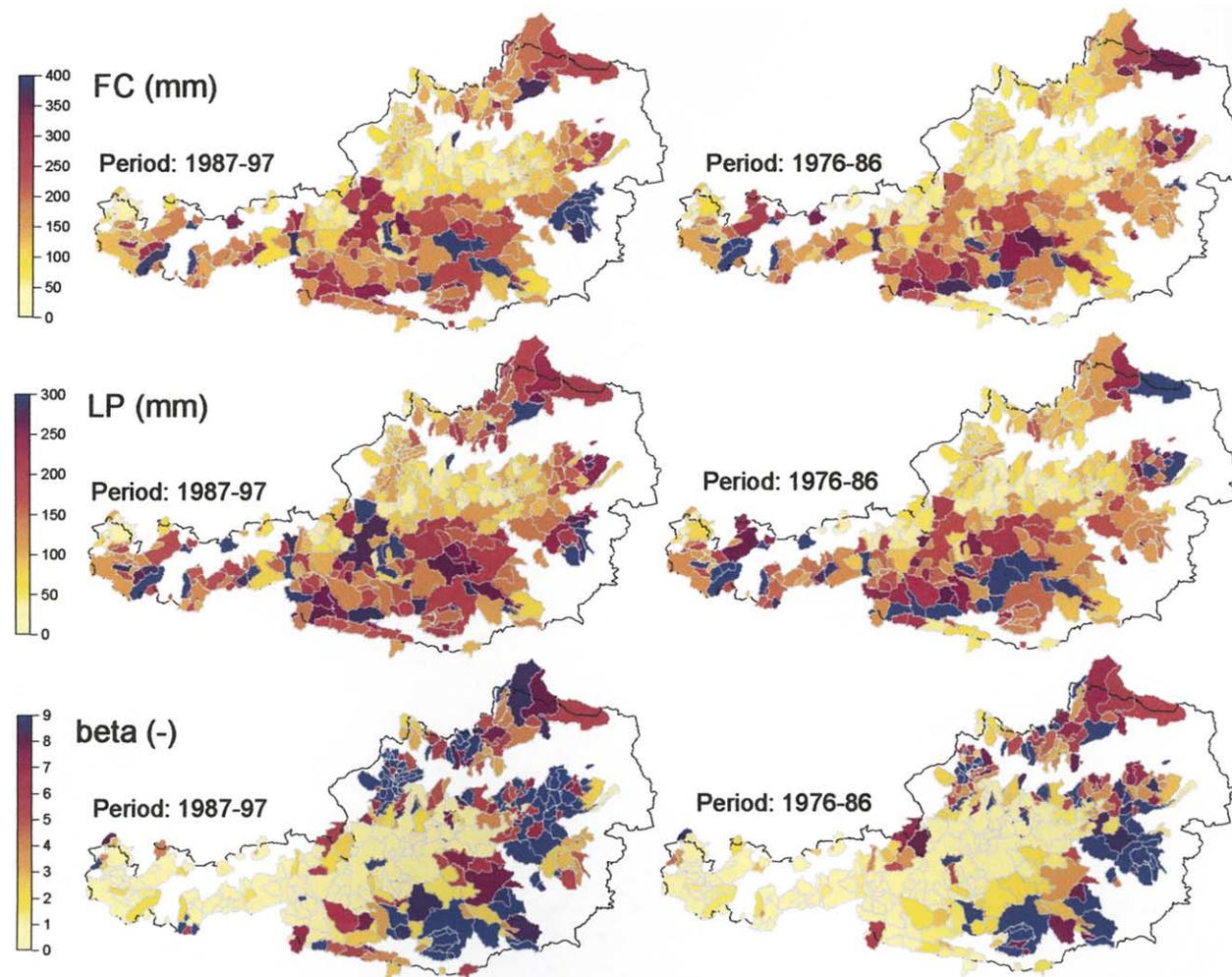


Fig. 5. Patterns of calibrated soil moisture model parameters (left: calibration period 1987–1997, right: calibration period 1976–1986). Top: Maximum soil moisture storage,  $FC$  (mm); centre: limit for potential evaporation,  $LP$  (mm); bottom: non-linearity parameter,  $\beta$  (-).

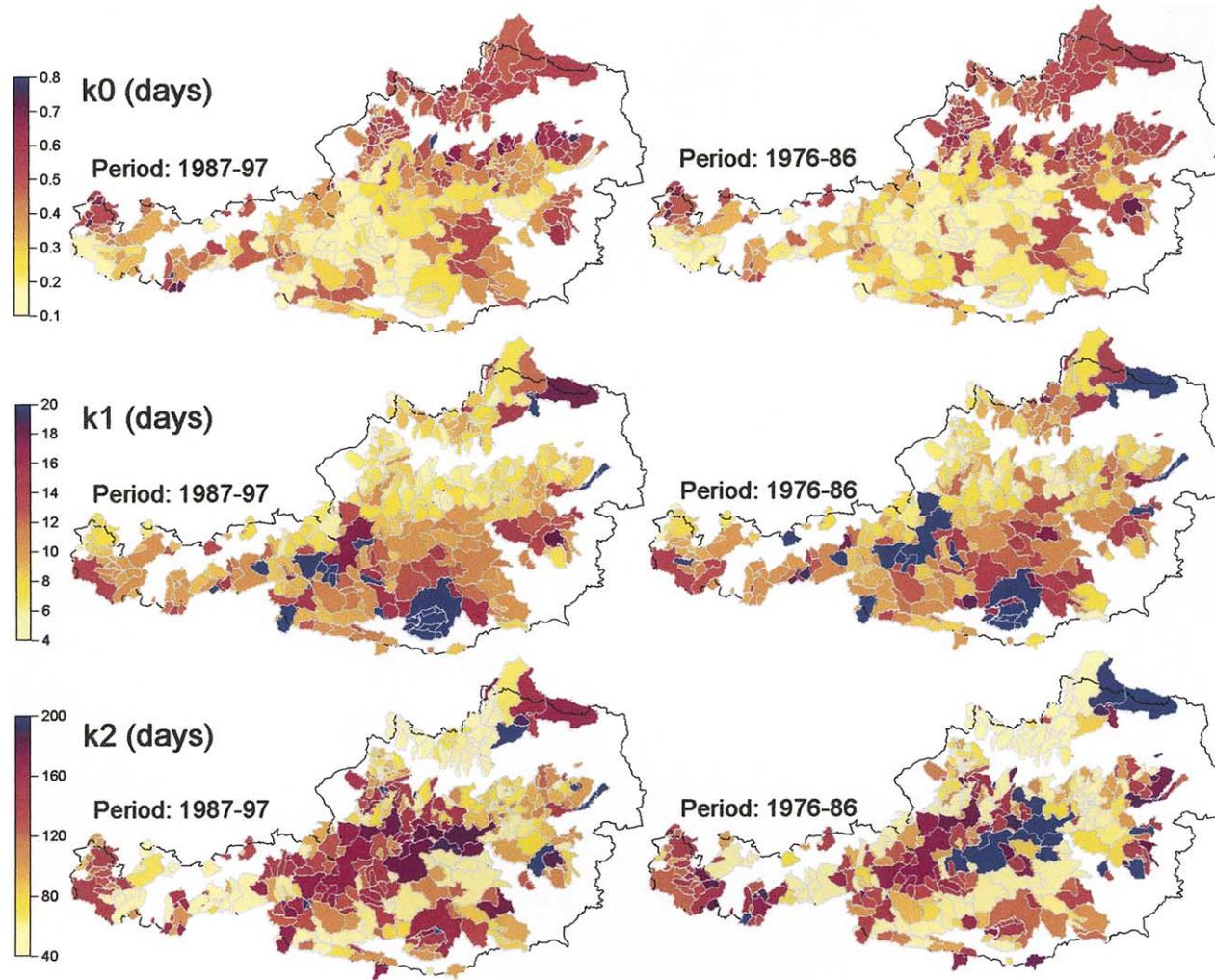


Fig. 6. Patterns of calibrated storage parameters (left: calibration period 1987–1997, right: calibration period 1976–1986). Top:  $k_0$  storage coefficient (days); centre: fast storage coefficient  $k_1$  (days); bottom: slow storage coefficient  $k_2$  (days).

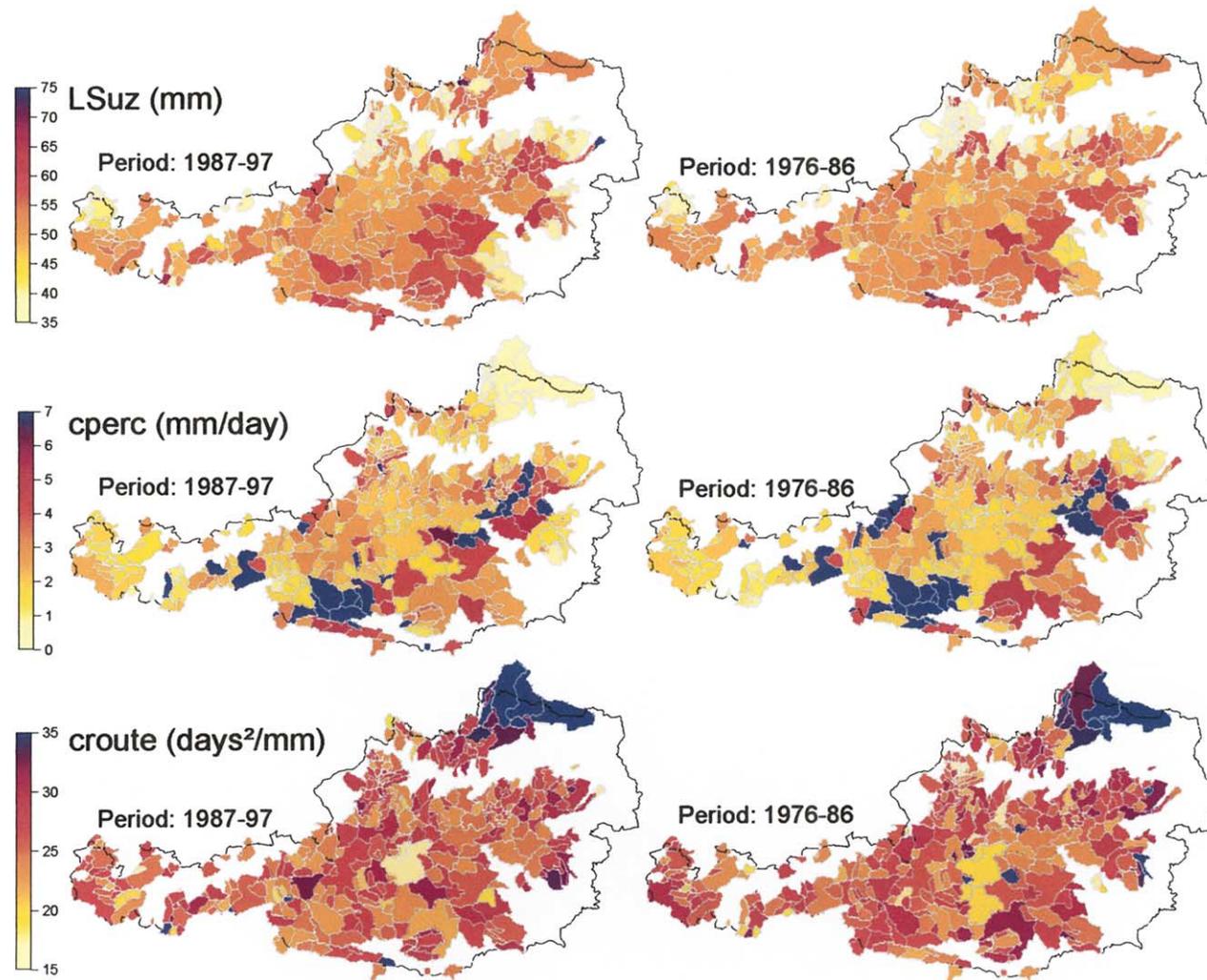


Fig. 7. Patterns of calibrated runoff response parameters (left: calibration period 1987–1997, right: calibration period 1976–1986). Top: storage threshold,  $LS_{uz}$  (mm); centre: percolation rate,  $c_{perc}$  (mm/day); bottom: routing parameter,  $c_{route}$  (days<sup>2</sup>/mm).

