Characteristic scales, temporal variability modes and simulation of monthly Elbe River flow time series at ungauged locations

D. Marković *, M. Koch

Department of Geohydraulics and Engineering Hydrology, University of Kassel, Kurt-Woltersstr. 3, 34109 Kassel, Germany

Received 19 August 2005; received in revised form 12 April 2006; accepted 14 April 2006
Available online 25 October 2006

Abstract

Spatial and temporal patterns of the long-range extreme monthly Elbe River flows across Germany are investigated, using various statistical methods, among others, principal component and wavelet analysis. Characteristic time scales are derived for various time series statistics. The wavelet analysis of the raw river discharge data as well as of the major principal component reveal the main oscillatory components and their temporal behavior, namely low frequency oscillations at interannual (6.9 yr) and interdecadal (13.9 yr) scales. The EOFs at ungauged stations are estimated from the principal components of the observed time series sampled over a limited time span whose length equals the major temporal variability scale (≈7 yr). The EOFs (empirical orthogonal functions) obtained in this way are subsequently used to simulate long-range flows at these locations. A comparison of this method with linear interpolation and ordinary kriging of the EOF shows the superiority of the former in representing the distributional properties of the observed time series. The simulated time series preserve also short and long-memory.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Temporal variability; Principal component analysis; Wavelets; Ungauged locations; Interannual to interdecadal variability; Elbe River

1. Introduction

Hydrologic engineers are mostly called upon to provide information on activities such as designing or operating hydraulic-conveyance and water-control facilities, preparing for and responding to floods, or regulating floodplain activities. For example, a flood-damage reduction study may require an estimate of the extreme flood peaks as well as estimates of an runoff increase for proposed changes to land use or climate scenarios. To achieve such tasks, characteristic flow quantiles, i.e., flows for a certain occurrence probability are required. A prerequisite for deriving such quantiles are long-range time series of flow observations which are, almost as a rule, not available at the planned construction site.

In spite of significant progress made in recent years in the understanding of the basic processes that govern the runoff formation through a watershed and propagation of a flood wave through a channel, practical application of a physically based hydrological model is often limited by a paucity of soil, vegetation or channel data. While this might be less of a problem for an event, lumped, empirical rainfall-runoff model where mostly only a short flood discharge time series is simulated, classical hydrological models are of a little help when long-term time series are considered. The situation is even more worrisome when there is lack of calibrating gauge data. In fact, many catchments are ungauged, particularly in less developed countries, which in return are often struck by extreme hydrological events. Hence, there is an urgent need to have easy, consistent, robust and reliable methods for simulating and predicting streamflow discharges within ungauged catchments and this is, thus, one of the most challenging issues in modern hydrology.
In this study we investigate the performance of the principal component analysis (PCA) regarding the simulation of the long-term river flows at ungauged locations. The PCA technique is a pure statistical method frequently also called EOF (empirical orthogonal function) analysis with widespread application for the multivariate data analysis in oceanography, atmospheric sciences and climatology (see, e.g. Emery and Thomson, 1997). Exemplarily, from the numerous PCA studies made up-to-date we note that of Rodriguez-Puebla et al. (2001) on the spatial and temporal patterns of winter precipitation over the Iberian Peninsula and their connection to teleconnection indices, and that of Brunetti et al. (2004), who explored changes in statistical properties of daily precipitation in Italy. An extensive PCA-based hydrological study is that of Sacco and Brunetti, 2003). A signal processing tool that performs simultaneous time-frequency localization is the wavelet with the non-dimensional frequency set to $w_0 = 6$. The wavelet transform is a common tool for performing a continuous time-frequency localization of a time series $f(t)$. The method has been applied in numerous studies in geosciences and climatic research (e.g. Foufoula-Georgiou and Kumar, 1994; Torrence and Compo, 1998; Zoppou et al., 2002; Gray et al., 2003). We performed the wavelet analysis using the continuous wavelet transform (CWT) as proposed by Torrence and Compo (1998). The CWT of a function $f(t)$ can be interpreted as the inverse Fourier transform (FT) of the product of $FT(f(t))$ and the FT of the scaled and translated basic wavelet $Ψ(t/s)$. The basic wavelet used in this study is the Morlet wavelet with the non-dimensional frequency set to $w_0 = 6$. The average of the wavelet power over all local wavelet spectra along the time axis is called the global wavelet spectrum (GWS):

$$W^2(s) = \frac{1}{N} \sum_{n=0}^{N-1} |W_n(s)|^2$$

(1)

