ELSEVIER

Flood frequency regionalisation—spatial proximity vs. catchment attributes

R. Merz*, G. Blöschl

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

Abstract

We examine the predictive performance of various flood regionalisation methods for the ungauged catchment case, based on a jack-knifing comparison of locally estimated and regionalised flood quantiles for 575 Austrian catchments, 122 of which have a record length of 40 years or more. The main result is that spatial proximity is a significantly better predictor of regional flood frequencies than are catchment attributes. A method that combines spatial proximity and catchment attributes yields the best predictive performance. This is a novel method proposed in this paper which is based on kriging and takes differences in the length of the flood records into account. It is shown that short flood records contain valuable information which can be exploited by the proposed method. A method that uses only spatial proximity performs second best. The methods that only use catchment attributes perform significantly poorer than those based on spatial proximity. These are a variant of the Region Of Influence (ROI) approach, applied in an automatic mode, and multiple regressions. It is suggested that better predictive variables and similarity measures need to be found to make these methods more useful. A stratified analysis suggests that in wet catchments all regionalisation methods perform better than they do in dry catchments.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Flood frequency; Regionalisation; Regional estimation; Kriging; Regression

1. Introduction

Flood peak estimates associated with a given exceedance probability or return period are needed for many engineering problems. As flood estimates are often required for catchments without streamflow measurements, they are usually obtained by some sort of regional transposition, or regionalisation, from gauged catchments to the site of interest (Cunnane, 1988; Bobée and Rasmussen, 1995; Hosking and Wallis, 1997). There are three main issues in regionalisation: what information is best transferred, what is the method to be used and from which catchments is the information to be taken for deriving the estimates at the site of interest. 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. The traditional measure of similarity is spatial proximity with the rationale that catchments that are close to

* Corresponding author. Fax: +43 1 588 01 233 99.
  E-mail address: merz@hydro.tuwien.ac.at (R. Merz).

each other will also behave similarly in terms of their flood frequency response as climate and catchment conditions will only vary smoothly in space. In the classical Index flood approach (Dalrymple, 1960) the domain is subdivided into regions and within each region the flood frequency response is assumed to be similar apart from a scaling factor, the index flood. There are alternative methods that use spatial proximity. One of them is based on geostatistical concepts (Merz and Blöschl, 1999; Merz et al., 2000a). The advantage of the geostatistical approach is that it provides a best linear unbiased estimator, but the disadvantage is that the spatial structure imposed by weather divides and geologic divides is more difficult to exploit than in the Index flood approach.

The analysis of observed flood frequency behaviour often reveals small scale variability but catchments that are far apart may still be hydrologically similar (e.g. Pilgrim, 1983), so alternative measures of similarity have been proposed. These measures are often based on catchment attributes. Streamflow is not used in these similarity measures to make them applicable to the ungauged catchment case. The catchment attributes are thought of as surrogates of the hydrological processes within a catchment. The rationale of this approach is that catchments with similar attributes are also likely to exhibit similar flood generation processes and hence may also behave similarly in terms of their flood frequency response (Acreman and Sinclair, 1986). Catchment attributes include catchment size, land-use, geology, elevation, soil characteristics as well as climate variables such as mean annual precipitation. The catchment attributes can be used in various ways. These include multiple regressions between flood quantiles (or flood moments) and catchment attributes (Tasker, 1987), and the pooling of catchments into a homogeneous group and the subsequent use of averages within each pooling group (IH, 1999). The latter approach is also known as the Region Of Influence (ROI) approach (Burn, 1990; Pfaundler, 2001) where each site has its own pooling group which is determined by the similarity of catchment attributes.

While each of these two genres of approaches, i.e. those based on spatial proximity and those based on catchment attributes, have their defendable rationales, we are unaware of any comprehensive comparison of these two general approaches in terms of their predictive power. For any practical application, the interest resides in how well flood quantiles, such as the 100-year flood, can be estimated for a given catchment. The aim of this paper therefore is to compare the predictive performance of various flood regionalisation methods pertaining to one of the two groups, as well as combinations thereof. We focus on the ungauged catchment case for which flood frequency regionalisation is particularly difficult. We use a comprehensive data set of small to medium sized catchments in Austria, where the flood regime is not or only slightly modified by anthropogenic effects. The main strategy to assess the predictive performance is a jack-knifing approach where, in a first step, flood frequencies are estimated for gauged catchments from regional information without using local flood information and, in the second step, these regional estimates are compared to the local flood frequency estimates. This jack-knifing comparison then gives us a reliable measure of how well each of the methods performs for the ungauged catchment case. Some of the flood frequency regionalisation methods used in practice use expert knowledge, for example in the decision which catchments to pool into a homogeneous group. While it is clear that local expert knowledge will always be valuable, in this paper we have chosen to use an objective, reproducible comparison that is solely based on the hard data available in the data set.

As a representative of the genre of approaches that are based on spatial proximity we have chosen kriging, for which we propose an extension that exploits information on record lengths. As representatives of the genre of approaches that are based on catchment attributes, we have chosen multiple regressions between flood peak moments and catchment attributes as well as a variant of the Region Of Influence (ROI) approach. As combined approaches that use both proximity and catchment attributes we have chosen external drift kriging, georegression, and another variant of the ROI approach. In addition to the main comparisons between the methods, we analyse possible controls on the relative predictive performance of the methods. These include the presence of nested catchments, record length, parameter estimation method, and climate (wet vs. dry catchments).

2. Data

The catchments used in this paper are all located in Austria. The physiography ranges from the lowlands in the east of the country, with mean catchment elevations of less than 200 m a.s.l., up to the high alpine catchments in the west of the country with mean catchment elevations of more than 2500 m a.s.l.. Mean annual precipitation ranges from less than 400 mm/year in the east to more than 3000 mm/year in the west, where orographic effects tend to enhance precipitation. The topography of Austria and the boundaries of the gauged catchments are shown in Fig. 1.

The analyses in this paper use (a) observed maximum annual flood peaks of catchments ranging in area from 10 to 1000 km^2, (b) the coordinates of the centroids of these catchments, and (c) catchment attributes. In a first step, the maximum annual flood peaks were screened for data errors, and stations affected by reservoir operation and other anthropogenic effects were removed from the data set (Blöschl et al., 2000). This screening resulted in flood series for 575 catchments with observation periods from 5 to 44 years. To reduce the effect catchment area may have on the regional patterns of flood frequency, the specific flood discharges were standardised to specific

discharges of a hypothetical catchment area of 100 km^2 according to

$$Q = Q_A A^{0.25} 100^{-0.25} \tag{1}$$

where Q is the specific discharge for a hypothetical 100 km^2 catchment and Q_A is the observed specific discharge of a catchment of area A (km^2). The exponent of -0.25 was found by a regression analysis between observed mean annual floods and catchment area for the Austrian catchments. The coordinates of the centroids of the 575 catchments were derived from digital catchment boundaries. Two data sets were used. Most of the catchment boundaries were taken from a digital database of catchment boundaries digitised from the Austrian 1:50000 scale map (ÖK 50) (Behr et al., 1998). The remaining boundaries were derived from a digital elevation model (Rieger, 1999). All catchment boundaries were checked manually using the ÖK 50, so the coordinates of the centroids should be subject to minimum error. A number of catchment attributes were used. Average catchment elevation and average topographic slope were derived from the digital elevation model. Mean annual precipitation (MAP) and mean maximum annual daily precipitation (MADP) (i.e. the long term mean of the series) were spatially interpolated using observed precipitation from more than 1000
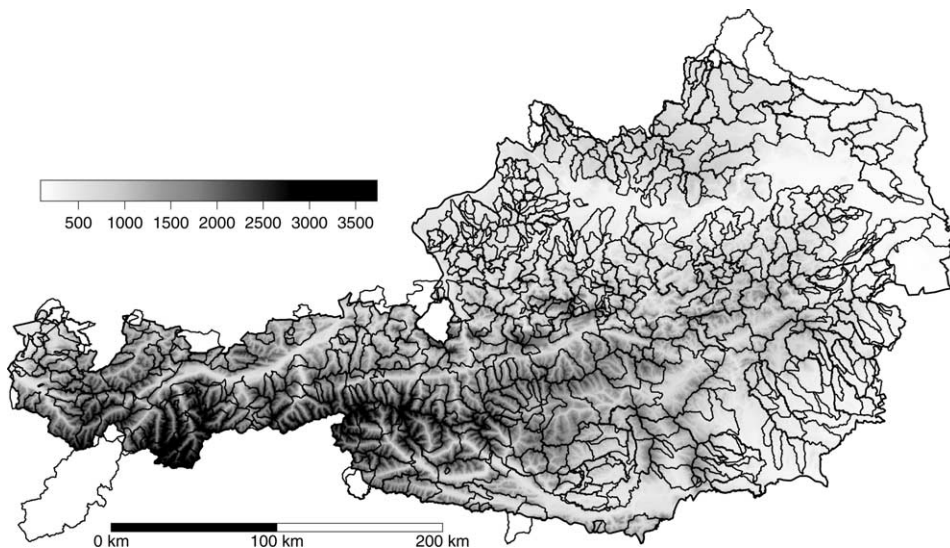

Fig. 1. Topography (m a.s.l.) of Austria and boundaries of the gauged catchments used in this paper.

raingauges with record lengths 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 (Behr et al., 1998) 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. Methods

3.1. Flood moments

For all regionalisation methods we used the first three moments of the annual flood peak series, specifically the first three product moments (mean annual flood (MAF), coefficient of variation (CV) and coefficient of skewness (CS))

$$
\begin{aligned}
\text{MAF} &= \frac{1}{m} \sum_{j=1}^{m} Q_j, \\
S^2 &= \frac{1}{m-1} \sum_{j=1}^{m} (Q_j - \text{MAF})^2, \\
\text{CV} &= \frac{S}{\text{MAF}}, \\
\text{CS} &= \frac{m \cdot \sum_{j=1}^{m} (Q_j - \text{MAF})^3}{(m-1)(m-2)S^3}
\end{aligned} \tag{2}
$$

where Q_j is the specific flood discharge of a hypothetical 100 km^2 catchment of year j, m is the number of years in the flood sample. The first three L-moments $(\lambda_1, \lambda_2, \lambda_3)$ were calculated according to

Hosking and Wallis (1997, p. 26)

$$
\begin{aligned}
b_k &= \frac{1}{m} \binom{m-1}{k}^{-1} \sum_{j=k+1}^{m} \binom{j-1}{k} Q_{j:m}, \\
\lambda_1 &= b_0, \quad \lambda_2 = 2b_1 - b_0, \\
\lambda_3 &= 6b_2 - 6b_1 + b_0
\end{aligned} \tag{3}
$$

where $Q_{j:m}$ is rank j of the ordered sample $Q_{1:m} \leq Q_{2:m} \leq \cdots \leq Q_{m:m}$. The flood peaks have all been standardised to a catchment area of 100 km^2 but this standardisation (Eq. (1)) will only change the first product moment (Eq. (2)) and the L-moments (Eq. (3)) as the second and third product moments are dimensionless. In all the regionalisation methods we used the logarithms of the mean annual flood and λ_1, i.e. $Z = \log(\text{MAF})$ and $Z = \log(\lambda_1)$, to reduce skewness, while we introduced the other moments without transformation.

3.2. Geostatistics (Kriging)

In the regionalisation approach based on kriging, we interpolated each moment independently by Ordinary Kriging In Ordinary Kriging the estimated value $\hat{Z}(\mathbf{x}_0)$ of a moment is a weighted linear combination of the moments of n gauged catchments Z_i

$$
\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i Z_i \tag{4}
$$

The weights λ_i are determined such as to minimize the estimation variance, while ensuring the unbiasedness of the estimator which leads to a system of linear equations called the kriging system. The only information of hydrological similarity required in the kriging system are the semivariogram values for different lags.