about 3–4 mm/(day °C) as shown for these catchments are in the upper range of values reported in the literature (WMO, 1986). In the Alpine catchments in the west of the country, rain on snow is of minor importance and radiation melt may contribute significantly to runoff, thus the *DDF* values are much lower (of the order of 1 mm/(day °C)). A similar pattern of the *DDF* has been found for the 1976–1986 period, however, the difference between the high and low altitude catchments is smaller.

The two patterns of the calibrated *SCF*, are similar to each other, with low values in the lowland and prealpine catchments of eastern Austria (see Fig. 1). There are a few outliers, i.e. individual small catchments with significantly larger *SCF* values than the surrounding catchments, which are a result of no rain gauges being located in those catchments. Large *SCF* values have been found for some of the high Alpine catchments in western Austria. In the higher altitude catchments, precipitation gauges are usually more exposed to wind, and snowfall tends to occur at lower temperatures, so one would expect larger deficits. Catch deficits of up to 50% are not unusual (Sevruk et al., 1998) which translate into *SCF* values of up to 1.5. The largest *SCF* values found here are 1.5.

In Fig. 5, the regional patterns of the soil moisture parameters for the 1987–1997 calibration period are shown left and those for the 1976–86 calibration period are shown right. The maximum soil moisture storage, *FC* (top), tends to exhibit large values in southern Austria. At the northern fringe of the high Alps and in most of the high alpine catchments of Tyrol in the west of the country, the *FC* values are small. The small *FC* values imply shallow hydrologically active soil depths, which may be realistic given that bare rock covers a substantial portion of the catchment areas in these regions. The patterns of the two calibration periods are reasonably similar. The patterns of the limit for potential evaporation, *LP*, (Fig. 5, centre) are similar to those of *FC*. *FC* has been defined as the upper limit of *LP* in the model and in most catchments, *LP* is equal to *FC* (also see Fig. 3). This means that there is a tendency for the evapotranspiration not to be at its potential rate most of the time. The non-linearity parameter, *beta*, (Fig. 5, bottom) shows distinct patterns of high values in eastern Austria and low values in western Austria for

both periods. Low values of *beta* are consistent with a linear rainfall-runoff relationship and large event runoff coefficients while the opposite is true for large values of *beta*. The regional differences in *beta* can thus be interpreted as implying a relatively linear rainfall-runoff relationship and large runoff coefficients in the wetter alpine catchment in the west, and a non-linear rainfall-runoff relationship and small runoff coefficients in the dryer lowland catchments in the east. These differences in the linearity are consistent with the general understanding of runoff generation processes in different climates (see, e.g. Goodrich et al., 1997).

In Fig. 6, the regional patterns of the storage coefficients for both periods are shown. The values of the  $k_0$  storage coefficient are smaller in the high alpine catchments in the south and west of the country than they are in the prealpine catchments of the north and in the lowlands of the north and east. This implies that, in the alpine catchments, flood runoff can be flashy once a threshold of  $LS_{uz}$  is exceeded. It is possible that the small  $k_0$  values are related to a large portion of surface flood runoff in these catchments. The fast storage coefficient,  $k_1$ , shows a tendency for an inverse pattern to that of  $k_0$  with faster responses in the prealpine catchments of the north than in the alpine catchments of the south. The inverse pattern is consistent with the weakly negative correlation between  $k_0$  and  $k_1$  found in Fig. 3. The patterns of  $k_1$  may be an indication that, in the alpine catchments, direct runoff penetrates deeper into the subsurface than in the rest of Austria. The slow storage coefficient,  $k_2$ , exhibits patterns with no obvious interpretation although the patterns of the two periods are similar.

In Fig. 7, the regional patterns of  $LS_{uz}$ ,  $c_{perc}$  and  $c_{route}$  for both calibration periods are shown. It is not easy to interpret these patterns from a hydrological perspective. Slightly lower values of  $LS_{uz}$  than the global mean in the prealpine Danube region in the north of the country, may indicate that it takes less millimetres of rainfall in these catchments to produce a flash flood with a response of  $k_0$  than in other parts of Austria. The  $c_{perc}$  parameter exhibits the largest values in East Tyrol. The values of  $c_{route}$  tend to be large in a few catchments in northern Austria implying a more non-linear channel response than in other catchments, i.e. faster response with increasing discharge, but it is

unclear what is the hydrologic reason for these patterns.

For all the parameters in Figs. 4–7, the regional patterns are similar for both periods although large local differences occur. Part of the consistency may be related to analyse nested catchments. However, a closer examination of upstream and downstream neighbours indicates that strong regional similarities also exist across catchment boundaries. This suggests that the calibrated parameters are able to represent the regional or large-scale differences in the hydrological conditions and hence the daily runoff regime in Austria. One would, therefore, assume that it is possible to derive regional relationships between the calibrated parameter values and catchment attributes, with the caveat that local differences, and to some degree parameter identifiability issues, may generate some noise.

## 6. Model parameters vs. catchment attributes

As a first step, we examined single correlations between the calibrated model parameters and each catchment attribute. The choice of catchment attributes has been guided by a general appreciation of the interaction between the runoff regime, climate and physiography. We examined catchment area, catchment average elevation, catchment average topographic slope, river network density, portion of catchment area with porous aquifers, the FARL lake index, the catchment average of mean annual precipitation, the catchment average of the long term mean of maximum annual daily precipitation, two land cover classes (portions of forest and glacier), three geologic units (portions of Tertiary + Quaternary, Calcareous Alps, Austroalpin crystalline) and two soil types (portions of Rendzina and Cambisol). Examples of the relationships between the calibrated model parameters and the catchment attributes are shown in Fig. 8. The ends of the error bars represent the parameter values found for the two calibration periods, and the full circles are the averages of the two periods. Short error bars represent similar parameters for the two periods, and hence reliable parameter values, while large error bars represent uncertain parameter values.

Fig. 8 shows the relationship between the fast storage coefficient,  $k_1$ , and catchment attributes. There is a tendency for small values of the fast storage coefficient not to occur in large catchments. This implies that large catchments never have a very flashy response, which is consistent with hydrologic experience. Similarly, high altitude catchments are never very flashy. For the other attributes, no obvious relationships exist. From a process-based reasoning, one would hope to find a relationship between  $k_1$  and attributes such as land use, geologic formation and/or soil type but this is not the case. For any of the attributes, the differences between the catchments are larger than what can be attributed to the uncertainty range of the error bars. This suggests that the lack of a relationship is not only due to the parameter uncertainty, but also due to the catchment attributes being poor indicators of  $k_1$ .

To examine this issue in more detail, for each catchment, we interpreted the calibrated parameter values of the two periods as the possible range of parameters within the uncertainty of parameter identifiability. We assumed that a true parameter value exists and lies within this range. If for any parameter value, within this range, a close correlation with the catchment attributes can be demonstrated, then the poor relationship is interpreted as a result of parameter uncertainty. There may exist an underlying relationship, which, however, is clouded by parameter uncertainty. If, in contrast, the correlation remains weak, then we suggest that there is no strong underlying relationship and the parameter uncertainty is relatively unimportant. We performed this analysis by an iterative approach. In a first step, we computed a simple linear regression between the average parameters of the two calibration periods and each catchment attribute. In a second step, we replaced the average parameter for each catchment by the point of intersection of the regression line and the parameter range, spanned by the calibrated values of the two periods. If the regression line did not intersect the parameter range, we used the nearest point of the parameter range instead. We then refitted a regression line to the changed data points and repeated the procedure until no improvement of the coefficient of determination was found. We repeated this procedure for each model parameter and each catchment attribute. The coefficients of determination found by