It can be shown that the GWS comes closest to a somewhat smoothed version of the classical Fourier spectrum. The significance of the periods observed by the GWS is tested by the $\chi^2$ test (Torrence and Compo, 1998) against the hypothesis of red noise.
2.2. Variogram analysis

The temporal variogram (Skøien and Blöschl, 2003) is defined as the average of variograms of all available Elbe River discharge time series over the selected time interval:

\[ \hat{\gamma}(h) = \frac{1}{\sum_{j=1}^{m} n_j(h)} \sum_{j=1}^{m} n_j(h) (Q(x_j, t_i + h) - Q(x_j, t_i))^2 \]  

(2)

where \( h \) is a time lag, \( m \) denotes the number of used time series and \( Q \) is a discharge at the location \( x_j \). The spatial variogram is analogous to the temporal whereas the time lag \( h \) is replaced by the spatial lag \( l \). To both variograms, theoretical variograms of the exponential potential, Weibull and the Sigmoid-type, respectively, are fitted. For the Weibull-type variogram

\[ \gamma(h) = ah^d(1 - \exp(-h/c^d)) \]  

(3)

the slope at short and long lags is \( b + d \) and \( h \), respectively (Skøien and Blöschl, 2003). The Sigmoid-type variogram

\[ \gamma(h) = a/(1 + b^{-c/h-d}) \]  

(4)

has a typical S-shaped form, where \( a, -d/c, \) and \( b \) control the sill, the slope at small legs, and the range, respectively.

2.3. Principal component analysis (PCA)

PCA compresses time series variability into a set of statistical, orthogonal modes called the principal components (PC) (cf. Emery and Thomson, 1997; Wilks, 1995). The PCs are defined as the projection of the \([N \times M]\) data matrix \( X \) onto the eigenvectors (EOFs) obtained solving the eigenvalue problem \( X'X \) EOF = \( \lambda \) EOF where \( X \) can be either the original matrix of observations or the matrix of anomalies or the matrix of standardized anomalies. Since the eigenvectors due to their orthonormality diagonalize the initial matrix \( (X'X) \) (Wilks, 1995), the sum of the diagonal elements of the \( (X'X) \) matrix (i.e. the trace) equals the sum of the eigenvalues. Consequently, if the \( X \) matrix is the original data matrix, the sum of the eigenvalues of \( (X'X) \) is equal to the sum of squares of the original time series. Since the PCs are simply the projection of the \( X \) matrix on the empirical orthogonal vectors (EOFs), it follows that they are mutually independent and that their total variance is equal to the total variance of the initial time series.

The EOFs define a new coordinate system for the studied data whereas the largest eigenvalues \( \lambda \) indicate the dominant directions of the time series joint variability. However, the EOFs of the original data matrix or of the variance-covariance matrix tend to align near the directions of the variables with the highest sum of squares and highest variance, respectively, wherefore in the PCA of variables in unlike physical units the correlation matrix is preferable (Wilks, 1995). For the time series reconstruction (5), only the \( k \leq M \) PCs which describe the major portion of the joint variance are usually used:

\[ Q(t)_m = \sum_{j=1}^{k} PC(t)_j EOF_j^m, \quad 1 \leq m \leq M, \quad 1 \leq k \leq M \]  

(5)

where, owing to the orthogonality property, the PCs are obtained by

\[ PC(t)_j = \sum_{m=1}^{M} Q(t)_m EOF_j^m. \]  

(6)

Eq. (6) can also be viewed upon as a simple regression relationship between the PC’s and the discharge \( Q \), wherefore the EOFs can be computed from the inversion of the normal equations as:

\[ EOF = (PC_{nj} EOF_{nj})^{-1} (PC_{nj} Q_{nj}) \]  

(7)

where \( t_i \) and \( t_j \) denote the beginning and the end of the selected interval and \( Q \) the sample discharge.