The moments of the flood peak series for each catchment are associated with some uncertainty or estimation error due to a relatively short sample size. This estimation error decreases with the size of the sample (number of years of the flood record) and increases with the order of the moment. The first moment (the mean) can be estimated from a flood peak record with relatively little error while the third moment (the skewness) is always associated with substantial error. We propose in this paper that this
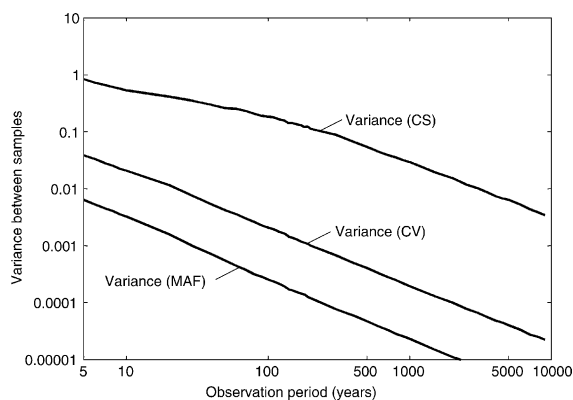
Fig. 2. Error variance due to short record lengths for three product moments estimated by a Monte Carlo analysis, as a function of record length.

local estimation error can be thought of as a kind of measurement error in the spatial interpolation and can hence be accommodated in the kriging system. If the errors associated with each flood moment Z_i are non-systematic, uncorrelated with each other, uncorrelated with the moment Z and have a known variance σ_i^2, the kriging system can be extended to account for these errors (de Marsily, 1986, p. 300)

$$\sum_{k=1}^{n} \lambda_k \gamma(\mathbf{x}_i - \mathbf{x}_k) - \lambda_i \sigma_i^2 + \mu = \gamma(\mathbf{x}_i - \mathbf{x}_0)$$
$$i = 1, \ldots, n \tag{5}$$

$$\sum_{i=1}^{n} \lambda_i = 1$$

where μ is the Lagrange parameter, the $\mathbf{x}_i$ and $\mathbf{x}_k$ are the coordinates of the catchment centroids, and $\gamma(\mathbf{x}_i - \mathbf{x}_k)$ is the true semivariogram (without representing errors) of the moment Z for the lag of catchment centroids $\mathbf{x}_i$ and $\mathbf{x}_k$. Note that each flood peak moment in each catchment i can have a different error. It is therefore possible to account for different record lengths in different catchments.

We term the proposed method Kriging with Uncertain Data (KUD). We used this method as an alternative geostatistical regionalisation procedure to Ordinary Kriging. Note that the ordinary kriging (OK) system is identical to Eq. (5) with $\sigma_i^2 = 0$.

We estimated the error variance due to short record lengths in a Monte Carlo analysis by drawing samples of a given size from a known distribution and estimating the variance of the moments between different samples (Fig. 2). The distribution was assumed to be Gumbel distributed with MAF=0.3 and CV=0.5. As a next step, we parameterised this error variance as a function of record length and the order of the moment by a power law of the form

$$\sigma_k^2(m) = a \cdot m^{-b} \tag{6}$$

where m is the number of years of record in a catchment and k is the order of the moment. We estimated the parameters a and b from a best fit to the variances as shown in Fig. 2, both for traditional product moments and L-moments (Table 1). As the L-moments put more weight on samples near the median than on extreme values, the variances between samples are smaller than those for the product moments, so the estimation of the moments is more robust, particularly for short observation periods. However, L-moments are based on the assumption that the observed values near the median have more predictive power for extreme situations than observed extreme values which may not always be justified from a hydrological perspective.

We estimated the variograms from the flood moments in the gauged catchments using the catchment centroids for determining the lags. We then fitted exponential variograms (Eq. (7)) to the data-based (experimental) variograms (also see Merz et al., 2000a)

$$\gamma = c(1 - \exp(-h/d)) \tag{7}$$

d was estimated as 20 km for all moments which means that the practical range is 60 km. Based on

Table 1
Parameters for the error variance σ_i^2 as a result of short record lengths as used in Eq. (6)

Moment	Log(MAF)	CV	CS	Log(λ_1)	λ_2	λ_3
A	1.383	1.187	1.992	1.383	0.012	0.020
B	−1.090	−0.959	−0.537	−1.090	−1.184	−1.714

these variograms and the error variance in Eq. (6) the first three product moments (log(MAF), CV, CS) and the first three *L*-moments (log(λ_1, λ_2, λ_3)) (Hosking and Wallis, 1997) were then interpolated independently. Three scenarios were examined to analyse the effect of record lengths on the geostatistical estimates. First, the flood moments were regionalised by ordinary kriging using all stations, irrespective of their record length (termed OK). In the second scenario, only stations with more than 20 years of observations were used in the ordinary kriging system to estimate the flood moments (termed OK_LS). In a third scenario, all stations were used for the regionalisation and the record lengths were taken into account by the Kriging with Uncertain Data approach (termed KUD). The flood quantiles were then estimated for given return periods using the Generalised Extreme Value (GEV) distribution for each of the three scenarios and each of the two moment types giving a total of six estimates for each catchment. These are estimates for which only regional information from neighbouring catchments have been used. These were then compared to the flood quantiles estimated from the local flood data. A comparison of various distribution functions in Merz et al. (2000a) suggests that the GEV distribution is suitable for the Austrian data set.

3.3. Multiple regression

For the regionalisation approach based on multiple regression we used, again, the first three moments of the annual flood peak series (both product moments and *L*-moments, Eqs. (2) and (3)) and related each of them to the catchment attributes. A linear multiple regression model with three predictive variables was used

$$\hat{Z}(\mathbf{x}_0) = a + b \cdot Y_1(\mathbf{x}_0) + c \cdot Y_2(\mathbf{x}_0) + d \cdot Y_3(\mathbf{x}_0) \qquad (8)$$

where $\hat{Z}(\mathbf{x}_0)$ is the flood sample moment (or a transformation thereof) to be estimated, $Y_1(\mathbf{x}_0)$, $Y_2(\mathbf{x}_0)$, $Y_3(\mathbf{x}_0)$ are the catchment attributes and a, b, c, d are the regression coefficients. Results of Merz et al. (2000b) suggested that the additional explained variance of using more than three variables was small.

The choice of catchment attributes as variables in the regression analysis in this paper has been guided by general knowledge on the relationship between

floods, climate and physiography as well as previous regression studies (e.g. Tasker, 1987; IH, 1999; Castellarin et al., 2001). We examined catchment area, catchment average elevation, river network density, the catchment average of mean annual precipitation, the FARL lake index, catchment average topographic slope, catchment average maximum annual daily precipitation, portion of catchment area with porous aquifers, two land use classes (portion of forest and glacier), two geologic units (portion of Tertiary + Quaternary and Calcareous Alps) and three soil types (portion of Austroalpin crystalline, Rendzina, Cambisol). The ordinary least squares method was used to estimate the regression coefficients. To reduce possible biases due to highly skewed explanatory data, all catchment attributes were transformed to be standard normally distributed. Only those catchments with distances less than 100 km to the site of interest were included in the regression system which resulted in regression systems with about 100–200 stations.

One of the most serious problems encountered with multiple regression is multicollinearity, i.e. when at least one of the explanatory variables is highly correlated with another explanatory variable or with some linear combination of other explanatory variables. If multicollinearity is present, the regression coefficient can be highly unstable and unreliable. A diagnostic for multicollinearity is the variance inflation factor (VIF) (Hirsch et al., 1992). For an explanatory variable *j*

$$\text{VIF}_j = \frac{1}{1 - r_j^2} \qquad (9)$$

where r_j^2 is the multiple correlation coefficient squared from a regression of variable *j* with all other explanatory variables. For the ideal case of orthogonal variables, i.e. no collinearity, $\text{VIF}_j = 1$, while for $\text{VIF}_j > 10$ the regression is no longer reliable (Hirsch et al., 1992).

To avoid the problem of multicollinearity, in this study, the multiple regressions only use three explanatory variables out of a set of 15 available catchment attributes. As it is not obvious which are the best explanatory variables we examined three scenarios. In the first scenario (termed MR_AP), three attributes, i.e. catchment elevation, river

network density and mean annual precipitation, have been selected a priori. These are the attributes one would expect to explain most of the variance of regional flood frequency based on experience in the literature (e.g. Tasker, 1987; IH, 1999). In the second scenario, out of the 15 available attributes, the set of three attributes with the largest multiple correlation coefficient for each station and each flood moment separately was used (termed MR_BEST). The rationale of this choice is that a high correlation coefficient may also be a good indicator of the predictive power of the attributes. However, if $VIF_j > 10$ this set of attributes was rejected and the scheme proceeded to the second best correlation. In the third scenario, for each flood moment, the set of attributes of the previous scenario that was used most in all the catchments was adopted (termed MR_MOST). The rationale of this scenario is that if the controls of the catchment attributes on flood frequency response are physically realistic they should perhaps be the same in the entire domain of Austria. The flood quantiles were then estimated for given return periods using the GEV distribution for each of the three scenarios and each of the two moment types which we then compared to the flood quantiles estimated from the local flood data.

3.4. External drift kriging and georegression

One could argue that spatial proximity and catchment attributes are not mutually exclusive pieces of information, so a combination of kriging and catchment attributes may provide more information than any of the individual regionalisation approaches alone. We used two methods of combining kriging with catchment attributes, external drift kriging and georegression. In external drift kriging, local regression with an auxiliary variable is directly implemented into the kriging system and the auxiliary variable is assumed to be error free and perfectly related to the primary variable (Deutsch and Journel, 1997, p. 67). We used mean annual precipitation (MAP) as the auxiliary variable as it is thought to be a surrogate of the rainfall input and the wetness state of the catchment. Also, MAP tends to be one of the best predictive catchment attributes for regional flood

frequency (IH, 1999). Using external drift kriging, all three flood moments were interpolated independently.

As an alternative we examined georegression which is a two step procedure. In a first step, flood moments were related to catchment attributes by a multiple regression with three catchment attributes

$$Z'(\mathbf{x}_i) = a + b \cdot Y_1(\mathbf{x}_i) + c \cdot Y_2(\mathbf{x}_i) + d \cdot Y_3(\mathbf{x}_i) \qquad (10)$$

where $Z'(\mathbf{x}_i)$ are the flood moments (or their logarithmic transformation in the case of the first moments), $Y(\mathbf{x}_i)$ are the catchment attributes and a, b, c, d are the regression coefficients. Only those catchments with distances less than 100 km to the site of interest were included in the regression system. In the second step, the corresponding residuals were spatially interpolated using ordinary kriging

$$\hat{Z}(\mathbf{x}_0) - Z'(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i [Z(\mathbf{x}_i) - Z'(\mathbf{x}_i)] \qquad (11)$$

where $\hat{Z}(\mathbf{x}_0)$ is the estimate of the flood moments (or a transformation thereof) of the site of interest, $Z(\mathbf{x}_i)$ are the observed flood moments (or a transformation thereof) of catchment i with coordinates of the centroids $\mathbf{x}_i$ and λ_i are the kriging weights. A priori it is not clear which attributes will improve the geostatistical regionalisation most, so we examined all possible combinations of catchment attributes and selected those three variables for each flood moment that exhibited the largest correlation coefficients. We then analysed three scenarios. In a first scenario, the residuals were interpolated by OK with zero nugget (termed GEOREG). To account for differences in the degree of correlation between the flood moments and the catchment attributes, kriging with uncertain data (KUD) was used to interpolate the residuals in the second scenario (termed GEOREG_KUD). The variance of the error in the kriging system (Eq. (5)) was set to

$$\sigma_k^2 = a \cdot m^{-b} \cdot (1 - r^2) \qquad (12)$$

where m is the number of years of record for a catchment, k is the order of the moment, a and b are the coefficients as given in Table 1, and r^2 is the multiple correlation coefficient squared of the regression system. In Eq. (12) we assume that r^2 is a measure of how well the flood moments can be predicted by the catchment attributes.