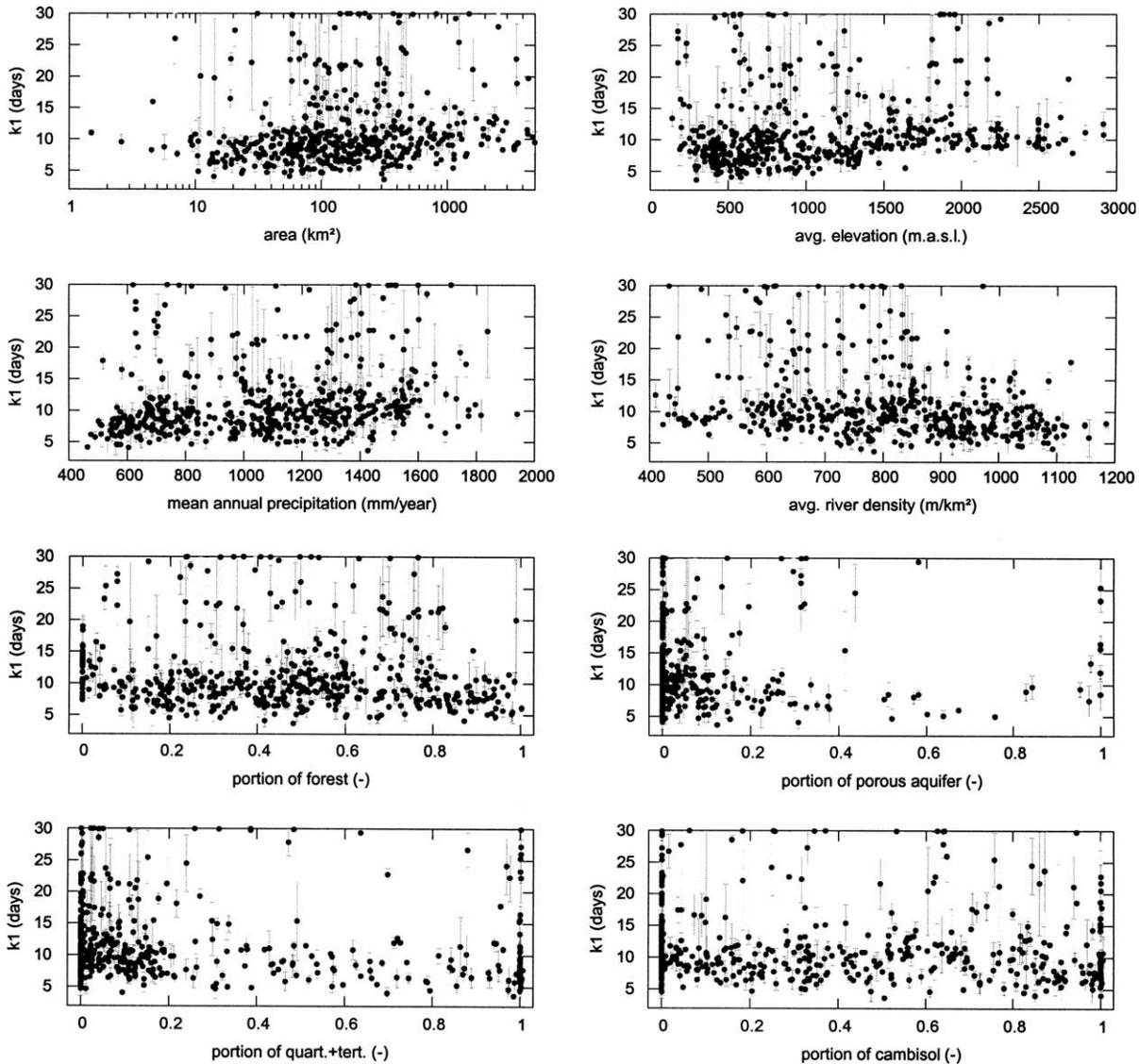


Fig. 8. Calibrated values of the fast storage coefficient,  $k_1$  (days), plotted against catchment attributes. The ends of the error bars represent the parameter values found for the two calibration periods (1987–1997 and 1976–1986) and the full circles are the averages of the two periods.

this iterative procedure are termed optimised coefficients of determination in this paper. They are always larger than the usual coefficients of determination as some of the parameter uncertainty is removed.

Table 2 shows the coefficients of determination,  $R^2$ , for the average parameters of the two periods and each catchment attribute as well as the optimised coefficients of determination in bold. Overall,

the correlations of the calibrated model parameters and the catchment attributes are rather weak. The attributes that are best related to the  $DDF$ , is the mean annual precipitation. The non-linearity parameter, beta, is mainly related to topographic elevation and topographic slope, the latter likely being a consequence of the interdependence of elevation and slope. As discussed above, large beta values stand for low

Table 2

Coefficients of determination,  $R^2$ , of single linear regressions between average calibrated model parameters of the two periods and catchment attributes (first numbers). Second numbers in bold are the  $R^2$  from optimised linear regression

$R^2$ (mean) $R^2$ (optimised)	DDF	SCF	FC	LP	Beta	$k_0$	$LS_{uz}$	$k_1$	$k_2$	$c_{perc}$	$c_{route}$
Area	0.01 <sup>-</sup> <b>0.02<sup>-</sup></b>	0.01 <sup>-</sup> <b>0.01<sup>-</sup></b>	0.01 <sup>-</sup> <b>0.03<sup>-</sup></b>	0.01 <sup>-</sup> <b>0.04<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.01<sup>-</sup></b>	0.07 <sup>+</sup> <b>0.14<sup>+</sup></b>	0.02 <sup>+</sup> <b>0.03<sup>+</sup></b>	0.01 <sup>-</sup> <b>0.02<sup>-</sup></b>	0.05 <sup>-</sup> <b>0.07<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.01<sup>+</sup></b>
Elevation	0.16 <sup>-</sup> <b>0.20<sup>-</sup></b>	0.10 <sup>+</sup> <b>0.12<sup>+</sup></b>	0.01 <sup>+</sup> <b>0.03<sup>+</sup></b>	0.05 <sup>+</sup> <b>0.14<sup>+</sup></b>	0.05 <sup>+</sup> <b>0.34<sup>-</sup></b>	0.26 <sup>-</sup> <b>0.31<sup>-</sup></b>	0.22 <sup>-</sup> <b>0.02<sup>+</sup></b>	0.01 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.06 <sup>-</sup> <b>0.18<sup>-</sup></b>
Slope	0.16 <sup>-</sup> <b>0.21<sup>-</sup></b>	0.04 <sup>+</sup> <b>0.05<sup>+</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.01 <sup>+</sup> <b>0.05<sup>+</sup></b>	0.25 <sup>-</sup> <b>0.31<sup>-</sup></b>	0.27 <sup>-</sup> <b>0.37<sup>-</sup></b>	0.02 <sup>+</sup> <b>0.02<sup>+</sup></b>	0.01 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.01 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.09 <sup>-</sup> <b>0.23<sup>-</sup></b>
RND	0.13 <sup>+</sup> <b>0.19<sup>+</sup></b>	0.04 <sup>-</sup> <b>0.05<sup>-</sup></b>	0.01 <sup>-</sup> <b>0.03<sup>-</sup></b>	0.01 <sup>-</sup> <b>0.02<sup>-</sup></b>	0.09 <sup>+</sup> <b>0.14<sup>+</sup></b>	0.00 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.12 <sup>-</sup> <b>0.22<sup>-</sup></b>	0.05 <sup>-</sup> <b>0.07<sup>-</sup></b>	0.01 <sup>-</sup> <b>0.04<sup>-</sup></b>	0.02 <sup>-</sup> <b>0.02<sup>-</sup></b>	0.01 <sup>-</sup> <b>0.01<sup>-</sup></b>
Porous aquifers	0.01 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.04 <sup>-</sup> <b>0.04<sup>-</sup></b>	0.01 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.00 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.01 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.06 <sup>+</sup> <b>0.06<sup>+</sup></b>	0.00 <sup>-</sup> <b>0.01<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.03 <sup>+</sup> <b>0.05<sup>+</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.00 <sup>-</sup> <b>0.01<sup>-</sup></b>
FARL	0.01 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.00 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.01 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.01 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.01 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.05 <sup>+</sup> <b>0.08<sup>+</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.02 <sup>-</sup> <b>0.02<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.01 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>
MAP	0.18 <sup>-</sup> <b>0.30<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.03 <sup>+</sup> <b>0.04<sup>+</sup></b>	0.06 <sup>+</sup> <b>0.11<sup>+</sup></b>	0.08 <sup>-</sup> <b>0.11<sup>-</sup></b>	0.19 <sup>-</sup> <b>0.29<sup>-</sup></b>	0.05 <sup>+</sup> <b>0.10<sup>+</sup></b>	0.05 <sup>+</sup> <b>0.06<sup>+</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.01 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.01 <sup>-</sup> <b>0.03<sup>-</sup></b>
MADP	0.08 <sup>-</sup> <b>0.12<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.07 <sup>-</sup> <b>0.11<sup>-</sup></b>	0.06 <sup>-</sup> <b>0.10<sup>-</sup></b>	0.06 <sup>-</sup> <b>0.07<sup>-</sup></b>	0.02 <sup>-</sup> <b>0.02<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.02 <sup>-</sup> <b>0.03<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.01 <sup>+</sup> <b>0.02<sup>+</sup></b>	0.07 <sup>-</sup> <b>0.20<sup>-</sup></b>
Forest	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.03 <sup>-</sup> <b>0.03<sup>-</sup></b>	0.01 <sup>-</sup> <b>0.04<sup>-</sup></b>	0.02 <sup>-</sup> <b>0.07<sup>-</sup></b>	0.04 <sup>+</sup> <b>0.08<sup>+</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.01 <sup>-</sup> <b>0.02<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.01<sup>-</sup></b>	0.01 <sup>+</sup> <b>0.02<sup>+</sup></b>	0.01 <sup>-</sup> <b>0.03<sup>-</sup></b>
Glacier	0.02 <sup>-</sup> <b>0.03<sup>-</sup></b>	0.32 <sup>+</sup> <b>0.33<sup>+</sup></b>	0.00 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.00 <sup>+</sup> <b>0.02<sup>+</sup></b>	0.06 <sup>-</sup> <b>0.10<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.10<sup>-</sup></b>	0.01 <sup>-</sup> <b>0.02<sup>-</sup></b>	0.01 <sup>-</sup> <b>0.03<sup>-</sup></b>
Tertiary + Quaternary	0.17 <sup>+</sup> <b>0.24<sup>+</sup></b>	0.01 <sup>-</sup> <b>0.02<sup>-</sup></b>	0.01 <sup>+</sup> <b>0.02<sup>+</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.13 <sup>+</sup> <b>0.17<sup>+</sup></b>	0.13 <sup>+</sup> <b>0.17<sup>+</sup></b>	0.09 <sup>-</sup> <b>0.18<sup>-</sup></b>	0.01 <sup>-</sup> <b>0.01<sup>-</sup></b>	0.07 <sup>+</sup> <b>0.11<sup>+</sup></b>	0.01 <sup>-</sup> <b>0.01<sup>-</sup></b>	0.02 <sup>+</sup> <b>0.05<sup>+</sup></b>
Calcareous Alps	0.02 <sup>-</sup> <b>0.02<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.05 <sup>-</sup> <b>0.08<sup>-</sup></b>	0.06 <sup>-</sup> <b>0.10<sup>-</sup></b>	0.02 <sup>+</sup> <b>0.02<sup>+</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.010 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.03 <sup>-</sup> <b>0.05<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.02 <sup>-</sup> <b>0.05<sup>-</sup></b>
Austroalpin crystalline	0.03 <sup>-</sup> <b>0.06<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.01<sup>+</sup></b>	0.10 <sup>+</sup> <b>0.12<sup>+</sup></b>	0.11 <sup>+</sup> <b>0.22<sup>+</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.04 <sup>-</sup> <b>0.06<sup>-</sup></b>	0.03 <sup>+</sup> <b>0.05<sup>+</sup></b>	0.03 <sup>+</sup> <b>0.05<sup>+</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>
Rendzina	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.04 <sup>-</sup> <b>0.05<sup>-</sup></b>	0.05 <sup>-</sup> <b>0.09<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.02 <sup>-</sup> <b>0.03<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.02 <sup>+</sup> <b>0.03<sup>+</sup></b>	0.01 <sup>-</sup> <b>0.04<sup>-</sup></b>
Cambisol	0.17 <sup>+</sup> <b>0.26<sup>+</sup></b>	0.03 <sup>-</sup> <b>0.04<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>	0.16 <sup>+</sup> <b>0.24<sup>+</sup></b>	0.01 <sup>+</sup> <b>0.02<sup>+</sup></b>	0.04 <sup>-</sup> <b>0.07<sup>-</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.01 <sup>+</sup> <b>0.03<sup>+</sup></b>	0.00 <sup>-</sup> <b>0.00<sup>-</sup></b>	0.00 <sup>+</sup> <b>0.00<sup>+</sup></b>

Plus and minus signs relate to direct and indirect relationships, respectively. Catchment attributes are: catchment area; catchment average elevation; catchment average topographic slope; river network density, RND; portion of catchment area with porous aquifers; the FARL lake index; the catchment average of mean annual precipitation, MAP; catchment average of the long term mean of maximum annual daily precipitation, MADP; two land cover classes (portions of forest and glacier); three geologic units (portions of Tertiary + Quaternary, Calcareous Alps, Austroalpin crystalline); two soil types (portions of Rendzina and Cambisol)

runoff coefficients and non-linear runoff generation behaviour, which prevail in the lowland catchments of eastern Austria, hence there is a negative relationship in Table 2. The  $k_0$  storage coefficient is negatively correlated with elevation and slope, implying that direct surface runoff may be particularly flashy in the high altitude catchments. There is also a tendency for the wetter catchments (large mean annual precipitation) to exhibit a flashy response and those catchment with a large portion of Tertiary and Quaternary deposits appear to have a tendency for

slower response. Both controls are consistent with hydrological reasoning. The storage coefficients,  $k_1$  and  $k_2$ , exhibit hardly any correlations. This is surprising as one would expect these two parameters to be related to soil type and geology.