Upon determination of the mathematical description of the EOFs spatial pattern, the \( Q \) at an unknown location \( x \) can be simulated using principal components obtained by the EOF analysis of the data from the neighbouring gauges. Our basic assumptions for this simulation procedure are: (a) At the location \( x \), either a short-range discharge time series is available or the location \( x \) was ungauged over the entire period of interest; (b) The selected discharge statistics \( Q \) remains constant (representative) for a selected sampling interval; (c) EOFs at the particular location \( x \) are time-invariant (which, by definition, they are); (d) For all separation distances \( l \) the increment EOF\( (x + l) - EOF(x) \) has a finite variance \( 2\gamma(l) \) that depends only on \( l \).

Although the best choice for analyzing the joint variability of the single variable time series like the river discharge \( Q \) would be the variance-covariance matrix, assuming that on a position \( x \) no observations are available favors the EOF analysis of the initial data matrix because the maximum number of unknowns is only equal to the length of the EOF-vector.

Since our goal is the time series simulation and not the simulation of their major variability modes, the truncation of the eigenvalues at some prescribed level of the joint variability is not necessary. Moreover, if the subject of interest is the simulation of the time series' large-scale behavior, i.e. of deterministic signals that govern the discharge process, then a better option would be to use the Multichannel Singular Spectrum Analysis (MSSA) of Golyandina and Stepanov (2005) or the classical version of this method described in detail in Ghil et al. (2002). Unlike the classical MSSA, the method of Golyandina and Stepanov (2005) offers a possibility of detecting deterministic components (slowly varying trend, different periodic components), their grouping and forecasting (cf. Marković and Koch, 2005a).

3. Data and study area

The German part of the Elbe River Basin (Fig. 1) comprises almost two thirds of its total Basin size
(148,268 km²), with most of the remaining part belonging to the Czech Republic and about 1% to Poland and Austria. The upper Elbe valley ends just before the Schwarze Elster mouth (199 km), while the lower Elbe valley reaches from the Havel river mouth (438 km) to the Elbe estuary at Cuxhaven (725 km). The hydrological discharge regime is characterised by a pronounced seasonal cycle whose rising limb is situated between November and April and the falling one between May and October.

The German Federal Agency for Hydrology (BAfG) provided the daily discharge time series for eight gauges along the Elbe River whose station name, geographic position, station altitude, km position and time series span are listed in Table 1. From the daily data we derived monthly means and monthly extremes, defined as the mean monthly discharges and maximum daily discharges for each month, respectively. Since the results for the two time series do not vary significantly, only the monthly extremes series will be presented.

### Table 1: Discharge stations used and record information

<table>
<thead>
<tr>
<th>Gauge</th>
<th>Long. [O]</th>
<th>Lat. [N]</th>
<th>H [m]</th>
<th>km</th>
<th>Starting year</th>
<th>N [month]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dresden</td>
<td>25.74</td>
<td>51.05</td>
<td>102.73</td>
<td>55.6</td>
<td>1852</td>
<td>1788</td>
</tr>
<tr>
<td>Torgau</td>
<td>22.01</td>
<td>51.55</td>
<td>75.18</td>
<td>154.1</td>
<td>1936</td>
<td>780</td>
</tr>
<tr>
<td>Wittenberg</td>
<td>21.65</td>
<td>51.86</td>
<td>62.48</td>
<td>214.1</td>
<td>1951</td>
<td>600</td>
</tr>
<tr>
<td>Aken</td>
<td>21.06</td>
<td>51.86</td>
<td>50.24</td>
<td>274.7</td>
<td>1936</td>
<td>780</td>
</tr>
<tr>
<td>Barby</td>
<td>20.89</td>
<td>51.99</td>
<td>46.15</td>
<td>294.8</td>
<td>1900</td>
<td>1212</td>
</tr>
<tr>
<td>Magdeburg</td>
<td>20.65</td>
<td>52.13</td>
<td>39.92</td>
<td>326.6</td>
<td>1931</td>
<td>840</td>
</tr>
<tr>
<td>Wittenberge</td>
<td>20.76</td>
<td>52.99</td>
<td>16.72</td>
<td>453.9</td>
<td>1900</td>
<td>1212</td>
</tr>
<tr>
<td>Neu Darchau</td>
<td>19.89</td>
<td>53.23</td>
<td>5.68</td>
<td>536.4</td>
<td>1875</td>
<td>1512</td>
</tr>
</tbody>
</table>

Fig. 1. The German part of the Elbe River Basin and the Elbe River gauges.