A preliminary analysis of flood moments indicated that the second and third flood moments (CV, CS) are not as well correlated with catchment attributes as the first moment (MAF). We therefore examined a third scenario (termed GEOREG_KUD/KUD) where we regionalised the mean annual floods by georegression as described above with the residuals interpolated by KUD, but the coefficient of variation and the coefficient of skewness we interpolated simply by KUD without using catchment attributes. As in the previous methods, we calculated flood quantiles from the regionalised flood moments by using a GEV distribution and compared them to the locally estimated values.

3.5. Region of Influence approach

The Region Of Influence (ROI) approach (Burn, 1990) is based on pooling stations into groups. Each site of interest (i.e. catchment for which flood quantiles are to be estimated) has its own pooling group, and this group is not necessarily spatially contiguous. The starting point of the ROI approach is the selection of a distance measure defining the hydrological similarity of catchments. The distance measure D_{i0} between station i and station 0 usually contains catchment attributes and has been defined as

$$D_{i0} = \left[\sum_{l=1}^{L} W_l (Y_l(\mathbf{x}_i) - Y_l(\mathbf{x}_0))^2 \right]^{1/2} \qquad (13)$$

where W_l is the weight of the relative importance of attribute l out of L and $Y_l(\mathbf{x}_i)$ is the value of attribute l for station i. Those catchments that are most similar to the site of interest in terms of D_{i0} are included in the group. We chose the number of catchments in each pooling group such that the sum of the observed years of all stations in the pooling group was about five times the return period of interest (IH, 1999, 3/p. 169). We chose 100 years as the return period of interest here, which means that, with an average record length of 30 years, 10–20 stations are pooled. The homogeneity of the pooling group was checked by the H test of Hosking and Wallis (1997) which is based on the hypothesis of homogeneity that the local frequency distributions are the same except for a site-specific scale factor. If a region is found to be very heterogeneous ($H > 4$) (IH, 1999, 3/p. 163) in this

study, the catchment with the largest distance measure to the mean of the group was removed from the pooling group and an alternative catchment was included. However, not more than three stations were removed by this procedure as in some cases a heterogeneous group may still contain valuable information (IH, 1999, 3/p. 172). In addition, the discordancy was examined and if a catchment exceeded the threshold given in Hosking and Wallis (1997, Table 3.1) the catchment was replaced by an alternative catchment. Once the pooling group was determined we calculated the regional averaged second and third moments for the group, weighted by the record length and a similarity ranking factor (IH, 1999, 3/p. 182), and assigned them to the site of interest. In most applications of the ROI approach (e.g. Zrinji and Burn, 1994) the first moment, or index flood, is calculated by multiple regressions and combined with a non-dimensional flood frequency curve (i.e. the growth curve) estimated by the ROI method. In this paper, we used multiple regressions to estimate the first moment in the same way as in the multiple regression approach described above. We then calculated the quantiles by combining the first moment so estimated and the second and third moments from the ROI method using the GEV distribution. The second and third moments are representative of the growth curve.

The selection of the attributes to be used in the distance measure and the determination of their weights is usually based on a screening process to identify their relative merits in terms of hydrological similarity (Burn, 1990; IH, 1999). In this paper, we analysed three scenarios. In the first scenario (termed ROI_BEST), out of the 15 available attributes, the set of three attributes with the highest multiple correlation coefficient between attributes and the second flood moment for each station was used. This is the same set as used in the multiple regression approach termed (MR_BEST) and is based on the assumption that the correlation coefficient is a meaningful indicator of hydrological similarity. The weights W_l were all set to unity, because the attributes have been standard normally transformed. In the second scenario, we used geographical distance alone in the distance measure (termed ROI_DIST). In a third scenario we combined catchment attributes used in ROI_BEST and geographical distance into a scenario

termed ROI_BEST + DIST. The weights W_l were set to 0.1, 0.1, 0.1 and 0.7 for the three attributes and geographical distance, respectively. These weights were found to give the lowest predictive errors in a preliminary analysis.

While, sometimes, selection of the pooling group in the ROI approach is supported by expert judgement (e.g. IH, 1999), in this paper, we have chosen to use a reproducible comparison that is solely based on the hard data available in the data set. One would expect that local expert knowledge will improve the predictive performance of the ROI approach, but this is difficult to quantify in an objective way.

3.6. Analysis of predictive performance

The aim of the study is to assess the predictive performance of various regionalisation methods of estimating flood quantiles in small to medium sized ungauged catchments. This assessment uses the jack-knifing method which simulates the case of ungauged catchments. In the jack-knifing approach, one gauged catchment is treated as ungauged. The flood quantiles for this catchment are then estimated based solely on the flood data in other catchments. In the second step, the flood quantiles so estimated are compared with the quantiles estimated from the local flood data. The difference between regional and local estimates then represents the error that is introduced by the regionalisation. The difference also includes the error of the local estimates. For small return periods this error will be small, so the difference represents the regionalisation error alone, while for large return periods both error sources are likely to be important. As the local estimates are always calculated by the same method (GEV, parameter estimation by the moment method), the local error will not change between the scenarios considered here. We therefore believe that the relative performance of the regionalisation methods can be very well assessed by this jack-knifing approach. The jack-knifing procedure is repeated for each catchment in turn, which gives an error estimate for all catchments.

The regionalisation error consists of a systematic component, or bias, and a random error component. The bias is a measure of whether a regionalisation method tends to overestimate or underestimate flood quantiles in all the catchments considered.

Non-negligible bias is an indication of poor model structure or inappropriate assumptions. In practical applications, biases, if known, can be removed from the estimates. The random error is a measure of the scatter of the regionalised values about the local values. Random errors are related to how much information a method can extract from the data. They cannot be removed from the estimates. We used the normalised mean error (nme) as a measure of the bias

$$\text{nme} = \frac{\sum_{i=1}^{n}(Q_i^{\text{reg}} - Q_i^{\text{loc}})}{\sum_{i=1}^{n} Q_i^{\text{loc}}} \qquad (14)$$

where Q_i^{loc} is the local flood quantile of station i as in Eq. (1), Q_i^{reg} is the regionalised quantile of station i out of n stations. nme can be positive and negative, and for a perfect regionalisation method nme = 0. We used the normalised standard deviation error (nsdve) as a measure of the random error:

$$\text{nsdve} = \frac{\sum_{j=1}^{n}\left\{(Q_i^{\text{reg}} - Q_i^{\text{loc}}) - \frac{1}{n}\sum_{i=1}^{n}(Q_i^{\text{reg}} - Q_i^{\text{loc}})\right\}^2}{\sum_{i=1}^{n} Q_i^{\text{loc}}} \qquad (15)$$

nsdve is always non-negative and for a perfect regionalisation method nsdve = 0. We used the root mean square error rmse) as a measure of the total regionalisation error:

$$\text{rmse} = \sqrt{\text{nme}^2 + \text{nsdve}^2} \qquad (16)$$

rmse is always non-negative and for a perfect regionalisation method rmse = 0.

In calculating the error measures only catchments with flood records longer than 40 years have been taken into account to reduce the uncertainty associated with the local flood frequency estimation. These are 122 catchments, so $m = 122$ in Eqs. (14) and (15). The three error measures were calculated for return periods from 2.3 to 500 years.

4. Results

4.1. Spatial proximity—Kriging and effect of record length

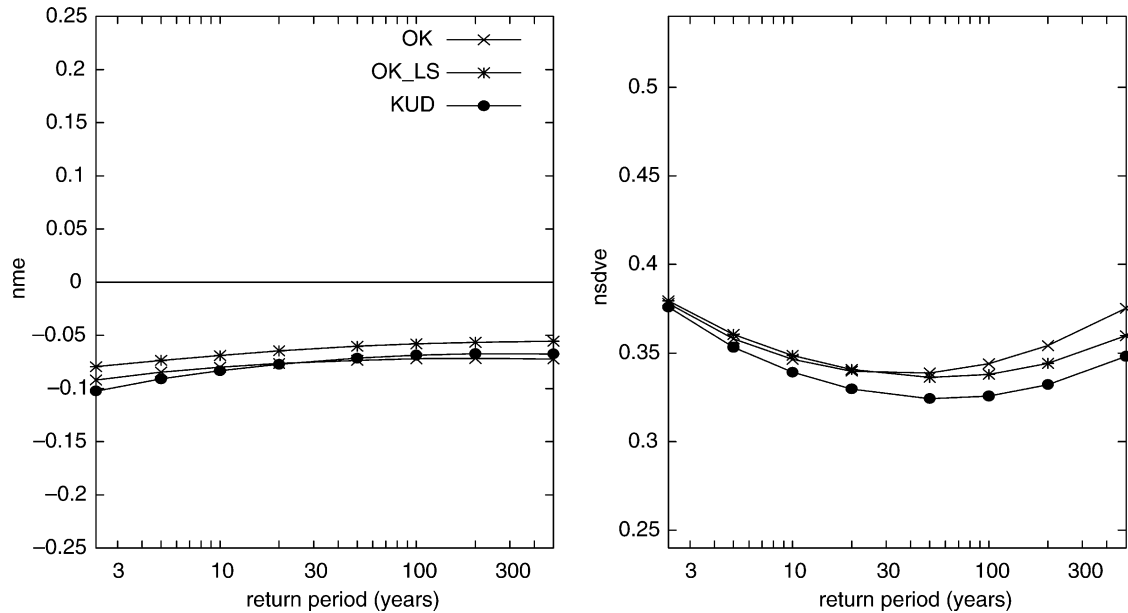In Fig. 3 the nme (bias) and nsdve (random error) of the kriging approaches have been plotted versus

Fig. 3. Bias (left) and random error (right) of flood quantile regionalisation for the Kriging approaches. Crosses (OK): Ordinary Kriging; Asterisks (OK_LS): Ordinary Kriging using only catchment with more than 20 years of record for the regionalisation; full circles (KUD): Kriging where different record lengths are taken into account. The GEV distribution and product moments are used.

the return period. Crosses (OK) refer to ordinary kriging using catchments with any record length for the regionalisation, asterisks (OK_LS) refer to ordinary kriging using only catchments with more than 20 years of record for the regionalisation and the full circles (KUD) refer to kriging where differences in the record length have been taken into account in the regionalisation. The bias of all methods is small (between −0.05 and −0.11) while the random error is large (between 0.32 and 0.38). The contribution of the bias to the total error (rmse) is only about 3% (Eq. (16)), implying that the bias is almost negligible. Kriging is an unbiased estimator, so small biases would be expected.

The random error decreases with increasing return period from 2.3 to 40 years. This suggests that the second and the third flood moments can be regionalised more accurately than the first moment. Given that kriging builds on the spatial correlations, this implies that the second and third moments are better correlated in space or, in other words, vary more smoothly in space than does the first moment. Conversely, the largest error source in flood regionalisation will likely be the regional transposition of

the mean annual flood. Beyond a return period of 40 years, the random error increases with return period. This increase reflects the increasing uncertainty of the local flood quantile estimation with increasing return period, as will be demonstrated more explicitly later in this paper. While the regionalisation error is likely to decrease monotonically for all return periods, the local estimation error will increase monotonically for all return periods. What is shown here is the sum of the two. The minimum in the random error occurs at a return period of about 40 years which is on the order of the average record length.