It is now interesting to examine whether the optimisation procedure improves the correlations significantly (Table 2, bold numbers). For most parameters, attribute combinations where some relationship exists, the correlations increase. The increase in the coefficient of determination is typically

on the order of one third. However, for combination with low or non-existing correlations, the optimisation hardly increases the coefficients of determination. This means that the parameter uncertainty does not cloud the relationship between calibrated model parameters and catchment attributes to a significant extent. It should be noted that [Table 2](#) shows single regression of each parameter with only one catchment attributes. It is possible that the relationships are more complex and involve more than one attribute. Multiple regressions are examined later in this paper.

The low correlations obtained in this study are similar to those found in other studies that have examined a large number of catchments such as [Peel et al. \(2000\)](#). It appears that studies involving only a few catchments typically yield significantly better correlations. One explanation may be that the latter studies (such as [Post and Jakeman, 1999](#); [Seibert, 1999](#)) have, perhaps, been performed in hydrologically more uniform regions than is the case here. It is possible that in more uniform regions, the relationships are better defined. Later in this paper, we will, therefore, also examine local regressions that allow for the regression coefficients to change with space as the study region of this paper is indeed very heterogeneous. Another explanation is that, for small sample sizes, spurious correlations are more likely to occur than for a large sample size as examined here.

The most informative attributes, for a particular model parameter, found by other authors (e.g. [Peel et al., 2000](#); [Sefton and Howarth, 1998](#); [Seibert, 1999](#)) are not the same as those obtained here. This is not surprising, as one would expect the relationships between parameters and attributes to be a function of climate region, model structure and data aspects. This finding corroborates the notion that it will be difficult, if at all possible, to find universal relationships between model parameters and catchment attributes, at least at the regional scale as examined in this study.

## 7. Model efficiencies of regionalised model parameters

In this chapter, we more closely examine the potential of using catchment attributes for predicting

model parameters and put it into the context of other regionalisation methods. Ultimately, the predictive power of catchment attributes can be assessed by how well runoff can be simulated if parameters are only derived from catchment attributes without making use of locally observed runoff data. This is the important case of runoff simulations in ungauged catchments. When using regionalised parameters from catchment attributes, it is likely that the model performance will decrease as compared to using locally calibrated parameters. The decrease in model performance when moving from gauged catchments with local calibration to ungauged catchments, in this paper, is termed the spatial loss in model performance. For the predictive case, one would expect an additional loss in model performance as a result of moving from the calibration period to the prediction. In a simulation study as in this paper, the predictive performance is assessed by an independent verification period. The decrease in model performance when moving from the calibration period to the verification period, in this paper, is termed the temporal loss in model performance. We first examine the temporal loss and then the spatial loss in model performance.

*Temporal loss in model efficiency.* In [Fig. 9](#), the Nash–Sutcliffe model efficiencies (Eq. (B1)) for the verification periods have been plotted vs. the efficiencies for the calibration periods. The left panel shows the efficiencies of the 1976–1986 verification period and the 1987–1997 calibration period and the right panel shows the efficiencies for the swapped periods. An efficiency of 1 implies a perfect match of simulated and observed daily streamflow hydrographs and lower values imply increasingly poorer matches. In the left panel, most points are below the 1:1 line, which means that the model efficiencies tend to decrease when moving from the calibration to the verification periods. For the swapped periods, the points cluster around the 1:1 line, so the performances of the calibration and verification periods are similar. For clarity, catchments with verification efficiencies smaller than 0.2 are not shown in the figure. In the left panel, there were 16 catchments and in the right panel, there were 12 catchments.

The median efficiencies over all catchments are given in [Table 3](#) for the at-site case. As can be seen from [Table 3](#), the median decreases from 0.69 to 0.61 in the case of the left panel and slightly increases from

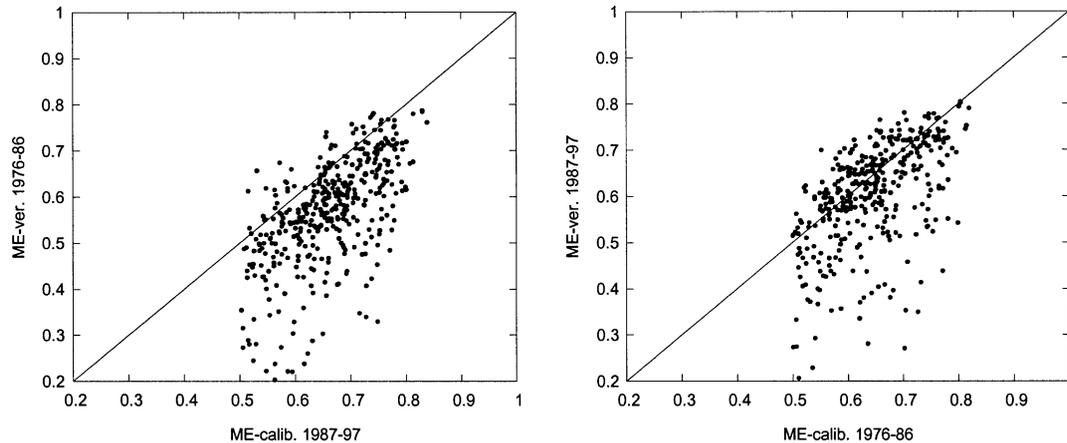


Fig. 9. Nash–Sutcliffe model efficiencies for the verification vs. calibration periods.

0.65 to 0.66 in the case of the right panel. This means that the temporal loss in the two cases is 0.08 and  $-0.01$ , respectively, or 0.04 on average over all years. There is a tendency for the model to perform better for the more recent period. This may be related to climate conditions that can be represented more accurately by the model or to a better data quality of the more recent period. The instrumentation in Austria has improved over the last years, which may contribute to the better model performance. It is, therefore, also of interest to compare the calibration and verification efficiencies for the same period. For the more recent period, the median efficiencies

decrease from 0.69 to 0.66 and for the earlier period, the median efficiencies decrease from 0.65 to 0.61, i.e. a loss of about 0.04 in both cases. This is a very small loss in model efficiency and allows us to draw important inferences on the potential of over-parameterisation of the model used. Bergström (1991, p. 129), for example, states: “If the model performance is significantly lower for the independent period used for validation than it was for the calibration period the modeller should seriously consider if there are problems of over-parameterisation. The model may simply have too many degrees of freedom for the information contained in the observed records.” We

Table 3

Model performance for gauged catchments (at-site) and ungauged catchments (various regionalisation procedures) both for the calibration and the verification periods

Median/scatter ( $ME_{med}/ME_{75-25\%}$ )	Cal.87-97	Ver.76-86	Cal.76-86	Ver.87-97	Cal. avg.	Ver. avg.
At-site	0.69/0.10	0.61/0.11	0.65/0.11	0.66/0.11	0.67/0.10	0.63/0.11
Preset	0.37/0.35	0.27/0.43	0.27/0.43	0.37/0.35	0.32/0.39	0.32/0.39
Global mean	0.42/0.27	0.32/0.35	0.33/0.33	0.41/0.25	0.37/0.30	0.36/0.30
Global regression	0.52/0.18	0.46/0.28	0.47/0.27	0.52/0.18	0.50/0.22	0.49/0.23
Local regression	0.55/0.19	0.48/0.28	0.50/0.28	0.54/0.21	0.53/0.23	0.51/0.25
Optimised local regression	0.56/0.19	0.49/0.27	0.49/0.26	0.56/0.17	0.53/0.22	0.53/0.22
Average of nested neighbours	0.60/0.15	0.54/0.20	0.55/0.22	0.57/0.16	0.57/0.18	0.56/0.18
Kriging	0.59/0.13	0.53/0.19	0.55/0.22	0.59/0.13	0.57/0.18	0.56/0.16
Kriging without nested neighbours	0.56/0.15	0.51/0.22	0.53/0.24	0.57/0.16	0.55/0.19	0.54/0.19

First value: median Nash–Sutcliffe efficiency. Second value: difference of 75 and 25% quantiles of efficiencies, i.e. a measure of scatter. High model performances are associated with large medians and small differences of the 75–25% quantiles. The columns denoted ‘avg.’ are the average efficiency measures of the 1987–1997 and 1976–1986 periods.

believe that the small decrease in median model efficiency when moving from the calibration to the verification period is a clear evidence for the model not being over-parameterised.

It is also interesting that Fig. 9 indicates that the calibration and verification efficiencies exhibit some correlation but there is not perfect relationship. This suggests that a good calibration efficiency does not necessarily entail a good predictive performance of the model. This results corroborates the generally held notion of the importance of split sample verification (e.g. Klemeš, 1986).

*Spatial loss in model efficiency.* The analysis of the regionalisation error, or spatial loss of model efficiency, is based on a jack-knifing approach, which emulates the case of ungauged catchments. In this approach, we treat one gauged catchment as ungauged and simulate the water balance dynamics using parameters estimated from regional information only. In a second step, we estimate the model performance by comparing the simulated and observed hydrographs. This comparison gives us a Nash–Sutcliffe efficiency and a volume error. We repeat the analysis for each catchment in turn and calculate the statistics of these error measures for all catchments. The comparison of these error measures with those for the locally calibrated case, both for the calibration and verification periods, indicates what decrease of model performance one would have to expect when moving from gauged to ungauged catchments.