4. Results

4.1. Temporal stationarity scales

Hydrological time series are considered stationary if they are free of trends, periodicities and shifts (Salas, 1992). Statistically, the strict stationarity condition imposes time invariance of the joint distribution of any possible subset of process realizations. When dealing with hydrological time series of moderate time spans, instead of a strict stationarity, a covariance stationarity can be assumed, which implies stationarity up to the second moment (Hipel and McLeod, 1994). Unlike the monthly time series which are due to the annual cycle seasonally nonstationary, it is generally assumed that the hydrological time series at annual scale are stationary. Nonstationarity can be however caused by an anthropogenic action or large-scale climatic variations which act at interannual to interdecadal and even larger scales, wherefore, prior to any kind of time series modelling, the time series temporal stationarity scales have to be investigated.

Classical statistical tests for the equivalence of the subsets moments like the Z-, T- or the F-test are based on the assumption that the data are normally distributed and independent. However, all studied time series are positively skewed and the skewness decreases along the Elbe River course (e.g. Dresden 2.47, Neu Darchau 1.52), whereas the short-range memory, i.e. the “lag-1” – correlations increase (e.g. Dresden 0.38, Neu Darchau 0.62). Although a logarithmic transformation appears to render the data approximately normal, the strong dependence within the data rules out the use of the classical statistical tests.

As a first step in the study of the temporal stationarity we analyze the influence of the time window width $T_{\text{win}}$ on the estimates of the time series’ mutual correlations, first and second moments. As representative gauges we selected Neu Darchau, Barby and Dresden (see Fig. 1). Fig. 2(a) indicates that above the time scale of about 6 yr ($T_{\text{win}} \approx 70$ months), the mutual correlations between the analyzed time series become stationary.

For the estimation of the “characteristic scale” for the time series’ first two moments, in the second step, the extreme absolute “error” of the moment estimate from all possible windows ($T_{\text{win}} \in \{1, N_{\text{max}}\}$) of this length is calculated. As the “true” value, the corresponding moments from the whole time series are used. The procedure is repeated for each of the representative gauges and, finally, the averages of these error estimates are calculated for each window length. Fig. 2(b) shows the results of this procedure for two different time spans, namely, 1960–2000 (1) and 1900–2000 (2) i.e. the corresponding maximal window length $N_{\text{max}}$ is set to 10 yr and 25 yr, respectively. One observes from Fig. 2(b) that for the 1900–2000 set (2) the estimate of the standard deviation decays slower than that of the mean. For all used window widths, the error estimate of the mean is larger for the 1900–2000 set (2) than for the 1960–2000 set (1).
1960–2000 set (1). This is not surprising and can be explained by the entropy principle which says that the uncertainty of the system (e.g. confidence intervals) increases with the number of possible system states, i.e. with the length of the used time series. This simple investigation shows that the “characteristic scales”, i.e., stationarity scales of the mean and the standard deviation are not equal but are approximately in the decadal to interdecadal range (10–15 yr) which is, however, far longer than the “characteristic scale” of the mutual correlations ($\alpha = 0.6$ yr).

For further inspection of the stationarity, the long-range records are split into 3 nonoverlapping sets of equal length and the first two moments (mean and variance) are calculated for each of them. At the gauge Dresden, the first moment is 10% smaller in the third data subset (1952–2001) than in the first subset (1852–1901). The second moment, i.e. the variance, shows a pronounced nonstationarity. Exemplarily, at the gauge Magdeburg, the variance is 15% larger in the third data subset (1967–1999) than in the first subset (1931–1953), while at the gauge Dresden the variance is 35% smaller in the third subset (1952–2001) than in the first subset (1852–1901). Furthermore, a slow decreasing trend of the mean characterizes the discharge time series of the gauge Neu Darchau which is closest to the North Sea estuary of the Elbe River and, therefore, indicates the resulting dynamics of the whole catchment.