The comparison of the methods in Fig. 3 shows that kriging with uncertain data (KUD), where differences in the record lengths are taken into account, has the smallest random error. This method appears to better exploit the information contained in the short flood records. Short records are introduced into the kriging system but they are associated with a relatively large variance σ_i^2. The case of using long series only (OK_LS) lacks the information contained in the short series, so the regionalisation error is significantly larger. The case of using all series but without taking differences in the record length into account (OK)

does use the information contained in the short series but it does not use it well, as too much credence is given to these short series thereby introducing noise and hence increasing the random error. These results suggest that, from a practical perspective, it will be essential to also use flood data from stations with very short record lengths (e.g. 5 years). Although floods with large return periods cannot be derived from these stations, the estimation of the mean annual flood is possible with relatively little uncertainty, which can improve the regional estimation of all flood quantiles substantially. This finding is consistent with the suggestions of the UK-flood estimation handbook of using very short flood records in the regionalisation in addition to the longer records (IH, 1999, 1/p. 18).

4.2. Catchment attributes—multiple regression

Fig. 4 shows the biases and the random errors versus return period for the various multiple regression approaches. Crosses (MR_AP) refer to using a priori selected attributes (catchment elevation, river network density and mean annual precipitation); full circles (MR_BEST) refer to using the three attributes with

the largest correlation coefficient for each catchment and each flood moment; and asterisks (MR_MOST) refer to using the same attributes for all catchments which are those selected in most of the catchments of the MR_BEST scenario. All the regression approaches show some bias (between -0.10 and -0.28), even though the attributes have been transformed to a normal distribution. The contribution of the bias to the total error (rmse) ranges between 4 and 10% (Eq. (16)). The biases are negative implying that the multiple regressions tend to underestimate the flood quantiles. The random errors are much larger (between 0.43 and 0.67). Similarly as with the geostatistical approaches, between return periods of 2.3 and about 20 years, the random errors decrease with increasing return period, at least for the MR_BEST and MR_AP scenarios. This implies that, again, the main error source is the transposition of the mean annual flood and the higher moments can be regionalised with better accuracy. Beyond 20 years, the random errors increase which is related to the uncertainty associated with the local flood quantile estimation.

The comparison of the multiple regression approaches in Fig. 4 shows that the method MR_BEST
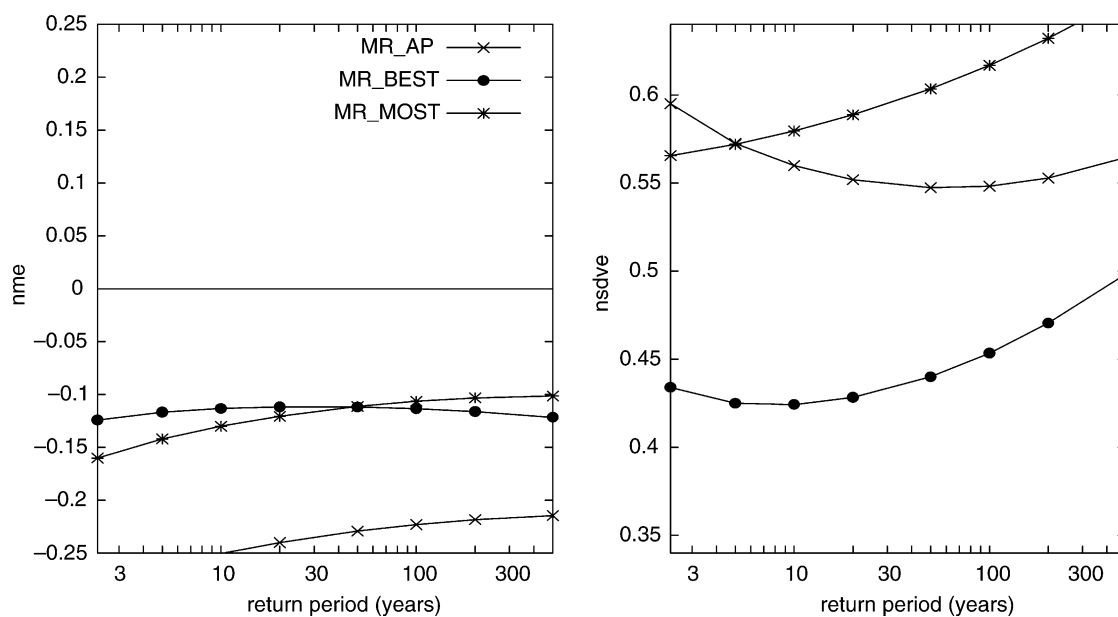

Fig. 4. Bias (left) and random error (right) of flood quantile regionalisation for the multiple regression approaches. Crosses (MR_AP): a priori attributes (catchment elevation, river network density and mean annual precipitation); full circles (MR_BEST): three attributes with the largest correlation coefficient for each catchment and each flood moment; Asterisks (MR_MOST): three attributes that are used in most of the catchments of the MR_BEST scenario. The GEV distribution and product moments are used.

gives the smallest random error and the smallest bias. This is the method where the three catchment attributes associated with the largest multiple correlation coefficient are used. This result supports the assumption that a regression with a large correlation coefficient has also a high predictive power.

The regression with a priori selected attributes (MR_AP) gives much larger random errors which are about 0.55 or more. The selected attributes were thought to be reasonable surrogates for flood controls—elevation for the hydrologic regime and climate; mean annual precipitation for the water input and wetness of the catchment; and river network density for climate, the development of soils and routing efficiency. It appears that the predictive performance is indeed poor which suggests that the relationships between flood moments and catchment attributes are not as tight as one would expect based on general hydrologic reasoning. It is interesting that the MR_MOST scenario performs still poorer for moderate to large return periods. This suggests that the controls of flood frequency vary regionally. A large correlation coefficient does not imply a reasonable correlation in other catchments, so the predictive power in other catchments is poor.

Table 2 shows how often each catchment attribute has been selected for the regression with each flood

moment as a result of a high multiple correlation coefficient. The numbers of positive correlations and negative correlations are given in brackets. The attributes used in most of the catchments can also be interpreted as those with the highest predictive power. For the mean annual flood (MAF) the three attributes used most are maximum annual daily precipitation (MADP), river network density, and the portion of a geological formation termed Austroalpine crystalline. As would be expected, MADP and river network density are positively correlated with the MAF for all catchments. Both attributes are a measure of the water input and the water availability in a catchment. In addition, river network density can be thought of as measure of the efficiency of runoff routing within a catchment, so one would expect floods to increase with both attributes. A negative correlation is found for Austroalpine crystalline which cannot be readily explained by hydrological reasoning. We believe that this is a spurious correlation which results from a co-location of this geological formation with a region of relatively dry catchments in southern Austria. Similarly, the positive correlation of the MAF with the portion of Calcareous Alps is likely a spurious correlation which results from the location of this formation at the northern fringe of the Alps where rainfall is enhanced by

Table 2
Number of instances each catchment attribute is used in the three parameter regression

Catchment attribute	MAF	CV	CS	λ_1	λ_2	λ_3
Area	2 (+2/−0)	55 (+0/−55)	16 (+0/−16)	2 (+2/−0)	9 (+5/−4)	39 (+0/−39)
Elevation	26 (+12/−14)	38 (+4/−34)	40 (+38/−2)	26 (+12/−14)	12 (+2/−10)	15 (+9/−6)
River network density	56 (+56/−0)	31 (+31/−0)	25 (+18/−7)	56 (+56/−0)	63 (+63/−0)	73 (+73/−0)
MAP (mean annual prec.)	8 (+0/−8)	18 (+14/−4)	44 (+1/−43)	8 (+0/−8)	16 (+2/−14)	25 (+3/−22)
FARL (Lake index)	3 (+3/−0)	9 (+9/−0)	23 (+23/−0)	3 (+3/−0)	6 (+6/−0)	7 (+7/−0)
Slope	13 (+11/−2)	11 (+7/−4)	12 (+8/−4)	13 (+11/−2)	5 (+4/−1)	7 (+2/−5)
MADP (max. annual daily prec.)	65 (+65/−0)	31 (+0/−31)	23 (+16/−7)	65 (+65/−0)	81 (+81/−0)	47 (+47/−0)
Porous aquifer	17 (+0/−17)	17 (+8/−9)	19 (+7/−12)	17 (+0/−17)	22 (+0/−22)	15 (+0/−15)
Forest	18 (+0/−18)	47 (+46/−1)	27 (+23/−4)	18 (+0/−18)	0 (+0/−0)	5 (+2/−3)
Glacier	34 (+34/−0)	0 (+0/−0)	24 (+19/−5)	34 (+34/−0)	21 (+21/−0)	36 (+36/−0)
Tertiary + quaternary	3 (+1/−2)	22 (+17/−5)	33 (+11/−22)	3 (+1/−2)	3 (+3/−0)	5 (+1/−4)
Calcareous Alps	33 (+33/−0)	25 (+4/−21)	20 (+3/−17)	33 (+33/−0)	53 (+34/−19)	36 (+26/−10)
Austroalpin Crystalline	46 (+0/−46)	8 (+4/−4)	21 (+21/−0)	46 (+0/−46)	44 (+0/−44)	12 (+1/−11)
Rendzina	35 (+23/−12)	26 (+25/−1)	13 (+11/−2)	35 (+23/−12)	27 (+27/−0)	35 (+35/−0)
Cambisol	7 (+3/−4)	28 (+26/−2)	26 (+24/−0)	7 (+3/−4)	4 (+2/−2)	9 (+9/−0)
Total	366	366	366	366	366	366

Numbers for positive and negative correlations are shown in brackets. MAF, CV and CS are the first three product moments. The λ are the first three L-moments.

orographic effects. Surprisingly, mean annual precipitation (MAP) is only chosen in eight catchments and the correlation is negative. This is in stark contrast with most correlation studies reported in the literature where MAP is usually the second most important predictor variable after catchment area (e.g. IH, 1999). It seems that MADP is a better surrogate for the water input during floods. The two catchment attributes are highly correlated, but not used for the regression at the same time. Note, that catchment area is only selected in two cases. This is because all flood peaks have been standardised by catchment area (Eq. (1)), so catchment scale effects have been removed from the MAF.

The three attributes that are used most for the correlation with the coefficient of variation (CV) are catchment area, the portion of forest and topographic elevation. The catchment area is negatively correlated. This is interesting in the context of the debate of whether CV changes with catchment scale (e.g. Smith, 1992). In fact, the decrease is consistent with the results of Blöschl and Sivapalan (1997) found by a derived flood frequency analysis for Austria. Note that, although the data set of Blöschl and Sivapalan (1997) was the same as in this paper, the method of analysis was completely different. Results of Merz and Blöschl (2004) suggest that the dependence of CV on catchment area depends on the flood process type and is most significant for floods produced by long synoptic rainfall events and least significant for flashy floods produced by thunderstorms. Forest cover is mainly positively correlated, but an interpretation based on local hydrologic reasoning is not obvious. Topographic elevation is mainly negatively correlated. This is due to the large CVs in the dry lowland catchments in eastern Austria. A moderate flood variability and very low mean annual floods give rise to high CVs in these catchments.

The two attributes that are used most for the correlation with skewness (CS) are mean annual precipitation (MAP) and catchment elevation. The correlation of CS and MAP is mostly negative. There is a clear interpretation for this. MAP is a measure of the wetness of the catchment. In dry catchments, with low MAP, often most of the floods are small but there are a few extreme events producing highly skewed flood samples. Conversely, in wet catchments the maximum annual floods

always tend to be large, so the samples are much less skewed. The mean catchment elevation is positively correlated with CS which can be explained by a small number of high alpine catchments with very skewed flood samples. Apparently, the presence of glaciers gives rise to a large number of moderate flood events and a small number of extreme events resulting in large CSs. The other attributes are selected less frequently than these two.