We examined eight regionalisation methods. In a first regionalisation method, we selected one parameter set a priori (preset parameters) and applied it to all catchments. The choice is based on expert judgement and is identical to the most likely parameter values as used in the a priori distribution of the calibration procedure (Table A1). In a second method, we selected each parameter as the mean value of all the calibrated values in Austria and applied this parameter set to all catchments (termed global mean). In a third, fourth and fifth method, we used catchment attributes. As a representative of regionalisation approaches that are based on catchment attributes, we used multiple regressions with three catchment attributes. As some of the catchment attributes are highly skewed, one would expect that the residuals are also skewed. Standard regression, however, requires

the residuals to be normally distributed. We, therefore, transformed all catchment attributes to be standard normally distributed. We estimated the regression coefficients by ordinary least squares. Out of the 15 available attributes, we only used the set of those three attributes that were associated with the largest multiple correlation coefficient for each station and each model parameter. The rationale of this choice is that a large correlation coefficient may also be a good indicator of the predictive power of the attributes. In the third regionalisation method, we used a global regression system based on all, but one catchments at the same time. As only one catchment out of 308 catchments is left out in turn in the jack-knifing approach, the regression coefficients are always very similar. The fourth method is similar to the third method but we only included a local neighbourhood of 50 km radius in the regression, which resulted in a regression system involving about 20–50 stations. The fifth method is similar to the fourth method but we optimised the correlation coefficients within the range of the parameters obtained from the two calibration periods analogously to the way we obtained the bold coefficients of determination in Table 2. We examined three alternative regionalisation methods that do not use catchment attributes (methods six to eight). In the sixth method, we used the average parameter values of the immediate upstream and downstream neighbours of each catchment (termed average of nested neighbours). For headwater catchments, we only used the parameters from the downstream neighbour. The rationale behind this method is that these are nested catchments so one would expect the parameters to be similar. The seventh and eighth methods are based on spatial interpolation by kriging (Deutsch and Journel, 1997). The kriging estimates are only based on spatial proximity between the gauged catchments and the ungauged site of interest and do not use catchment attributes. We measured the spatial distance between two catchments by the spatial distance of the respective catchment centroids. We used ordinary kriging based on an exponential variogram with a nugget of 10% of the observed variance, a sill equal to the variance, and a range of 60 km. This is consistent with the empirical variograms of most of the calibrated model parameters. In the seventh method, we used all catchments for the spatial interpolation

(with the exception of the ungauged catchment), while in the eighth method, we left out the immediate upstream and downstream neighbours to assess the effect of nested catchments. We termed the eighth method kriging without nested neighbours.

In Fig. 10, the Nash–Sutcliffe model efficiencies using regionalised parameters have been plotted against the model efficiencies using calibrated parameters for the 1987–1997 period. The labels (a)–(h) relate to the eight methods described above. Table 3 gives the medians and, as a measure of scatter, the differences of the 75 and 25% quantiles of the efficiencies for each of the regionalisation methods. For comparison, Table 4 gives the attending volume errors for all cases. The vertical axes in Fig. 10 relate

to the columns labelled ‘Cal.87-97’ (lines 2–9) in Tables 3 and 4. For almost all catchments, the model efficiencies based on regionalised parameters are significantly lower than those based on calibrated parameters (Fig. 10). This is also true of the other period as indicated in Tables 3 and 4. This finding is not surprising as the Nash–Sutcliffe efficiency is an important component of the objective function in the calibration procedure (Appendix B), so one would expect a significant decrease of the optimised value.

Using preset parameters (Fig. 10a, and second line in Tables 3 and 4) produces the poorest regionalisation results, both in terms of median efficiencies, scatter of the efficiencies and volume errors. The median efficiencies for the two periods are 0.37 and

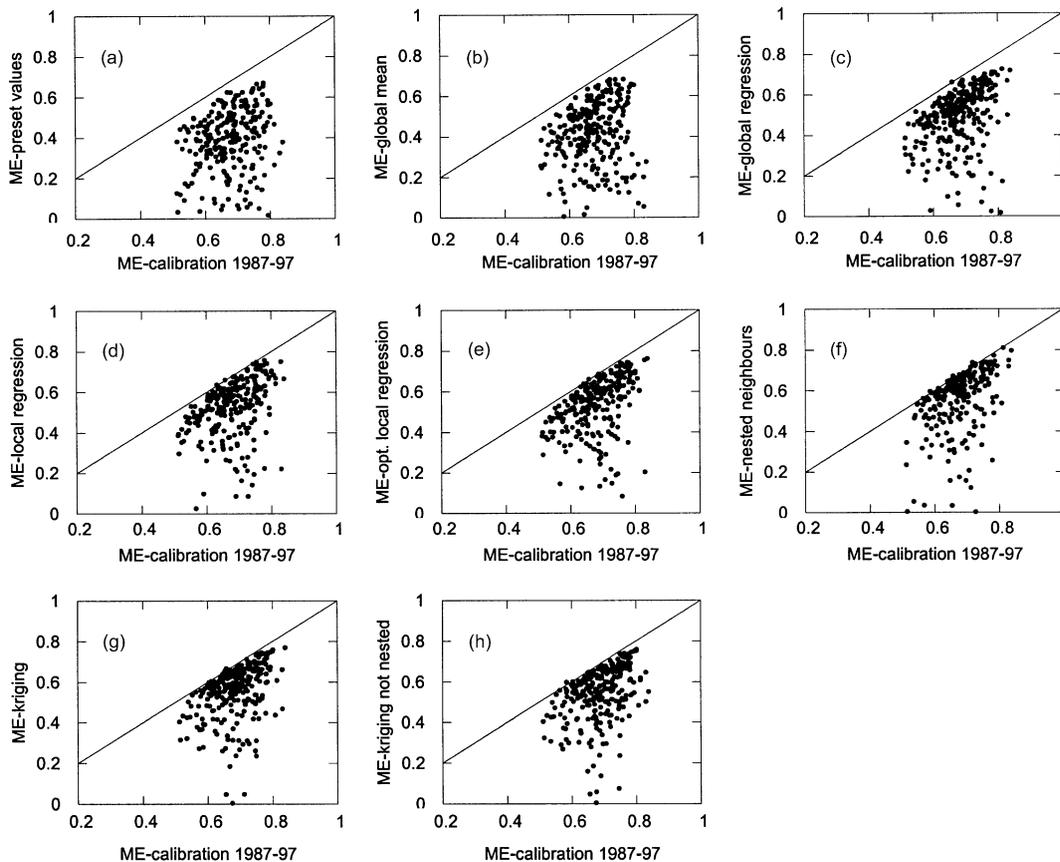


Fig. 10. Nash–Sutcliffe model efficiencies of regionalised versus calibrated parameters for the period 1987–1997. (a) Preset parameters, (b) global mean of all catchments, (c) global regression with catchment attributes using all catchments, (d) local regression within a 50 km neighbourhood, (e) optimised local regression, (f) average parameters of immediate upstream and downstream (nested) neighbours, (g) kriging, (h) kriging without nested neighbours.

Table 4

Volume errors for gauged catchments (at-site) and ungauged catchments (various regionalisation procedures) both for the calibration and the verification periods

Median/scatter ( $VE_{\text{med}}/VE_{75-25\%}$ )	Cal.87-97	Ver.76-86	Cal.76-86	Ver.87-97	Cal. avg.	Ver. avg.
At-site	0.00/0.04	-0.05/0.10	0.00/0.10	0.04/0.05	0.00/0.04	-0.01/0.10
Preset	0.13/0.32	0.03/0.33	0.03/0.34	0.13/0.32	0.08/0.33	0.08/0.32
Global mean	0.07/0.31	-0.03/0.33	-0.04/0.32	0.06/0.31	0.02/0.32	0.02/0.32
Global regression	0.08/0.30	-0.02/0.27	-0.02/0.24	0.06/0.27	0.03/0.27	0.02/0.27
Local regression	0.05/0.25	-0.04/0.24	-0.01/0.22	0.07/0.26	0.02/0.24	0.02/0.25
Optimised local regression	0.05/0.26	-0.04/0.24	-0.02/0.23	0.06/0.25	0.02/0.24	0.01/0.24
Average of nested neighbours	0.07/0.22	-0.02/0.20	0.01/0.21	0.09/0.23	0.04/0.21	0.04/0.21
Kriging	0.07/0.22	-0.02/0.27	0.00/0.25	0.10/0.21	0.04/0.23	0.04/0.24
Kriging without nested neighb.	0.06/0.23	-0.02/0.26	0.00/0.26	0.10/0.23	0.03/0.25	0.04/0.24

First value: median volume error. Second value: difference of 75 and 25% quantiles of volume errors, i.e. a measure of scatter. High model performances are associated with small volume errors (both median and scatter). The columns denoted 'avg.' are the average volume error measures of the 1987–1997 and 1976–1986 periods.

0.27. This means that the predictive power of parameters selected on the basis of expert judgement is rather poor. Using the global mean parameter set (Fig. 10b, and third line in Tables 3 and 4) slightly improves the predictive performance on all scores but the performance is still poor. It is clearly very important to account for differences between catchments, and using the same parameter set for the entire study region is inappropriate. Fig. 10c and fourth line in Tables 3 and 4 represent the case of using a global regression with catchment attributes. The median efficiencies increase to 0.52 and 0.47 for the two calibration periods and Fig. 10c indicates that quite a few of the points are closer to the 1:1 line. This suggests that the catchment attributes do contain valuable information that can be used to improve the parameter estimates beyond global mean values. Using the local regressions yet slightly improves the efficiencies and there is also a slight decrease in the scatter of the volume errors as indicated by Table 4. This result suggests that it is useful to account for regional differences in the regression equations. As compared to the at-site calibrated simulations, the loss in median model efficiency for the regionalisation based on local regressions is  $0.67 - 0.53 = 0.14$  averaged over the two calibration periods. This is about three times the temporal loss of 0.04. The optimised local regression performs only slightly better than the local regression, which suggests that accounting for parameter uncertainty does not

improve the regionalisation performance much. The regionalisation method that uses the average parameters of immediate upstream and downstream (nested) neighbours performs best on most scores (Fig. 10f, and seventh line in Tables 3 and 4). The median model efficiencies for the two calibration periods are 0.60 and 0.55, which are the largest values of all methods and the scatter is only 0.15 and 0.22. Similarly, the scatter of the volume errors is the smallest of all methods (0.22 and 0.21). As compared to the at-site calibrated simulations, the loss in median model efficiency for the regionalisation based on nested neighbours is  $0.67 - 0.57 = 0.10$ , averaged over the two calibration periods. This is about twice the temporal loss of 0.04. Kriging performs only slightly poorer. There is a slight decrease in performance for the case of kriging, where immediate (nested) neighbours are not used for the interpolation (Fig. 10h, and last line in Tables 3 and 4). For this method, the median efficiencies are still 0.56 and 0.53. This means that the favourable performance of kriging is not only a result of the same portion of the landscape draining into nested catchments. There appear to exist important similarities of model parameters across catchment boundaries. It is likely that these similarities are related to real hydrological controls that vary smoothly in space. An important finding is that the methods based on spatial proximity alone (the last three methods) all perform better than any of the regression methods based on catchment