We additionally show in Fig. 3 the log-wavelet spectrum of the extreme monthly river discharges from the northeast Elbe River gauge (Neu Darchau, 1875–2000). One observes that before 1920 the seasonal cycle (1 yr scale) is the only statistically significant periodic component whereas after that time a broad low frequency spectrum with dominant variability scales equal to 6.9 yr and 13.9 yr, respectively, is present. Though the time series’ variability at scales higher than 1 yr is obviously nonstationary, there is a clear quasiperiodic pattern, also called the Joseph effect (Mandelbrot, 1977; Beran, 1994) characterized by approximately 7 yr long and dry and wet periods (see Fig. 4). As this pattern has also been found by Marković and Koch (2005a) applying the singular spectrum analysis of Golyandina et al. (2001) on the extreme monthly discharge time series of the Elbe River gauge Dresden, we have enough confidence on the existence of the named periodicity. The most probable source of these quasiperiodic cycles is the North Atlantic Oscillation (NAO) since similar interannual to interdecadal oscillations as well as significant correlations with the NAO are found in the extreme monthly precipitation across Germany (Marković and Koch, 2005b). However, apart from an apparent “skin effect” of the groundwater aquifer which feeds the baseflow to the stream-sections under study, further causes of the low frequency amplification of the river discharge are still a matter of investigation (Marković, 2006).

4.2. Spatial and temporal variograms

For the variogram analysis, the data should nearly follow a normal distribution and be free of periodicities (Skroien and Blöschl, 2003). This can be achieved approximately by logarithmizing the data. Subsequently the wave-
let filter (Torrence and Compo, 1998; Marković and Koch, 2005a) is used to remove the periodicities. In Fig. 5 we show the scatter plots of the temporal (a) and the spatial (b) variability as well as the corresponding theoretical variograms described by Eqs. (3) and (4), respectively.

For the temporal variogram the smallest least squares fitting error is obtained with the Weibull-type model (RMS = 0.01), whereas the spatial variogram is fitted better by the Sigmoid-type function (RMS = 0.02) than by the Weibull variogram (RMS = 0.024). However, unlike the temporal variogram which converges relatively fast (≈1 yr) to its asymptotic value, the spatial variogram does not have any characteristic scale. This property alludes to a long-memory in the spatial domain and short-memory in the time domain (cf. Beran, 1994). The parameters of the spatial variogram can further be used for the geostatistical interpolation (kriging) of the discharge parameters.
4.3. Principal component analysis of the Elbe River flow variability

As mentioned in the theoretical section, for the detection of the joint variability modes the PCA will be applied to the original multivariate data from all available gauges, which means that the data are not transformed prior to the analysis.

In Fig. 6 we show the global wavelet spectra of the first three principal components and the corresponding 95% significance levels assuming a red noise background.
significance levels derived from the $\chi^2$ test assuming a red noise background (see Torrence and Compo, 1998). The percentage of the joint variability (percentage of the total time series sum of squares) explained by the first, second and the third principal components are 98.7%, 0.86% and 0.17%, respectively.

The significance level for the spectra of the first PC indicates two statistically significant variability scales, namely, 1 yr and 6.9 yr, but, a clear peak at 13.9 yr is also noticeable. If one keeps in mind that the global wavelet spectra is the time-averaged wavelet power it becomes clear that the slowly varying oscillation at the scale 13.9 yr is present in the time series, (but not over the entire investigated time interval). Something similar can be observed from the wavelet spectrum of the discharge time series at gauge Neu Darchau (see Fig. 3) where the low frequency oscillations at scales larger than 10 yr are present only between 1920 and 1970. The second PC consists of the periodic component at the scale 5.2 yr, and probably some time-dependent slowly varying components at scale 10–11 yr. The third PC does not have any significant peaks and is therefore considered as a noise. Interestingly, the seasonal (annual) component appears only in the spectra of the dominant PC while the second PC seems to describe the additional low frequency joint variability. Finally, since the 13.9 oscillation is not time-independent, the major joint variability scale of the German Elbe River must be the 6.9 yr period. ($\approx$7 yr).