As the first L-moment, λ_1, is identical with the first product moment, MAF, the correlations are also identical. For the second and third L-moments, λ_2 and λ_3, the three attributes selected most of the times are river network density, MADP and portion of Calcareous Alps, all with a positive correlation as well as catchment area with a negative correlation. This is similar to the results for the first moment. It appears that, as the higher L-moments are derived from the first moment, the selected catchment attributes are also similar.

4.3. Spatial proximity and catchment attributes—external drift kriging and georegression

Fig. 5 shows the biases (left) and random errors (right) versus return period for the approaches that combine kriging and catchment attributes. Open circles (EXTDK) refer to external drift kriging using mean annual precipitation; crosses (GEOREG) refer to georegression with the three catchment attributes exhibiting the best correlation coefficient; asterisks (GEOREG_KUD) refer to the same regionalisation approach as in GEOREG but differences in the record length have been taken into account for the interpolation of the residuals; full circles (GEOREG_KUD/KUD) refer to a regionalisation where the mean annual flood is interpolated by georegression (as in GEOREG_KUD) while CV and CS are interpolated by kriging taking record lengths into account (KUD). All the approaches show small negative biases (between −0.09 and −0.02) and much larger random errors (between 0.29 and 0.42). The contribution of the bias to the total error (rmse) is less than 2% (Eq. (16)). Kriging is an unbiased estimator, so small biases would be expected.

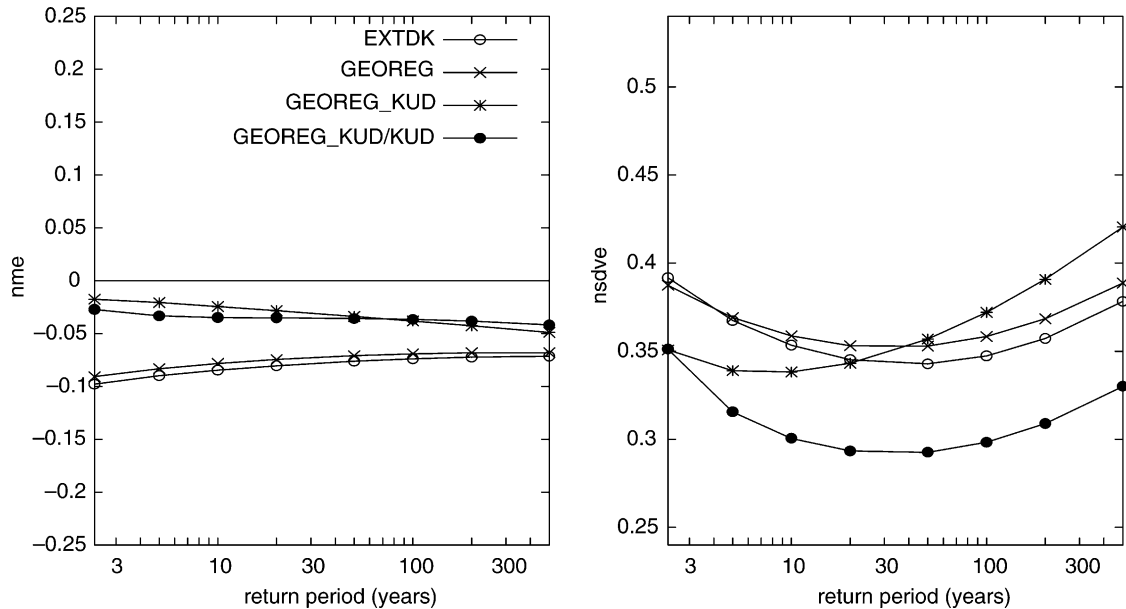Similar to the previous cases, the random errors decrease with increasing return period between 2.3

Fig. 5. Bias (left) and random error (right) of flood quantile regionalisation for the kriging approaches using catchment attributes. Open circles (EXTDK): external drift kriging with mean annual precipitation; Crosses (GEOREG): georegression with the three catchment attributes exhibiting the best correlation coefficient; asterisks (GEOREG_KUD): as in GEOREG but record length taken into account for residuals; Full circles (GEOREG_KUD/KUD): as in GEOREG_KUD for mean annual flood but kriging (KUD) interpolation for CV and CS. The GEV distribution and product moments are used.

and about 40 years, reflecting smaller errors in the higher moments than in the first moment, and increase beyond 40 years reflecting the increasing uncertainty of the local flood estimates. The comparison of the four methods gives the following results: GEOR-EG_KUD/KUD gives the smallest random error and a small bias. This is the method where catchment attributes are only used in the georegression of the mean annual flood, while CV and CS are interpolated using KUD without using the catchment attributes. GEOREG_KUD, where all three flood moments have been interpolated assisted by information on the catchment attributes, yields larger random errors. This means that the use of catchment attributes for the second and third moments deteriorates the predictive performance as compared to the method that uses only spatial proximity for the second and third moments. Not only do the correlations with catchment attributes for the second and third moment add no information, but they apparently add noise as a result of spurious correlations. While this result is not intuitive, it is consistent with common practice in flood

regionalisation where the mean annual flood is estimated from catchment attributes while the higher moments are assumed to be uniform across a region, as is the case in the index flood method.

The georegression where differences in record lengths are not taken into account (GEOREG) yields larger random errors than the method where record lengths are taken into account (GEOREG_KUD), at least for small to moderate return periods. Also, the biases are higher. External drift kriging (EXTDK) with mean annual precipitation gives similar results as the georegression (GEOREG) with three attributes, with slightly smaller random errors and slightly larger biases. The main difference of these two methods is the number of catchment attributes used (one in the case of external drift kriging and three in the case of georegression). This means that, adding information on the catchment attributes, results in hardly any change in the predictive performance. In contrast, accounting for the uncertainty due to different record lengths (GEOREG_KUD/KUD) improves the predictive performance much more.

4.4. Spatial proximity and catchment attributes—Region of Influence approach

In Fig. 6, the biases and random errors for various variants of the Region Of Influence (ROI) approach have been plotted versus the return period. For all variants, the mean annual flood was estimated by multiple regression with three catchment attributes exhibiting the best correlation coefficient and the variants differ in terms of how the second and third moments have been estimated in the ROI approach. Asterisks (MR_BEST/ROI_BEST) refer to the ROI approach where the distance measure uses the three catchment attributes exhibiting the best multiple correlation coefficient. Crosses (MR_BEST/ROI_-DIST) refer to the ROI approach where the distance measure uses geographical distance only. Full circles (MR_BEST/ROI_BEST+DIST) refer to a combination of the previous two variants.

All the ROI approaches show small biases (between 0.07 and 0.14), and relatively large random errors (between 0.41 and 0.52). The contribution of the bias to the total error (rmse) is always less than 5% (Eq. (16)). The random errors for the variant that uses catchment attributes alone is largest and varies from 0.43 to 0.52. If catchment attributes are replaced by geographical distance, the random errors decreases significantly for all return periods larger than 5 years. The combined variant that uses geographical distance and catchment attributes performs still slightly better. These results clearly indicate that for the data set used here, spatial proximity alone is a better predictor of flood frequency than catchment attributes alone and a combination of them yields still better results.

4.5. Comparison of methods

We now compare the four genres of regionalisation methods discussed above. For each genre, the predictive performance of the variant with the smallest errors is shown in Fig. 7. The biases are shown on the left, the random errors are shown on the right. Most striking in Fig. 7 is that the errors of the two geostatistical approaches (open circles—KUD; full circles—GEOREG_KUD/KUD) are significantly smaller than those of the other approaches. For example, for a 100-year flood, the random errors of the geostatistical approaches are 0.30 and 0.33 while
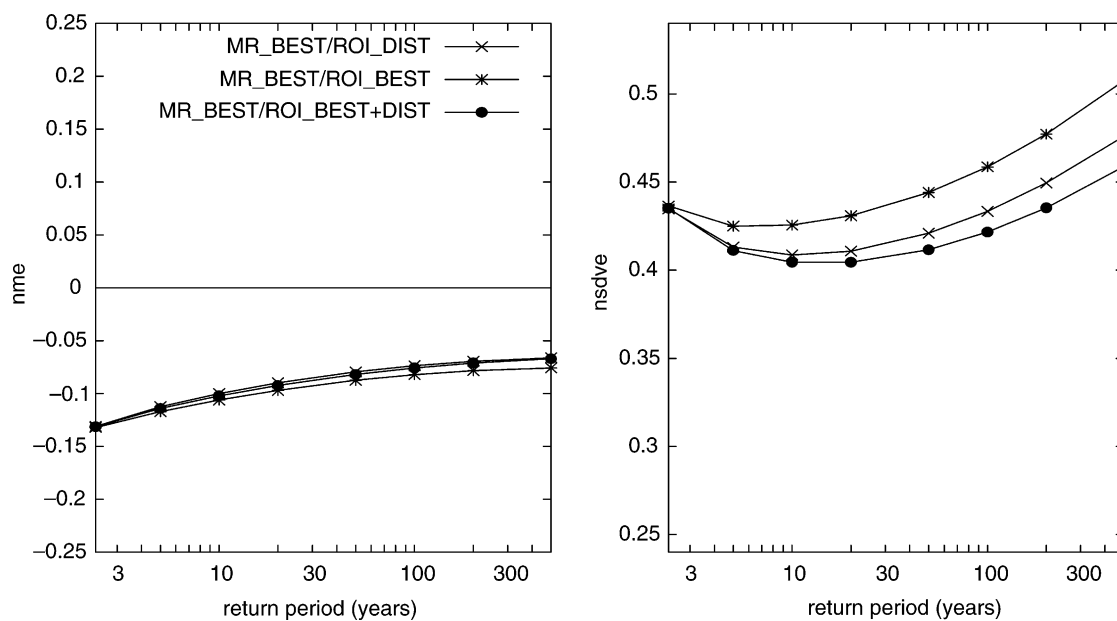

Fig. 6. Bias (left) and random error (right) of flood quantile regionalisation for the Region Of Influence (ROI) approaches (CV and CS). In all variants, the mean annual flood is regionalised by multiple regression (MR_BEST as in Fig. 4). Asterisks (ROI_BEST): ROI with the three catchment attributes exhibiting the best correlation coefficient; crosses (ROI_DIST): ROI with geographical distance: full circles (ROI_BEST+DIST): ROI with catchment attributes and geographical distance. The GEV distribution and product moments are used.
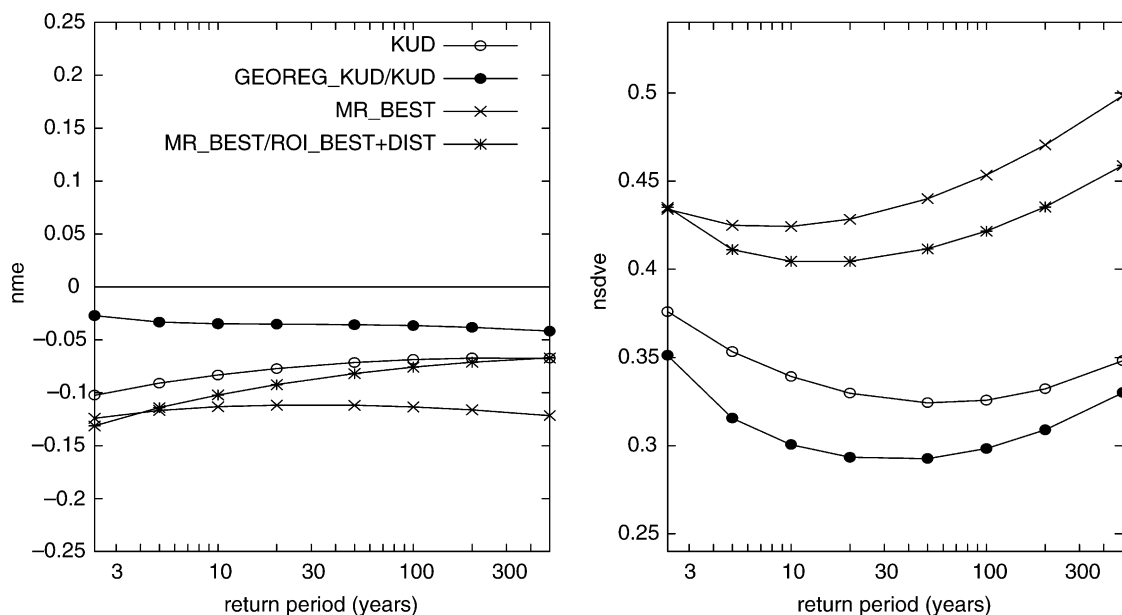
Fig. 7. Comparison of bias (left) and random error (right) of flood quantile regionalisation for the best of the regionalisation types considered in Figs. 3–6. Open circles (KUD): variant of kriging; full circles (GEOREG_KUD/KUD): variant of georegression; crosses (MR_BEST): variant of multiple regression; asterisks (MR_BEST/ROI_BEST+DIST): combination of multiple regression and Region Of Influence Approach, the latter using catchment attributes and geographical distance. The GEV distribution and product moments are used.