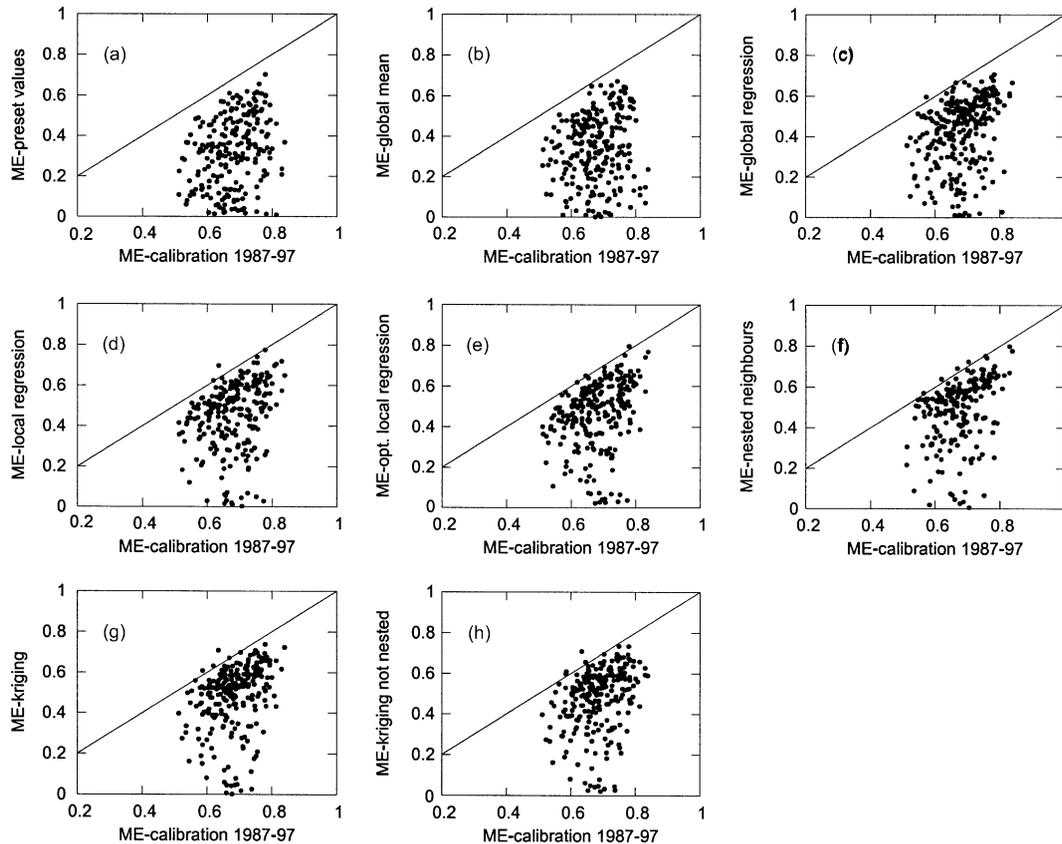


Fig. 11. Nash–Sutcliffe model efficiencies. Vertical axes: verification efficiencies for the period 1976–1986 using regionalised parameters estimated from the period 1987–1997. Horizontal axes: calibration efficiencies for the period 1987–1997. (a–h) are as in Fig. 10.

attributes. Apparently, the catchment attributes used here are not very representative of the real physical controls on the water balance dynamics.

*Spatio-temporal loss in model efficiency.* Fig. 11 shows the combined effect of temporal and spatial losses in the model performance. Specifically, the vertical axes in Fig. 11 relate to the columns labelled ‘Ver.76-86’ in Tables 3 and 4, i.e. verification efficiencies for the period 1976–1986 using regionalised parameters estimated from the period 1987–1997. The relative performance of the regionalisation methods in the combined case is similar to that in the spatial case. The poorest performances are obtained for preset parameters and the global mean parameter values. The local regressions perform slightly better than the global regressions and the optimised local regressions are still a little better.

As compared to the at-site calibrated simulations, the combined loss in median model efficiency for the regionalisation based on local regressions is  $0.67 - 0.51 = 0.16$  averaged over the two verification periods. The nested neighbour regionalisation and kriging yield the best results. As compared to the at-site calibrated simulations, the combined spatio-temporal loss in median model efficiency is  $0.67 - 0.56 = 0.11$  averaged over the two verification periods, which is slightly more than the loss of 0.10 in spatial model regionalisation only. Again, all the methods based on spatial proximity perform better than those based on catchment attributes.

*Catchment attributes used in the regression method.* To assist in the interpretation of the regression results in Figs. 10 and 11 as well as Tables 3 and 4, Table 5 shows how often each

Table 5  
Number of instances each catchment attribute is used in the three parameter multiple local regression

	<i>DDF</i>	<i>SCF</i>	<i>FC</i>	<i>LP</i>	$\beta$	$k_0$	$LS_{uz}$	$k_1$	$k_2$	$c_{perc}$	$c_{route}$
Area	0/165	1/136	22/12	37/7	21/54	0/108	124/3	22/13	40/15	21/48	16/42
Elevation	0/66	4/50	61/15	59/16	9/77	0/66	50/12	29/28	10/16	6/56	21/29
Slope	0/75	1/91	5/61	2/65	29/52	0/83	83/9	21/15	23/41	32/55	26/66
RND	0/85	2/78	24/60	29/45	32/17	0/46	7/44	26/88	12/95	11/69	17/43
Porous aquifers	6/37	8/42	64/28	69/29	37/24	3/48	24/26	14/49	36/28	16/39	20/50
FARL	12/33	8/68	38/47	42/44	29/27	8/70	4/78	11/98	55/64	10/90	23/46
MAP	0/35	2/83	32/12	36/12	10/21	0/45	10/28	35/10	15/37	9/45	29/53
MADP	1/55	1/72	6/61	8/39	13/53	1/35	16/31	13/54	42/30	7/62	2/66
Forest	0/82	3/44	33/9	29/9	26/43	0/79	121/17	20/24	13/17	27/57	20/50
Glacier	0/66	0/9	4/22	5/19	17/13	0/37	4/18	13/11	3/13	3/15	7/16
Tertiary + Quaternary	0/29	1/76	15/22	21/26	30/38	0/41	19/21	4/47	56/36	6/31	39/59
Calcareous alps	1/44	1/60	18/46	9/54	11/64	1/88	12/38	5/77	17/52	24/45	15/39
Austroalpin Crystalline	0/44	0/17	66/4	64/5	11/23	0/55	11/22	38/32	10/18	7/26	6/36
Rendzina	5/23	3/40	9/75	10/79	30/26	3/65	9/32	25/36	20/31	13/65	9/30
Cambisol	0/60	0/23	25/28	29/26	54/33	0/42	39/12	13/53	24/55	2/27	15/34
Total	924	924	924	924	924	924	924	924	924	924	924

Number of instances is a measure of the predictive power of each attribute. The two numbers shown relate to positive and negative regression coefficients, respectively. As there are 399 catchments and three variables the total number of instances is 1197. Calibration period 1987–1997

catchment attribute has been selected for the local regression with each model parameter as a consequence of exhibiting the best multiple correlation coefficient. Two numbers are shown. The first is the number of instances an attribute was selected associated with a positive regression coefficient, the second is the number of instances associated with a negative regression coefficient. The attributes used in most of the catchments can be interpreted as those with the highest predictive power. Some but not all of the attributes with the highest predictive power can be interpreted on the basis of hydrological reasoning. The *SCF* is negatively related to catchment area implying that which may be related to raingauge density. The maximum soil moisture content, *FC*, is positively related to the portion of porous aquifers in 64 out of 308 catchments. This suggests that porous aquifers tend to increase the storage capacity of a catchment. *FC* is positively related to Austroalpin crystalline implying particularly deep soils for this geologic formation. The non-linearity parameter, beta, is negatively related to elevation and topographic slope. As discussed above, this is likely due to a tendency for the alpine catchments to be associated with larger runoff coefficients and more linear runoff

generation than lowland catchments. For the  $k_0$  storage coefficient and its threshold value,  $LS_{uz}$ , two attributes used frequently are catchment area and the portion of forest. Small catchments are associated with small values of  $LS_{uz}$  and hence tend to produce surface runoff more easily than large catchments. In contrast, forested catchments are associated with large values of  $LS_{uz}$  and hence tend to produce surface runoff less frequently than catchments with other land uses. The controls of  $k_0$  and  $LS_{uz}$  tend to have inverse signs which is because of a tendency of these two parameters to be negatively related to each other (see Fig. 3). The fast storage coefficients  $k_1$  is negatively related to the FARL lake index which implies that catchments with a large portion of lakes tend to respond more slowly than catchments without lakes. The percolation parameter,  $c_{perc}$ , is negatively related to river network density. This suggests that in catchments with few streams a larger portion of water penetrates deep into the subsurface than is the case for catchments with a large river network density.

*Discussion.* The model efficiencies for the calibration and verification periods of the gauged catchment case found in this paper (median values

of 0.67 and 0.63) are reasonable but they are not as high as one would expect them to be from the literature (e.g. WMO, 1986). We believe there are three main reasons for the efficiencies not to be higher. The first is that the objective function used in this paper consists of five terms, only one of which represents the Nash–Sutcliffe model efficiency. This objective function has been chosen in favour of more robust parameter estimation. It should be noted that the Nash–Sutcliffe model efficiency only contributed about two thirds to the total objectives function in the case of the current study. The calibration efficiencies are, therefore, not as high as they were when only using Nash–Sutcliffe model efficiency in optimising the model parameters. The second reason is that, unlike many studies in the literature, we have not hand picked the model for each catchment. For some of the catchments, the model structure may be less than perfect. We do believe that the model structure is adequate for the purposes of the paper as both the calibration and verification efficiencies are acceptable. The third reason is that there may also be some data problems which we have not detected but in an individual case study for a small set of catchments, as commonly reported in the literature, one would remove outliers and focus on the data for which the model gives consistent results. The efficiencies obtained here are similar to those of Perrin et al. (2001) who analysed a similar number of catchments. It should also be noted that, in this paper, the model efficiencies have been calculated for daily runoff values. In some studies, the efficiencies are reported for weekly or monthly averages of runoff (e.g. Chiew and McMahon, 1994; Vandewiele et al., 1994) for which the efficiencies would be better. In test simulations, not shown here, we have evaluated the efficiencies for different aggregation time steps. The median efficiencies increased from about 0.65 to 0.75 when moving from daily time steps to monthly time steps. This suggests that the seasonal cycle can be more easily represented by the model than the day-to-day fluctuations.