The EOFs derived from the PCA analysis constitute the directions which explain most of the variability of the system, i.e. in the present application the discharge time series along the Elbe River course. By construction the EOFs are stationary, i.e. they do not evolve in time. It is only the principal component attached to the corresponding EOF that provides the sign and the overall amplitude of the EOF as a function of time. Moreover, the position of these vectors is defined for the $M$-dimensional space in the EOF matrix itself, but if there is a statistically or mathematically derivable dependence of the eigenvectors on the gauge location, we could derive the position of these vectors for the $M+1$ dimensional space. Such a dependence is shown in Fig. 7 where the scatter plots of the EOF1 (circles) and the EOF2 (squares) as a function of the gauge position are drawn. The third PC is not considered because of the negligible portion of variance explained by this component as well as to the fact that this component has a white noise spectrum. Obviously, the dependence of EOF1 and EOF2 on the gauge location can easily be described by a second to third order polynomial alluding to a possibility of simulating flows at ungauged locations using interpolated or modeled EOFs.

### 4.4. Simulation of Flow time series at ungauged locations

With an increase of the number of available discharge time series the uncertainty of the multivariate-based description of the discharge process along the river course decreases. However, in practice, often only a small number of gauges is available, therefore, limiting the application of multivariate methods. The easiest possible method for calculating values between two neighbouring locations would be linear interpolation (LI). If data from at least three gauges are available it becomes reasonable to apply the PCA while for a higher number ordinary kriging (OK)
can be used. We will discuss all these possibilities by comparing their outputs with respect to the time series’ statistical, namely, its distributional properties. Similarly, Haberlandt et al. (2001) compared the performance of the regression method, ordinary- and external drift kriging when regionalizing the base flow index in the Elbe River Basin. The study of Helms et al. (2002) deals with the simulation of Elbe River flows along the river course between 1985 and 1990 using the ELBA model, which is a diffusion–translation model based on the hourly data. However, the authors argue that the difference is negligible if instead of each hour of the day the one mean daily value is used.

For the application of the LI and the PCA method we selected the three gauges Neu Darchau, Barby and Dresden (see Fig. 1) because the gauge Dresden is the most southern gauge, the gauge Neu Darchau is closest to the estuary of the Elbe River and the gauge Barby is about “half way” between these two gauges. The results show that the first two PCs explain 99.7% of the joint variance and, like in the PCA of all available time series (Fig. 6), the spectrum of the third PC does not have any statistically significant peaks. For the calibration of the EOFs, only the first two PCs and a data subsample with a length of 7 yr (1953–1960) are used. This approach is based upon the following assumptions: (a) the third PC represents negligible noise, and (b) the calibration sample should be at least as long as the major temporal variability scale.

To estimate the influence of sampling on the EOF estimate, we have repeated the subsample selection and the least squares procedure (LSR) 1000 times and calculated the corresponding mean and standard deviations. It turned out that for the analyzed data sets the standard deviation of the EOF estimates are only up to 5% and 15% of the mean EOF1 and EOF2, respectively. Assuming that the means of the LSR and OK simulations are seasonally biased, the correction of the models is straightforward, i.e. seasonally dependent (winter and summer) correction factors are simple added to the simulated $Q$ values.

A simulation study is performed for one gauge located in the upper part of the Elbe River Basin (gauge Torgau) and one located in its lower part (gauge Wittenberge) (Fig. 1). In Fig. 8 we show the corresponding observed time series (1984–1986), including those for the selected reference gauges (Neu Darchau, Barby and Dresden). One