they are 0.42 and 0.46 for the Region Of Influence (ROI) multiple regression and the approach, respectively. These differences very clearly indicate that the use of the spatial correlation structure of the flood moments in the geostatistical methods can significantly improve the regionalisation over approaches that do not use spatial correlations (multiple regression and ROI approaches). The relative merits of using spatial proximity and catchment attributes also becomes apparent. The approach that uses catchment attributes alone (multiple regression) performs poorest. The approach that uses spatial proximity alone (KUD) performs significantly better. The ROI approach that uses both pieces of information performs better than the multiple regression approach and the georegression that uses both pieces of information performs better than the KUD approach. For the comprehensive data set used here, spatial proximity is a significantly better predictor of flood frequency than are catchment attributes. Apparently, the catchment attributes are not representative of the real physical controls of the flood frequency processes.

It is also interesting that the biases of the geostatistical methods are smaller than those of

the other approaches. Kriging is an unbiased estimator, and the biases are indeed small. All biases are negative, implying that all the approaches tend to underestimate flood quantiles. The negative biases can be explained by the selection of stations for the regionalisation and the jack-knifing verification. The regionalisation uses catchments with flood records of any length while the bias is only calculated from catchments with more than 40 years of observation. The average specific mean annual flood (MAF) for catchments with less than 40 years of data is MAF=0.31 m^3/s/km^2 while for catchments with more than 40 years MAF=0.39 m^3/s/km^2 which can explain the sign of the bias. It is likely that these differences in the MAF are due to climate fluctuations, as the short records mainly cover the most recent years.

A more detailed comparison of the predictive performance of the methods in Fig. 7 indicates that the smallest random errors and the smallest biases are obtained by GEOREG_KUD/KUD, where a georegression for the mean annual flood using KUD and the three attributes with the highest correlation coefficient is combined with only KUD for the regionalisation of
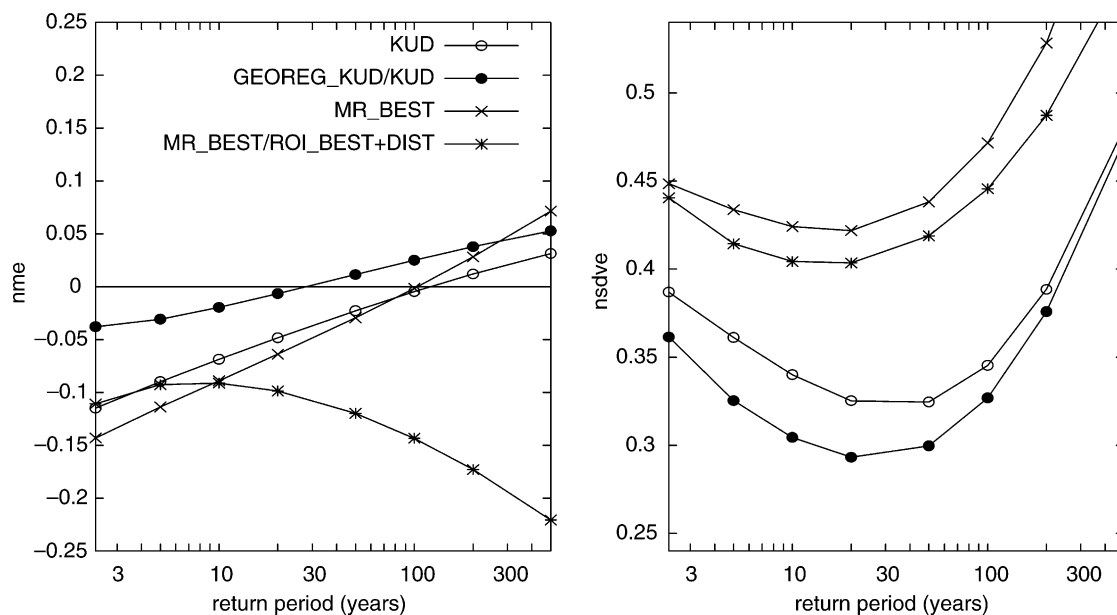
Fig. 8. Comparison of bias (left) and random error (right) of flood quantile regionalisation as in Fig. 7 but using the GEV distribution with *L*-moments rather than product moments.

the higher moments. For a return period of 100 years, the random error and the bias are 0.30 and −0.04, respectively. Using KUD for all flood moments, i.e. without using catchment attributes, gives slightly larger random errors and slightly larger biases (0.33 and −0.07 for a return period of 100 years). This means that information on catchment attributes will improve the geostatistical regionalisation for all return periods but the improvement is not very large. Both the random errors and the biases of the ROI approach are slightly smaller than those of the multiple regression approach. A comparison with Fig. 6 suggests that the main improvement stems from the use of geographical distance in the case of the ROI approach.

In this paper, we have regionalised the flood moments by different approaches and then estimated flood quantiles to judge the predictive performance of the regionalisation approaches against local quantile estimates. It is likely that the selection of the distribution type and the parameter estimation method will affect the results to some degree. To examine the magnitude of these effects we performed a similar comparison, but used *L*-moments instead of product moments and examined other distribution functions in addition to the Generalised Extreme Value (GEV)

distribution used above. Some of the results are shown here. Fig. 8 shows a similar comparison of the regionalisation performance (bias and random error) for the various methods as in Fig. 7 but *L*-moments rather than product moments have been used. Figs. 7 and 8 give similar results. The *L*-moments yield slightly larger random errors and smaller biases for most of the methods, particularly for moderate to large return periods. However, the magnitude of the relative performance of the methods remains unchanged. We found similar results (not shown here) for other distributions such as the Gumbel distribution, both when using product moments and *L*-moments. These comparisons strongly suggest that the findings of this papers also apply to other parameter estimation methods and other distribution functions and are a general result for a hydrologic environment such as the Austrian catchments examined here.

4.6. Analysis of error statistics

To investigate the error sources of the various regionalisation methods in more detail, we performed a number of stratified analyses of the error statistics.

The comparison of the methods in this paper showed that the approaches based on spatial proximity alone give much smaller random errors and smaller biases than the methods based on catchment attributes. One could argue that the predictive power of spatial proximity comes mainly from transposing flood data on the same stream, i.e. from using nested catchments where stream gauges are upstream and downstream neighbours of the site of interest. If this were the case, the predictive power of the geostatistical approach for ungauged and not nested catchments would be poorer. To address this issue we repeated the analysis of the predictive performance for not nested catchments. Specifically, in estimating the flood quantiles for a particular site we did not use the immediate upstream and downstream neighbours in the regionalisation approaches. In Fig. 9 the biases and random errors for this analysis of not nested catchments have been plotted vs. return period. The results show a slight increase in the random error but the relative performance of the methods remains the same. This suggests that the main reason for the relatively good performance of the geostatistical approaches is the presence of spatial correlations of flood characteristics across catchment boundaries and between different streams

in addition to the correlations on the same stream. The geostatistical approaches are therefore likely to perform similarly well in nested and not nested ungauged catchments.

In all the jack-knifing assessments of regionalised against locally estimated flood quantiles in this paper we have so far only used catchments with flood records of more than 40 years to minimise the local estimation error of extrapolating quantiles to large return periods. It is also of interest to perform a similar assessment for a larger number of catchments including those with shorter records. Fig. 10 shows the error statistics for 518 catchments with more than 10 years of observation. As compared to Fig. 7 the random errors are larger while the biases are smaller. The smaller biases are due to the climate effects discussed above as, here, the years of record used in the regionalisation are more similar to those used in the assessment than is the case for Fig. 7. The random errors in Fig. 10 are a measure of the difference between local and regional estimates (Eq. (15)), so larger local estimation errors will also increase these random errors. The location of the minimum of the random error is of most interest. While this minimum was around 40 years in Fig. 7, it shifts to about 5 years in Fig. 10. Five years is where the local estimation
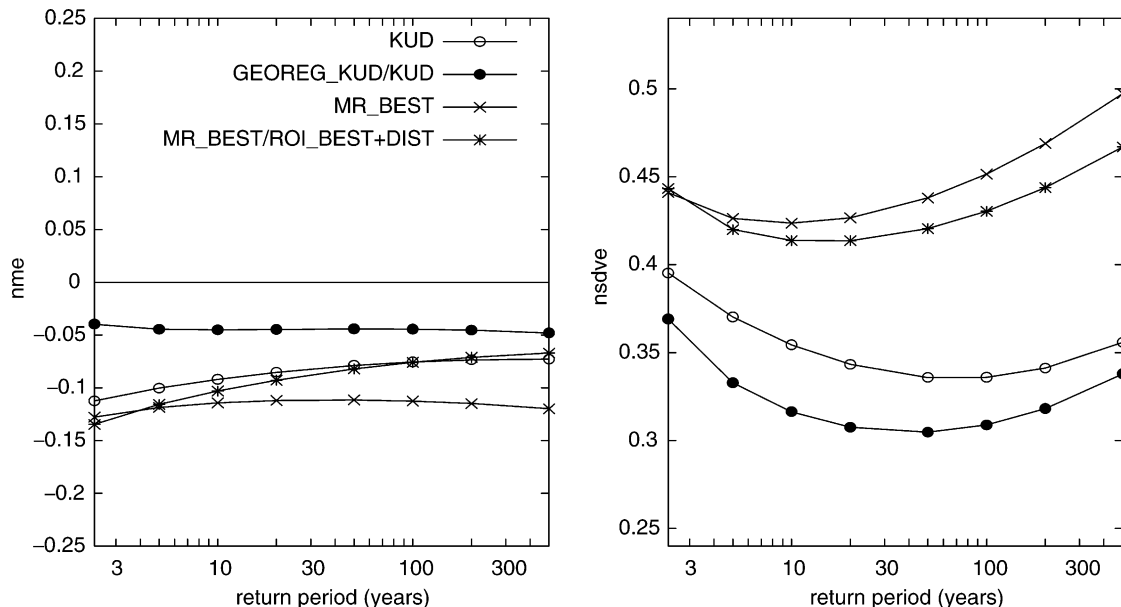