The ungauged catchment case, even for the best regionalisation method of using nested neighbours, results in median efficiencies of 0.57 and 0.56, which is a significant decrease from the gauged catchment case. This supports the notion that model calibration, if at all possible, is by far better than model parameter

regionalisation. Indeed, Chiew and McMahon (1994, p. 383), based on a similar but less extensive study, pointed out that it is “difficult to estimate the values of some of the important parameters” and hence the models “must always be calibrated in all modelling applications”. Using catchment attributes for parameter regionalisation yields still lower efficiencies than methods based on spatial proximity. This finding is counter intuitive, as one would hope the attributes as used in this paper to represent the physiographic and hence hydrologic characteristics of catchments well. This result adds a caveat to the use catchment models that focus on land use characteristics or other catchment attributes for estimating parameters (e.g. Kite, 2000), at least for environments such as the study region of this paper. It appears that the attributes available at the regional scale are not very good predictors of the hydrological dynamics. We believe that for parameter regionalisation, better predictive variables and similarity measures need to be found than those currently used. We are currently examining dynamic indicators, as opposed to the usual static attributes used in this paper, that are hoped to better represent catchment processes during individual events and should therefore also be better regional predictors (e.g. Merz and Blöschl, 2003). These indicators include seasonality measures, catchment state variables such as soil moisture and snow cover, as well as storm type indicators.

## 8. Conclusions

The comparison of model parameters calibrated on two independent 11 year periods suggests that all model parameters are associated with some uncertainty. The coefficients of determination of the parameters from the two periods range from 0.09 to 0.64. For 50–93% of all catchments, the differences in parameters between the two periods are less than 10% of the possible parameter range. The most uncertain parameters are the slow storage coefficient,  $k_2$  and the routing parameter,  $c_{\text{route}}$ . The parameters with the smallest uncertainties are the fast storage coefficient,  $k_1$  and the threshold storage coefficient,  $k_0$ . Overall, the interdependences of the parameters are weak, if at all present. There is a tendency for the  $k_0$  and  $k_1$  storage coefficients to be negatively related and for  $k_0$

and its threshold value  $LS_{uz}$  to be negatively related. These interdependences can be interpreted based on the model structure. The lack of significant interdependences suggests that the number of model parameters cannot be reduced easily.

The regional patterns of the calibrated model parameters are similar for the two periods although large local differences occur. Some of the large-scale variability of parameters can be interpreted based on hydrological process reasoning. For example, large DDF are found in regions of frequent rain-on-snow events; large SCF, are found in high alpine catchments; small values of the non-linearity parameter beta, which imply a linear rainfall-runoff relationship and large runoff coefficients, are found in the alpine high precipitation catchments. A closer examination of upstream and downstream neighbours indicates that strong regional similarities not only exist for nested catchments but also across catchment boundaries. This suggests that the calibrated parameters are able to represent the regional or large-scale differences in the hydrological conditions and hence the daily runoff regime in Austria. One would, therefore, assume that it is possible to derive regional relationships between the calibrated parameter values and catchment attributes, with the caveat that local differences, and to some degree parameter identifiability issues, may generate some noise.

An analysis of the calibrated model parameters versus catchment attributes suggests that there is no tight relationship between the parameters and any of the attributes. However, the attributes do explain some of the spatial variability of the parameters. For example, the  $k_0$  storage coefficient is negatively correlated (coefficient of determination,  $R^2 = 0.22$  and  $0.27$ ) with elevation and slope, implying that direct surface runoff may be particularly flashy in the high altitude catchments. The coefficients of determination for the other model parameters are smaller. An optimised correlation procedure that accounts for some of the parameter uncertainty resulted in increase of the coefficient of determination by about one third. This means that the parameter uncertainty does not cloud the relationship between calibrated model parameters and catchment attributes to a significant extent. The underlying hydrological relationship seems to be rather weak as the catchment attributes are relatively poor indicators of the model parameters.

An analysis of the model performance suggests that, when moving from the calibration period to the verification period in the gauged catchment case, the median Nash–Sutcliffe model efficiency decreases from 0.67 to 0.63 on average over both periods, which implies a temporal loss in model efficiency of 0.04. This is a small decrease, which suggests that problems with over-parameterisation of the model are unlikely. A comparison of various regionalisation methods indicates that the spatial loss in model performance, when moving from gauged catchments to ungauged catchments, is significantly larger. Using the same parameter set for all catchments (either preset parameters or globally averaged parameters) produces the poorest regionalisation results. It is clearly very important to account for differences between catchments. Using multiple regressions with catchment attributes produces better regionalisation results. This suggests that the catchment attributes do contain valuable information that can be used to improve the parameter estimates beyond global mean values. Local regressions slightly improve the performance over global regressions indicating that it is useful to account for regional differences in the regression equations. The median model efficiency of local regressions is 0.53, implying a spatial loss of 0.14, which is about three times the temporal loss of 0.04. The optimised local regression performs only slightly better than the local regression, which suggests that parameter uncertainty does not affect the regionalisation significantly. In the multiple regressions, the set of the three attributes with the best correlations is used. Some but not all of these attributes can be interpreted on the basis of hydrological reasoning.

The regionalisation method that uses the average parameters of immediate upstream and downstream (nested) neighbours performs best on most scores. The median model efficiency of simulations using parameters from nested neighbours is 0.57, implying a spatial loss of 0.10, which is about twice the temporal loss of 0.04. Similarly, the scatter of the efficiencies and the scatter of the volume errors between catchments are relatively small. Kriging performs only slightly poorer. There is only a slight decrease in performance for the case of kriging, where immediate (nested) neighbours are not used for the interpolation. This means that the favourable

performance of kriging is not only a result of the same portion of the landscape draining into nested catchments. There appear to exist important similarities of model parameters across catchment boundaries. An important finding is that the methods based on spatial proximity alone (the last three methods) all perform significantly better than any of the regression methods based on catchment attributes. Apparently, catchment attributes are not very representative of the real physical controls on the water balance dynamics. Spatial proximity may be a useful surrogate for unknown controls on the runoff regime and hence on the model parameters but we believe that better predictive variables based on dynamic catchment characteristics are needed to make headway in regionalising catchment model parameters.

## Acknowledgements

The authors would like to thank the Austrian Science Foundation (FWF), project no. P14478-TEC, and the Austrian Academy of Sciences, project HÖ 18, for financial support. We would also like to thank the Austrian Hydrographic Service (HZB) and the Austrian Central Institute of Meteorology and Geodynamics (ZAMG) for providing the hydrographic and meteorological data. We would like to thank three anonymous reviewers for valuable comments on the manuscript.

## Appendix A

### A.1. Snow model

The snow routine represents snow accumulation and melt by a simple degree day concept

$$P_s = x \cdot P, \quad x = \begin{cases} 0, & \text{if } T \geq T_r \\ 1, & \text{if } T \leq T_s \\ \frac{T_r - T}{T_r - T_s}, & \text{if } T_s < T < T_r \end{cases} \quad (\text{A1})$$

$$P_r = (1 - x) \cdot P$$

where  $P$  is the daily value of catchment average precipitation,  $P_s$  is snowfall,  $P_r$  is rainfall,  $T$  is the mean daily air temperature, and  $T_s$  and  $T_r$  are lower and upper threshold temperatures, respectively. Melt starts at temperatures above a threshold  $T_m$

$$P_m = (T - T_m) \cdot DDF, \quad \text{if } T > T_m \text{ and SWE} > 0 \\ P_m = 0, \quad \text{otherwise} \quad (\text{A2})$$

where  $P_m$  is the amount of melt water per time step,  $DDF$  is the degree day factor and SWE is the snow water equivalent. Catch deficit of the precipitation gauges during snowfall is corrected by  $SCF$ , and changes in the snow water equivalent from days  $i - 1$  to  $i$  are simply accounted for by

$$\text{SWE}_i = \text{SWE}_{i-1} + (\text{SCF} \cdot P_s - P_m) \Delta t \quad (\text{A3})$$

where  $\Delta t$  is the time step of one day.

### A.2. Soil moisture accounting

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

$$S_{\text{sm},i} = S_{\text{sm},i-1} + P_r + P_m - E_a \quad (\text{A4})$$

where  $S_{\text{sm}}$  is the soil moisture of a top soil layer controlling runoff generation and actual evaporation,  $E_a$ . The contribution  $\Delta S_{\text{uz}}$  of rain and snowmelt to runoff is calculated as a function of soil moisture of the top layer,  $S_{\text{sm}}$ , using a non-linear relationship with two free parameters,  $FC$  and  $\beta$ :

$$\Delta S_{\text{uz}} = \left( \frac{S_{\text{sm}}}{FC} \right)^{\beta} \cdot (P_r + P_m) \quad (\text{A5})$$

$FC$  is the maximum soil moisture storage.  $\beta$  controls the characteristics of runoff generation and is a non-linearity parameter. If the top soil layer is saturated, i.e.  $S_{\text{sm}} = FC$ , all rainfall and snowmelt contributes to runoff. Actual evaporation,  $E_a$ , is calculated from potential evaporation,  $E_p$ , by a piecewise linear function of soil moisture of the top layer:

$$E_a = E_p \cdot \frac{S_{\text{sm}}}{LP}, \quad \text{if } S_{\text{sm}} < LP \\ E_a = E_p, \quad \text{otherwise} \quad (\text{A6})$$

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

### A.3. Response and transfer functions

The response function represents runoff routing on the hillslopes and consists of two reservoirs, representing two soil zones. The storage states of the upper and lower zones are  $S_{uz}$  and  $S_{lz}$ , respectively.  $\Delta S_{uz}$  enters the upper zone reservoir and leaves this reservoir through three paths, outflow from the reservoir with a fast storage coefficient of  $k_1$ , percolation to the lower zone with a constant percolation rate  $c_{perc}$ , and, if a threshold  $LS_{uz}$  of the storage state is exceeded, through an additional outlet with a storage coefficient of  $k_0$ . Water leaves the lower zone with a slow storage coefficient of  $k_2$ . The outflow from both reservoirs,  $Q_g$ , is then routed by a triangular transfer function, which represents runoff routing in the streams

$$B_q = B_{max} - c_{route} \cdot Q_g, \quad \text{if } (B_{max} - c_{route} \cdot Q_g) \geq 1$$

$$B_q = 1, \quad \text{otherwise}$$
(A7)

where  $B_q$  is the base of the transfer function,  $B_{max}$  is the maximum value of  $B_q$  and  $c_{route}$  is a free parameter. Four model parameters were set to constant values ( $T_s = 0^\circ\text{C}$ ,  $T_r = 2^\circ\text{C}$ ,  $T_m = 0^\circ\text{C}$ ,  $B_{max} = 10$  days) on the basis of sensitivity analyses while the remaining 11 parameters were found by calibration.