![Fig. 8. Elbe River discharge time series: Neu Darchau (dashed line), Barby (triangles), Dresden (circles), Wittenberge (solid line), Torgau (squares).](image_url)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$Q_{obs}$</th>
<th>LI</th>
<th>PC-LSR</th>
<th>PC-LSRc</th>
<th>OK</th>
<th>OKc</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>–</td>
<td>0.985</td>
<td>0.985</td>
<td>0.985</td>
<td>0.984</td>
<td>0.984</td>
</tr>
<tr>
<td>$\overline{Q}$</td>
<td>533.9</td>
<td>659.3</td>
<td>512.0</td>
<td>529.2</td>
<td>657.2</td>
<td>509.2</td>
</tr>
<tr>
<td>$S_Q$</td>
<td>388.1</td>
<td>478.0</td>
<td>378.3</td>
<td>377.3</td>
<td>480.3</td>
<td>473.6</td>
</tr>
<tr>
<td>$C_s$</td>
<td>1.57</td>
<td>1.58</td>
<td>1.60</td>
<td>1.61</td>
<td>1.59</td>
<td>1.61</td>
</tr>
<tr>
<td>$r_1$</td>
<td>0.51</td>
<td>0.53</td>
<td>0.52</td>
<td>0.52</td>
<td>0.53</td>
<td>0.52</td>
</tr>
<tr>
<td>$r_2$</td>
<td>0.25</td>
<td>0.28</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
</tr>
<tr>
<td>$r_3$</td>
<td>0.16</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>$H$</td>
<td>0.81</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.81</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>–</td>
<td>138.93</td>
<td>29.97</td>
<td>64.06</td>
<td>117.17</td>
<td>123.65</td>
</tr>
<tr>
<td>RMS</td>
<td>–</td>
<td>171.19</td>
<td>71.24</td>
<td>67.95</td>
<td>172.30</td>
<td>118.59</td>
</tr>
</tbody>
</table>

Abbreviations are: $Q_{obs}$ – sample data; LI – linear interpolation between two closest referent gauges; PC-LSR – statistics of the PCA based simulation; PC-LSRc – statistics after correction of the PC-LRC; OK – geostatistical interpolation of EOFs; OKc – results after correction of OK simulations; $R^2$ – coefficient of determination; $\overline{Q}$ – mean; $S_Q$ – standard deviation; $C_s$ – skew; $r_1$ – $r_3$ are lag-1 to lag-3 autocorrelations; $H$ – DFA Hurst parameter estimate; $\chi^2$ – test statistic ($\chi^2_{0.05} = 38.88$); RMS – the root mean squares error of each model.
notices only few exceptions from the rule that the discharge increases with increasing downstream distance (Dresden to Barby to Neu Darchau). In Tables 2 and 3 we list the statistical properties of the sample and simulated data. The calibration period for the PCA-based determination of the EOFs at the “ungauged” locations is 1953–1960, and the simulation period is the subsequent time interval (1960–2000).

Interestingly, regarding the portion of variance explained by the simulated time series, all methods provide almost equal results ($R^2 > 0.98$) (cf. Tables 2 and 3). However, the simple linear interpolation (LI) between the two closest reference gauges significantly overestimates (Torgau) or underestimates (Wittenberge) the first two moments. All simulation methods turn out similar results concerning the lag-1 to lag-3 correlations and the Hurst parameter $H$. Although the latter are $>0.5$, this does not mean a priori that a long-term persistence should be assumed, as $H$-parameter estimations based on the DFA technique are usually upwards-biased when the analysed time series contain periodic signals (cf. Hu et al., 2001; Marković and Koch, 2005a). The skewness of the simulated discharge matches better that of the sample at location Torgau than at Wittenberge where it is consistently too high. Similar to LI, the OK method either overestimates (Torgau) or underestimates (Wittenberge) the first two moments. In any case, the PC-LSR method exhibits the best performance regarding the distributional properties of the original sample data and, additionally, results also in the lowest RMS.

In Figs. 9 and 10 we show traces of the sample data, the corrected PC-LSR-$Q_{LSRc}$ and of the OK-simulated time series $Q_{OKc}$, as well as of the absolute errors ($Q_{obs} - Q_{sim}$) obtained with both simulation methods for two Elbe River stations. Expectedly, after correction for the mean, the errors have approximately a zero mean, however, a weak (<5% of the mean) 13.9 yr periodicity was further noticed from the inspection of the corresponding global wavelet spectra of the error terms, confirming the results of the variability scales study of the previous section.