Fig. 9. Comparison of bias (left) and random error (right) of flood quantile regionalisation as in Fig. 7 but without using upstream and downstream neighbours for the regionalisation.
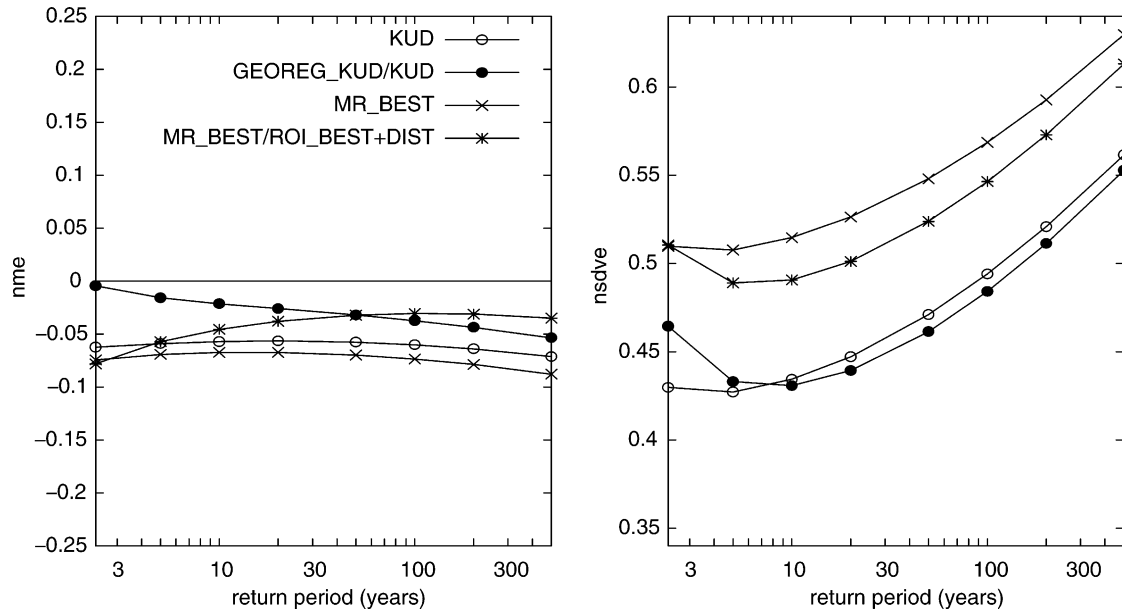
Fig. 10. Comparison of bias (left) and random error (right) of flood quantile regionalisation as in Fig. 7 but jack-knifing for catchments with more than 10 years of observation rather than 40 years of observation.

error becomes more important than the regionalisation error which is perfectly consistent with the shorter record lengths used for the jack-knifing assessment. This finding corroborates the interpretation of the two types of error. For small return periods, the main source of error is the regional transposition of the mean, while for higher return periods the local estimation error gets increasingly more important.

It is also of interest to examine whether the regionalisation performance differs with the hydrologic regimes of the catchments. In Fig. 11 the set of jack-knifing catchments has been stratified into dry and wet catchments according to their mean annual flood. Catchments with mean annual floods smaller than the median of all the catchments have been termed dry catchments and are shown in Fig. 11 (top). Catchments with mean annual floods larger than the median of all the catchments have been termed wet catchments and are shown in Fig. 11 (bottom). The random errors for the wet catchments are much smaller than those for the dry catchments. Note that all the errors in this paper are shown as normalised errors (Eqs. (14) and (15)). The absolute errors in the wet catchments would, of course, be larger than in the dry catchments. The smaller relative errors for

the wet regimes are consistent with hydrologic reasoning. In wet regimes, often, rainfall input is the main control of the magnitude of floods. In terms of its statistical characteristics, rainfall tends not to be as spatially heterogeneous as catchment characteristics, so flood frequency characteristics are not too heterogeneous in space. In contrast, dry catchments tend to respond more non-linearly to rainfall inputs. The catchment state prior to the flood event and soil characteristics, both highly variable in space (Western et al., 2002), tend to produce much more heterogeneous flood frequency patterns, so the spatial correlations between catchments tend to be lower and catchment attributes are less representative of flood frequency. Positive biases for the dry catchments and negative biases for the wet catchments are a result of the stratification.

For the wet catchments, the relative performance of the regionalisation methods is similar to that for all catchments (Fig. 7). However, for the dry catchments the geostatistical approach that does not use catchment attributes (KUD, open circles in Fig. 11) yields the smallest random errors. For these regimes, adding catchment attributes (GEOREG_KUD/KUD, full circles in Fig. 11) deteriorates the predictive
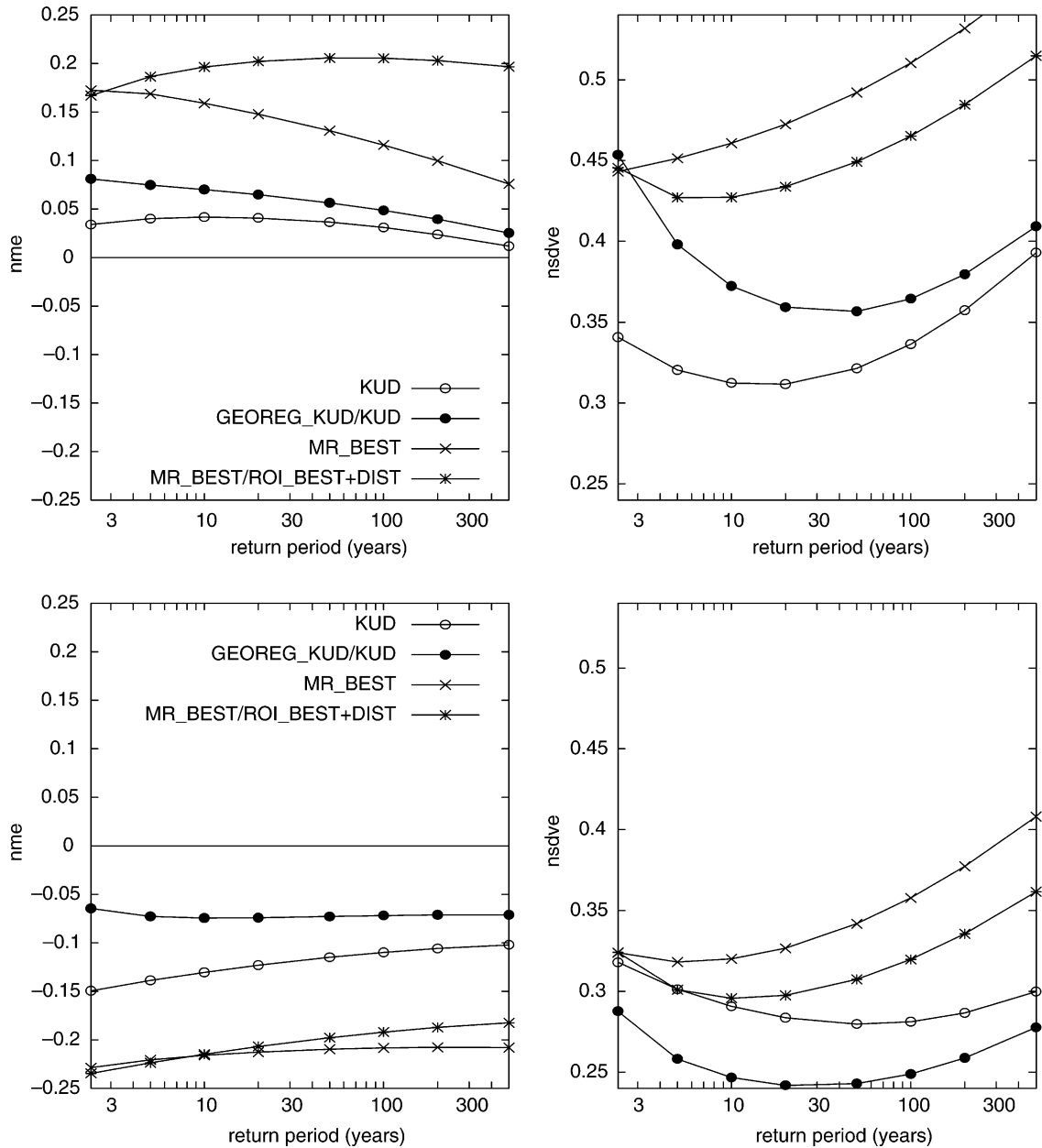
Fig. 11. Comparison of bias (left) and random error (right) of flood quantile regionalisation as in Fig. 7 but stratified by dry (top) and wet (bottom) catchments. Normalised by the group mean.

performance. It is likely that the use of catchment attributes representing the catchment state during flood events would improve the regionalisation for these catchments but these are not the type of attributes available in this paper.

The error statistics presented above represent the average predictive performance over many catchments. An interesting point is to see whether spatial patterns of the predictive performance exist. Fig. 12 shows the locally estimated 100 year flood quantiles
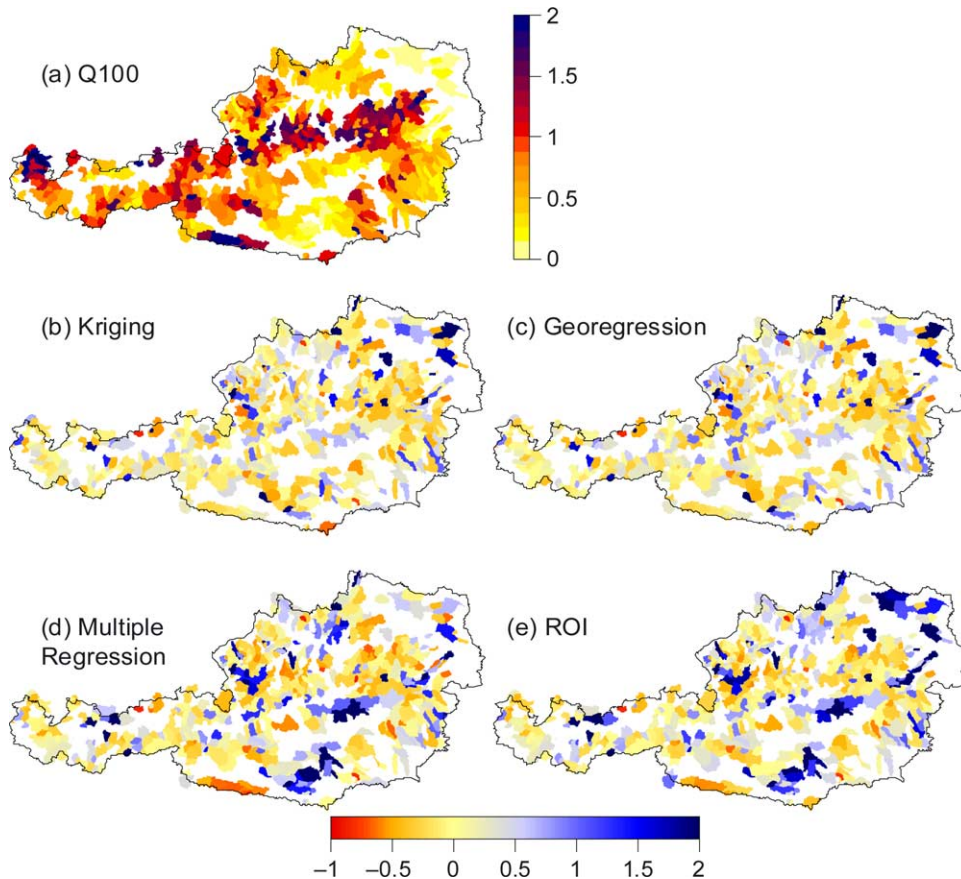
Fig. 12. (a) Locally estimated specific 100 year flood, standardised to a catchment area of 100 km^2 (m^3/s/km^2), (b–e) Normalised difference of the regionalised and local 100 year floods for different regionalisation methods. (b) KUD: variant of kriging; (c) GEOREG_KUD/KUD: variant of georegression; (d) MR_BEST: variant of multiple regression; (e) MR_BEST/ROI_BEST+DIST: combination of multiple regression and Region Of Influence Approach.