## Appendix B

### B.1. Objective function

The first component of the objective function involves the Nash and Sutcliffe (1970) coefficient of efficiency, ME

$$ME = 1 - \frac{\sum_{i=1}^n (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^n (Q_{obs,i} - \overline{Q_{obs}})^2}$$
(B1)

where  $Q_{obs,i}$  and  $Q_{sim,i}$  are observed and simulated runoff on day  $i$ , respectively, and  $\overline{Q_{obs}}$  is the mean of observed runoff over the calibration period of  $n$  days. The second component involves the volume error, VE:

$$VE = \frac{\sum_{i=1}^n Q_{sim,i} - \sum_{i=1}^n Q_{obs,i}}{\sum_{i=1}^n Q_{obs,i}}$$
(B2)

To avoid snow and moisture to accumulate without bounds over the years, we introduced a penalty,  $\epsilon_s$  once the normalised snow water equivalent of the last third of the calibration period is different by more than 30% from that of the first third

$$\epsilon_s = \frac{\left( \sum_{i=1}^{n/3} SWE_i - \sum_{i=2n/3}^n SWE_i \right)}{\sum_{i=1}^{n/3} SWE_i}, \quad \text{if } \epsilon_s > \epsilon_{min},$$

$$\epsilon_s = 0, \quad \text{otherwise}$$
(B3)

with  $\epsilon_{min} = 0.3$ . An analogous penalty  $\epsilon_{min}$  was adopted for the accumulation of soil moisture  $S_{sm}$ . The third and fourth components involve these penalties.

In calibration procedures, the parameter values are usually bounded between two limits (Duan et al., 1992) and otherwise no a priori assumptions are made about the parameters. This implies that the a priori distribution of the parameters is a uniform distribution. We believe that it is possible to make a more informed guess about the shape of the a priori distribution and introduced a fifth component which involves deviations from a priori assumptions on the parameters. We used a penalty function,  $\epsilon_p$ , based on a Beta distribution for each parameter

$$\epsilon_p = \sum_{j=1}^k \frac{f_{max,j} - f_j \left( \frac{p_j - p_{l,j}}{p_{u,j} - p_{l,j}} \right)}{f_{max,j}}$$
(B4)

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

Table A1  
A priori distribution of parameter values

Parameter	$\alpha$	$\beta$	$p_l$	$p_u$	$p_{\max}$	
DDF	Degree day factor (mm/(day °C))	2	4	0	5	1.25
SCF	Snow correction factor (–)	1.2	4	1	$0.0002z + 1.2$	
FC	Maximum soil moisture storage (mm)	1.1	1.5	0	600	100
LP/FC	Ratio of limit for potential evaporation and FC (mm/mm)	4	1.2	0	1	0.93
beta	Non-linearity parameter of runoff generation (–)	1.1	1.5	0	20	3.3
$k_0$	$k_0$ storage coefficient (days)	2	4	0	2	0.5
$LS_{uz}$	Storage capacity threshold (mm)	3	3	1	100	50
$k_1$	Fast storage coefficient (days)	2	4	2	30	9
$k_2$	Slow storage coefficient (days)	1.05	1.05	30	250	105
$c_{\text{perc}}$	Percolation rate (mm/day)	2	4	0	8	2
$c_{\text{route}}$	Routing parameter (days <sup>2</sup> /mm)	1.05	1.05	0	50	25

$\alpha$  and  $\beta$  are the parameters of the Beta function (Eq. (B5)),  $p_l$  and  $p_u$  are the lower and upper bounds of the parameter space and  $p_{\max}$  is the parameter value at which the Beta distribution is at a maximum (Eq. (B4)).  $z$  is the mean catchment elevation in m a.s.l.

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

$$f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1},$$

for  $0 < x < 1$ ,  $\alpha > 0$ ,  $\beta > 0$  with  $B(\alpha, \beta)$  (B5)

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

We assumed values of  $\alpha$ ,  $\beta$ ,  $p_l$  and  $p_u$  for each parameter  $j$  based on our own assessment of the hydrologic characteristics of the study region as well as on literature values (Bergström, 1976; Seibert, 1997) (Table A1). The  $\alpha$ ,  $\beta$ ,  $p_l$  and  $p_u$  values were the same for all catchments with the exception of the upper bound  $p_u$  of  $SCF$  which was assumed to be a linear function of elevation (Table A1) as the catch deficit of snowfall is often higher in mountainous catchments than it is in low elevation catchments. The entire objective function now consists of the following parts

$$Z = w_1(1 - \text{ME}) + w_2|\text{VE}| + w_3(1 + \varepsilon_s)^2 + w_4(1 + \varepsilon_m)^2 + w_5\varepsilon_p \quad (\text{B6})$$

where the weights  $w_i$  were set to  $w_1 = 1$ ,  $w_2 = 0.1$ ,  $w_3 = 0.01$ ,  $w_4 = 0.01$  and  $w_5 = 0.5$ . The values of  $w_1$  and  $w_2$  follow the recommendations of Lindström et al. (1997). The third and fourth terms have been taken as squared terms using small weights in order to provide very small penalties for small deviations from the prescribed trend but exceedingly larger penalties once the deviations get larger. The value of  $w_5$  we obtained from sensitivity analyses by examining the distribution functions of the calibrated parameter values for different values of  $w_5$ . Large values of  $w_5$  resulted in calibrated parameters clustering around  $p_{\max}$  while small values resulted in similar distributions as when using  $w_5 = 0$ . A value of  $w_5 = 0.5$  provided balanced distribution functions between these two extremes for all parameters.

## References

- Beldring, S., Roald, L.A. and Voksø, A., 2002. Avrenningskart for Norge (Runoff map for Norway, in Norwegian). Norwegian Water and Energy Directorate, Report No. 2, Oslo, Norway.
- Bergström, S., 1976. Development and application of a conceptual runoff model for Scandinavian catchments. Dept. Water Resour. Engng, Lund Inst. Technol./Univ. Lund, Bull. Ser. A 52, 134.
- Bergström, S., 1991. Principles and confidence in hydrological modelling. Nordic Hydrol. 22, 123–136.
- Beven, K.J., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298.

- Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling—a review. *Hydrol. Process.* 9, 251–290.
- Chiew, F., McMahon, T., 1994. Application of the daily rainfall-runoff model MODHYDROLOG to 28 Australian catchments. *J. Hydrol.* 153, 383–416.
- Deutsch, C.V., Journel, A.G., 1997. *Geostatistical Software Library and User's Guide*. Oxford University Press, Oxford, 384 pp.
- Duan, Q., Sorooshian, S., Gupta, V.K., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* 28, 1015–1031.
- DVWK, 1996. *Ermittlung der Verdunstung von Land- und Wasserflächen*, DVWK-Merkblätter, Heft 238, Bonn.
- Ecker, R., Kallianay, R., Steinnocher, K., 1995. Fernerkundungsdaten für die Planung eines Mobilfunknetzes. *Österr. Zeitschr. f. Vermessung und Geoinformation* 83, 14–25.
- Geologische Bundesanstalt, 1998. *Metallogenetische Karte 1:500.000 auf CD-ROM*. GBA, FA-ADV, Wien.
- Goodrich, D.C., Lane, L.J., Shillito, R.M., Miller, S.N., Syed, K.H., Woolhiser, D.A., 1997. Linearity of basin response as a function of scale in a semiarid watershed. *Water Resour. Res.* 33 (12), 2951–2965.
- Gottschalk, L., 2002. Advances in observational hydrology—field experiments and modelling. In: Takeuchi, K., (Ed.), *Proceedings of Workshop on the Prediction of Ungaged Basins (PUBs) held at 28–29 March 2002 at the Yamanashi University, Kofu, Japan*, International Association of Hydrological Sciences.
- Grayson, R., Blöschl, G. (Eds.), 2000. *Spatial Patterns in Catchment Hydrology: Observations and Modelling*, Cambridge University Press, Cambridge, p. 404.
- HZB, 2000. *Hydrologisches Jahrbuch von Österreich 1997*. Hydrographisches Zentralbüro im BMLF, Band 105, Wien.
- IH, 1999. *Flood Estimation Handbook*. Institute of Hydrology, Wallingford.
- Kite, G., 2000. Using a basin-scale hydrological model to estimate crop transpiration and soil evaporation. *J. Hydrol.* 229, 59–69.
- Klemeš, V., 1986. Operational testing of hydrological simulation models. *Hydrol. Sci. J.* 31, 13–24.
- Lindström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. *J. Hydrol.* 201, 272–288.
- Merz, R., Blöschl, G., 2003. *A process typology of regional floods*. *Water Resour. Res.* in press.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I. A discussion of principles. *J. Hydrol.* 10, 282–290.
- ÖBG, 2001. *Bodenaufnahmesysteme in Österreich*. *Mitteilungen der Österreichischen Bodenkundlichen Gesellschaft*. Heft 62.
- Peel, M.C., Chiew, F.H.S., Western, A.W., McMahon, T.A., 2000. Extension of unimpaired monthly streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments. Report prepared for the National Land and Water Resources Audit. In *Australian Natural Resources Atlas website*, 37 pp.
- Perrin, C., Michel, C., Andréassian, V., 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.
- Piock-Ellena, U., Blöschl, G., 1998. *Abflusstypen-Klassifizierung als Basis für die Regionalisierung von Hochwässern*. Technical report, Austrian Academy of Sciences (HÖ-12/97) Institute of Hydraulics, Vienna University of Technology.
- Post, D.A., Jakeman, A.J., 1999. Relationships between catchment attributes and hydrological response characteristics in small Australian mountain ash catchments. *Hydrol. Process.* 10 (6), 877–892.
- Sefton, C.E.M., Howarth, S.M., 1998. Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales. *J. Hydrol.* 211, 1–16.
- Seibert, J., 1997. Estimation of parameter uncertainty in the HBV model. *Nordic Hydrol.* 28, 247–262.
- Seibert, J., 1999. Regionalisation of parameters for a conceptual rainfall-runoff model. *Agric. Forest Meteorol.* 98/99, 279–293.
- Sevruk, B., Paulais, M., Roulet, Y.-A., 1998. Correction of precipitation measurement using fresh snow as reference. *Proceedings of the WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation, TECO-98, WMO/TD-No. 877*, Geneva, Switzerland, pp. 349–352.
- Uhlenbrook, St., Seibert, J., Leibundgut, Ch., Rodhe, A., 1999. Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. *Hydrol. Sci. J.* 44 (5), 779–797.
- Vandewiele, G.L., Elias, A., 1995. Monthly water balance of ungauged catchments obtained by geographical regionalisation. *J. Hydrol.* 170, 277–291.
- Vandewiele, G.L., Chong-Yu, X., Ni, L.W., 1994. Methodology and comparative study of monthly water balance models in Belgium, China and Burma. *J. Hydrol.* 134, 315–347.
- WMO, 1986. *Intercomparison of models of snowmelt runoff*. WMO Operational Hydrology Report No. 23, WMO Publication No 646, Geneva.