Since the results of the $\chi^2$ test, as well as the basic statistics listed in Tables 2 and 3 indicate the PC-LSR as the best method, we conclude that the PCA can indeed be used

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$Q_{obs}$</th>
<th>LI</th>
<th>PC-LSR</th>
<th>PC-LSRc</th>
<th>OK</th>
<th>OKc</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>-</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
</tr>
<tr>
<td>$\bar{Q}$</td>
<td>952.5</td>
<td>907.1</td>
<td>934.9</td>
<td>935.7</td>
<td>904.6</td>
<td>934.2</td>
</tr>
<tr>
<td>$S^2_Q$</td>
<td>590.6</td>
<td>578.0</td>
<td>588.3</td>
<td>588.4</td>
<td>573.7</td>
<td>597.9</td>
</tr>
<tr>
<td>$C_s$</td>
<td>1.31</td>
<td>1.46</td>
<td>1.44</td>
<td>1.44</td>
<td>1.45</td>
<td>1.45</td>
</tr>
<tr>
<td>$r_1$</td>
<td>0.65</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.63</td>
</tr>
<tr>
<td>$r_2$</td>
<td>0.38</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.35</td>
</tr>
<tr>
<td>$r_3$</td>
<td>0.22</td>
<td>0.20</td>
<td>0.20</td>
<td>0.21</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>$H$</td>
<td>0.81</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>-</td>
<td>47.64</td>
<td>25.89</td>
<td>25.29</td>
<td>51.69</td>
<td>26.92</td>
</tr>
<tr>
<td>RMS</td>
<td>-</td>
<td>75.01</td>
<td>61.50</td>
<td>61.30</td>
<td>77.32</td>
<td>62.76</td>
</tr>
</tbody>
</table>

See Table 2 for explanations.
for the simulation of discharge time series at ungauged locations.

5. Conclusions

Regarding a hydrological process as stationary, equals believe that the process is predetermined, whereas when assuming an asymptotic nonstationarity, dynamic spontaneity of the system is allowed (Klemes, 1974). Various statistical analyses performed up-to-date confirmed that, instead of an a priori hypothesis on either stationarity or nonstationarity, basic temporal or spatial variability scales should be investigated.

Applying temporal and spatial variogram analysis and PCA (EOF) analysis, in combination with the wavelet tool, we were able to (a) extract the characteristic time scales of the monthly Elbe River flows, (b) to determine the main deterministic signals which describe the joint variability of the Elbe River flow across its Basin within the German territory, and (c) show that the EOF- eigenvectors computed from a small subset of the river gauge locations can be used for the simulation of flow data at ungauged locations.

The results indicate that the stationarity scales of the first two moments of the analyzed time series are not equal but are approximately in the decadal to interdecadal range (10–15 yr). Although there is no overall linear trend in the data, a slow decreasing tendency of the mean characterizes the discharge time series of the Northern gauge (Neu Dar- chau), reflecting the aggregate dynamics of the Elbe River Basin. Moreover, in addition to the nonstationary character of the interdecadal oscillations (13.9 yr), there is a clear quasiperiodic pattern distinguished by an approximately 7 yr cycle. A wavelet analysis of the first principal component which describes more than 95% of the joint variability also shows low frequency oscillations at the interannual (6.9 yr) and interdecadal (13.9 yr) scales, with the corresponding EOFs depending on the gauge location.

As for the simulations of flow at ungauged locations, three methods are applied and compared, namely linear interpolation (LI), PC-EOF least squares regression (PC-LSR) and ordinary kriging (OK). With respect to the portion of variance explained by the simulated time series, as well as the lag-1 to lag-3 correlations and the Hurst parameter \( H \) (computed by DFA), all methods appear to fare equally well. However, only the PC-LSR method fits the first two moments of the observed time series satisfactorily and results also in the lowest RMS. We conclude that the PCA can be used as an interesting alternative to deterministic methods for the long-range simulation of discharge time series at ungauged locations, provided that a minimum calibration period of the order of the major temporal variability scale is used.

Acknowledgement

We would like to thank the German Federal Agency for Hydrology (BAFG) for providing the discharge time series data.
and the Potsdam Institute for Climate Impact Research (PIK) for the GIS data of the Elbe River Basin. We would also like to thank the two anonymous reviewers for helpful comments and suggestions.

References