(a), and the spatial patterns of the relative errors of the regionalised 100 year flood quantiles (b–e). All catchments with records longer than 5 years have been used in this assessment. The locally estimated 100 year flood quantiles in Fig. 12 are based on a GEV distribution using the local product moments and are plotted in terms of specific discharges standardised to a catchment area of 100 km^2 (Eq. (1)). The 100-year flood quantiles exhibit pronounced regional patterns. Most striking are the large specific floods at the northern fringe of the Alps (see Fig. 1). This is also a region of high rainfall which is a result of the Alps acting as a topographic barrier to north-westerly airflows. The inner part of the high Alps, to the south of this fringe, and the lowlands in eastern Austria

show lower specific floods. There is also a small number of catchments that do not fit into the regional pattern. It is likely that at least part of these outliers are a result of short flood samples in these catchments where one or two extreme floods may dominate the locally estimated quantiles.

The relative errors, shown in Fig. 12(b–e) have been calculated as the difference of the regionalised and locally estimated 100-year quantiles divided by the local value

$$\text{Err} = \frac{Q_{100}^{\text{reg}} - Q_{100}^{\text{loc}}}{Q_{100}^{\text{loc}}}, \qquad (17)$$

where Q_{100}^{loc} is the locally estimated specific 100 year flood and Q_{100}^{reg} is the specific 100 year flood estimated

by one of the regionalisation methods. All specific floods relate to hypothetical 100 km² catchments as of Eq. (1).

All error maps in Fig. 12 are quite patchy and show a lot of small scale variability. This means that none of the regionalisation methods are very good at predicting the small scale spatial variability of flood frequency. Most of the catchments are however, coloured in yellow indicating small to moderate predictive errors. There are a small number of dark blue catchments, particularly in regions of low specific floods. In these catchments, the regionalisation methods massively overestimate the local flood quantiles. Kriging, based only on spatial proximity, and georegression, based on both spatial proximity and catchment attributes (Fig. 12b and c) give the smallest errors. The two patterns are very similar. The smaller errors of the georegression approach shown in Fig. 7 are apparently due to a slightly better predictive performance in many catchments rather than due to a few outliers. The pattern of the Region Of Influence (ROI) approach and the multiple regression are similar (Fig. 12d and e). Both approach tend to underestimate floods in regions of high discharges and overestimate floods in regions of low discharges. In areas where the spatial gradients in the flood quantiles are large, the errors are particularly large, in a number of catchments more than 100%. It is interesting that in the north-eastern part of Austria, where flood discharges are very small, the multiple regression approach performs better than the kriging and georegression approaches. However, in southern Austria, where flood discharges are also very small, it performs poorer.

5. Conclusions

In this paper, we have examined the predictive performance of flood regionalisation methods. The assessment of the performance is based on a jack-knifing comparison of locally estimated and regionalised flood quantiles for 575 Austrian catchments, 122 of which have a record length of 40 years or more. This assessment is a measure of how well the regionalisation methods can estimate floods in ungauged catchments. For all methods, the bias is relatively small while the random errors are

substantial. The main error source in flood quantile regionalisation is the spatial transposition of the mean annual flood. The second and the third moments can be regionalised with better accuracy.

A method that combines spatial proximity and catchment attributes yields the best predictive performance. For the first moment, this method is based on a georegression that uses both spatial proximity and catchment attributes while for the second and third moments the method is based on kriging and uses spatial proximity alone. Including information on catchment attributes in the regionalisation of the second and third moments deteriorates the predictive performance.

A method that uses only spatial proximity performs second best. This is a novel method proposed in this paper which is based on kriging and takes differences in the length of the flood samples into account. This method performs significantly better than conventional kriging as it is able to exploit the information in short records. By applying this method, it is also shown that short flood records (e.g. 5 years) contain valuable information and should be used in flood frequency regionalisation. The predictive performance of the methods based on spatial proximity is shown to be due to the spatial correlation of flood frequency characteristics across catchment boundaries in addition to the spatial correlations on the same stream, so the method should work equally well both for nested and not nested ungauged catchments. Spatial proximity is a surrogate for unknown controls on flood frequency that vary smoothly in space.

The methods that only use catchment attributes perform significantly poorer than the methods based on spatial proximity, including those that use spatial proximity alone. The Region Of Influence (ROI) approach and multiple regression yield similar regionalisation errors if only catchment attributes are used in both cases. Inclusion of the spatial distance in the ROI approach reduces the random errors. This suggests that the catchment attributes available at the regional scale are not very good predictors of flood frequency characteristics. The multiple regressions are used to examine which of the catchment attributes have the highest predictive power for flood regionalisation. A hydrological interpretation of the predictive power is possible for some attributes

while it is not for others. All regionalisation methods perform better for wet catchments than they do for dry catchments.

In the engineering practice, flood frequency regionalisation is often supported by expert judgement (e.g. IH, 1999). In this paper, we have chosen to use a reproducible comparison that is solely based on the hard data available in the data set. One would expect that local expert knowledge will improve the predictive performance of all methods, particularly for the ROI approach, but this is difficult to quantify in an objective way.

We suggest that for improving flood frequency regionalisation better predictive variables and similarity measures need to be found than those currently used. We are currently examining dynamic catchment attributes that represent flood processes during individual events and should therefore also be better regional predictors (e.g. Merz et al., 1999; Merz and Blöschl, 2003). These indicators include seasonality measures, catchment state variables such as antecedent moisture and snow cover, as well as storm type indicators. One would hope that these dynamic catchment attributes will allow a more realistic representation of flood processes and will help go beyond simple proximity measures in regionalising flood frequencies.

Acknowledgements

References

Acreman, M.C., Sinclair, C.D., 1986. Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. J. Hydrol. 84, 365–380.

Behr, O., Blöschl, G., Piock-Ellena, U., 1998. A digital data base of catchment characteristics for assessing anthropogenic effects on flood processes. Ann Geophys. Eur. Geophys. Soc. Supplement II to vol. 16, C507.

Blöschl, G., Sivapalan, M., 1997. Process controls on regional flood frequency: coefficient of variation and basin scale. Water Resour. Res. 33 (12), 2967–2980.

Blöschl, G., Piock-Ellena, U., Merz, R., 2000. Abflußtypen-Klassifizierung als Basis für die Regionalisierung von Hochwässern. (Flood type classification as a basis for regionalising floods). Endbericht 2000, Österreichische Akademie der Wissenschaften (HÖ-12). Institut für Hydraulik, TU-Wien.

Bobée, B., Rasmussen, P., 1995. Recent advances in flood frequency analysis, U.S. Natl. Rep. to Int. Union Geod. Geophys.: 1991–1994. Rev. Geophys. (Suppl.) 33, 1111–1116.

Burn, D.H., 1990. Evaluation of regional flood frequency analysis with a region of influence approach. Water Resour. Res. 26 (10), 2257–2265.

Castellarin, A., Burn, D.H., Brath, A., 2001. Assessing the effectiveness of hydrological similarity measure for flood frequency analysis. J. Hydrol. 241, 270–285.

Cunnane, C., 1988. Methods and merits of regional flood frequency analysis. J. Hydrol. 100, 269–290.

Dalrymple, T., 1960. Flood Frequency Analysis, Water Supply Paper 1543_a. US Geological Survey, Reston, VA.

de Marsily, G., 1986. Quantitative Hydrogeology. Academic Press, San Diego, CA.

Deutsch, C.V., Journel, A.G., 1997. GSLIB, Geostatistical Software Library and User's Guide. Oxford University Press, New York. 384 pp.

Ecker, R., Kalliany, R., Steinnocher, K., 1995. Fernerkundungsdaten für die Planung eines Mobilfunknetzes. Österr. Zeitschr. f. Vermessung und Geoinformation 83, 14–25 (Jahrgang, Heft 1+2).

Geologische Bundesanstalt, 1998. Metallogenetische Karte 1:500.000 auf CD-ROM. GBA, FA-ADV, Wien.

Hirsch, R.M., Helsel, D.R., Cohn, T.A., Gilroy, E.J., 1992. Statistical analysis of hydrological data, in: Maidment, R. (Ed.), Handbook of Hydrology. McGraw-Hill, New York, pp. 17.1–17.55.

Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis—An Approach based on L-Moments. Cambridge University Press, Cambridge.

Hydrographisches Zentralbüro (HZB), 2000. Hydrologisches Jahrbuch von Österreich 1997. Hydrographischres Zentralbüro im BMLF, Band, 105, Wien, 2000.

Institute of Hydrology (IH), 1999. Flood Estimation Handbook. Institute of Hydrology, Wallingford.

Merz, R., Blöschl, G., 1999. Regionalisation of flood frequency, in: Rosso (Ed.), Third Interim Framework Report. Institut für Hydraulik Gewässerkunde und Wasserwirtschaft, Technische Universität Wien, , pp. 23–27.

Merz, R., Blöschl, G., 2003. A process typology of regional floods. Water Resour. Res. 30, 1340.

Merz, R., Piock-Ellena, U., Blöschl, G., Gutknecht, D., 1999. Seasonality of flood processes in Austria, in: Gottschalklz, L., Olivry, J.C., Reed, D., Rosbjerg, D. (Eds.), Hydrological Extremes: Understanding, Predicting, Mitigating Proceedings of Birmingham Symposium, July 1990. IAHS Publication no. 255, pp. 273–278.

Merz, R. Blöschl, G., Piock-Ellena, U., Rieger, W., 2000a. Regionalisierung von Bemessungshochwässern mit geostatistischen Verfahren (Regional estimation of design floods by geostatistical techniques). Interpraevent 2000, June 26–30, 2000 in Villach, Austria, vol. 1, pp. 71–84.

Merz, R., Piock-Ellena, U., Blöschl, G., Kirnbauer, R., 2000b. Skalierungsprobleme bei der Regionalisierung von Hochwässern. (Scale problems in regionalising floods). Endbericht 2000, Österreichische Akademie der Wissenschaften (HÖ-18). Institut für Hydraulik, TU-Wien.

ÖBG, 2001. Bodenaufnahmesysteme in Österreich. Mitteilungen der Österreichischen Bodenkundlichen Gesellschaft. Heft 62.

Pfaundler, M., 2001. Adapting, Analysing and Evaluating a Flexible Index Flood Regionalisation Approach for Heterogeneous Geographical Environments. Schriftenreihe des Institutes für Hydromechanik und Wasserwirtschaft ETH Zürich, Zürich.

Pilgrim, D.H., 1983. Some problems in transferring hydrological relationships between small and large drainage basins and between regions. J. Hydrol. 65, 49–72.

Rieger, W., 1999. Topographischer Feuchteindex für ganz Österreich, in: Strobl, Blaschke (Eds.), Angewandte geographische Informationsverarbeitung, vol. XI. AGIT, Salzburg, pp. 436–447.

Smith, J.A., 1992. Representation of basin scale in flood peaks distributions. Water Resour. Res. 28 (11), 2993–2999.

Tasker, G.D., 1987. Regional analysis of flood frequencies: I, in: Singh, V.P. (Ed.), Regional Flood Frequency Analysis. Reidel, Dordrecht, pp. 1–9.

Western, A., Grayson, R., Blöschl, G., 2002. Scaling of soil moisture: a hydrologic perspective. Ann. Rev. Earth Planet. Sci. 30, 149–180.

Zrinji, Z., Burn, D.H., 1994. Flood frequency analysis for ungauged sites using a region of influence approach. J. Hydrol. 153 (1-4), 1–21.