Appears in: InExtremis: Trends, Correlations and Extremes in Hydrology and Climate. eds. J.P. Kropp & H.J. Schellnhuber, Springer, Berlin, 2008

11. Confidence Intervals for Flood Return Level Estimates assuming Long-Range Dependence

Henning W. Rust, Malaak Kallache, Hans-Joachim Schellnhuber and Jürgen P. Kropp

Standard flood return level estimation is based on extreme value analysis assuming independent extremes, i.e. fitting a model to excesses over a threshold or to annual maximum discharge. The assumption of independence might not be justifiable in many practical applications. The dependence of the daily runoff observations might in some cases be carried forward to the annual maximum discharge. Unfortunately, using the autocorrelation function, this effect is hard to detect in a short maxima series. One consequence of dependent annual maxima is an increasing uncertainty of the return level estimates. This is illustrated using a simulation study. The confidence intervals obtained from the asymptotic distribution of the Maximum-Likelihood estimator (Mle) for the generalised extreme value distribution (Gev) turned out to be too small to capture the resulting variability. In order to obtain more reliable confidence intervals, we compare four bootstrap strategies, out of which one yields promising results. The performance of this semi-parametric bootstrap strategy is studied in more detail. We exemplify this approach with a case study: a confidence limit for a 100-year return level estimate from a run-off series in southern Germany was calculated and compared to the result obtained using the asymptotic distribution of the Mle.

11.1 Introduction

Many achievements regarding extreme value statistics and the assessment of potential climate change impacts on frequency and intensity of extreme events have been made in the last years, summarised for instance in [11.34], [11.49] or [11.45]. The IPCC stated that it is very likely for the frequency of intense

◂ **Fig. 11.0.** Here Figure caption for an illustrative introductory images of your chapter precipitation to increase [11.32] with increasing global mean temperature. This implies changes in precipitation patterns, a major factor – among some others – for the intensity and frequency of floods, which ultimately can cause tremendous consequences for nature and societies in a catchment area. This has been already observed in many regions of the world [11.37].

A pressing question is thus whether heavy rain or severe floods become more frequent or intense. Some concepts make use of non-stationary models, e.g., Chapter 5, [11.16, 11.33], or try to identify flood producing circulation patterns [11.2]. A variety of approaches assess changes by comparing windows covering different time spans [11.3]. This procedure is especially useful for getting an impression of possible further developments by comparing GCM control and scenario runs [11.38, 11.57]. A useful indicator for changes in flood frequency and magnitude is the comparison of return level estimates. For this purpose a reliable quantification of the uncertainty of return level estimates is crucial.

Alerted by an seemingly increasing flood risk, decision makers demand for quantitative and explicit findings for readjusting risk assessment and management strategies. Regional vulnerability assessments can be one strategy to deal with the threat of extremes, such as floods or heat waves (e.g., [11.36]). Other approaches try to anticipate extreme scenarios by the help of GCM model runs. The development of risk assessment concepts, however, has still a long way to go, since forecasting of extreme precipitation or floods is highly uncertain (cf. for instance [11.42]). Another potential problem in the risk assessment framework is the quantification of uncertainty in extreme value statistics. In situations where common statistical approaches might not be applicable as usual, e.g., dependent records, specification of uncertainty bounds for a return level estimate cannot be made on the basis of the mathematically founded asymptotic theory. The simplified assumption of independent observations usually implies an underestimation of this uncertainty [11.4, 11.13, 11.35]. The estimation of return levels and their uncertainty plays an important role in hydrological engineering and decision making. It forms the basis of setting design values for flood protection buildings like dikes. Since those constructions protect facilities of substantial value or are by themselves costly objects, it is certainly of considerable importance to have appropriate concepts of estimation and uncertainty assessment at hand. Otherwise severe damages, misallocation of public funds, or large claims against insurance companies might be possible. Thus, the approach presented in this contribution focuses on an improvement of common statistical methods used for the estimation of return levels with non-asymptotic bootstrap method.

In the present article, we focus on the block maxima approach and investigate the maximum-likelihood estimator for return levels of autocorrelated run-off records and its uncertainty. In a simulation study, the increase in uncertainty of a return level estimate due to dependence is illustrated. As a result of comparing four strategies based on the bootstrap, we present a concept which explicitly takes the autocorrelation into account. It improves the estimation of confidence intervals considerably relative to those provided by the asymptotic theory. This strategy is based on a semi-parametric bootstrap approach involving a model for the autocorrelation function (ACF) and a resampling strategy from the maxima series of the observations. The approach is validated using a simulation study with an autocorrelated process. Its applicability is exemplified in a case study: we estimate a 100-year return level and a related 95% upper confidence limit under the different assumptions of independent and dependent observations. The empirical run-off series was measured at the gauge Vilsbiburg at the river Große Vils in the Danube catchment.

The paper is organised as follows: Section 11.2 describes the basic theory of the block maxima approach of extreme value statistics and the associated parameter estimation. Section 11.3 illustrates the effect of dependence on the variability of the return level estimator. In Sect. 11.4 the bootstrap strategies are presented including the methodological concepts they require. The performance of the most promising approach is evaluated in Sect. 11.5, followed by a case study in Sect. 11.6. A discussion and conclusions in Sects. 11.7 and 11.8 complete the article. Details regarding specific methods used are deferred to the appendix 11.10.

11.2 Basic Theory

11.2.1 The Generalised Extreme Value Distribution

The pivotal element in extreme value statistics is the three types theorem, discovered by Fisher and Tippett [11.22] and later formulated in full generality by Gnedenko [11.23]. It motivates a family of probability distributions, namely the general extreme value distributions (GEV), as models for block maxima from an observed record, e.g., annual maximum discharge. We denote the maxima out of blocks of size n as M_n. According to the three types theorem, for n large enough the maxima distribution can be approximated by

$$\Pr\{M_n \leq z\} \approx G(z), \tag{11.1}$$

where $G(z)$ is a member of the GEV family (cf. App. 11.10.1).

The quality of the approximation in Eq. (11.1) depends in the first place on the block size n, which in hydrologic applications naturally defaults to one year, $n = 365$. Further influencing factors are the marginal distribution of the observed series and – a frequently disregarded characteristic – its autocorrelation. Fortunately, the three types theorem holds also for correlated records under certain assumptions (cf. App. 11.10.1). The quality of approximation, however, is affected by the correlation as demonstrated in the following.

We compare records of white noise and a simple correlated process (AR[1], cf. Sect. 11.4.3) with the same (Gaussian) marginal distribution. For different block sizes n, we extract 2 000 block maxima from a sufficiently long record. Subsequently, the maxima are modelled with a Gumbel distribution being the appropriate limiting distribution in the Gaussian case [11.21]. We measure the quality of approximation for different n using the negative log-likelihood l (cf. Sect. 11.2.2). Figure 11.1 shows a decreasing negative log-likelihood with increasing block sizes n for the uncorrelated and the correlated record. This implies that the approximation in general ameliorates with block size n. However,

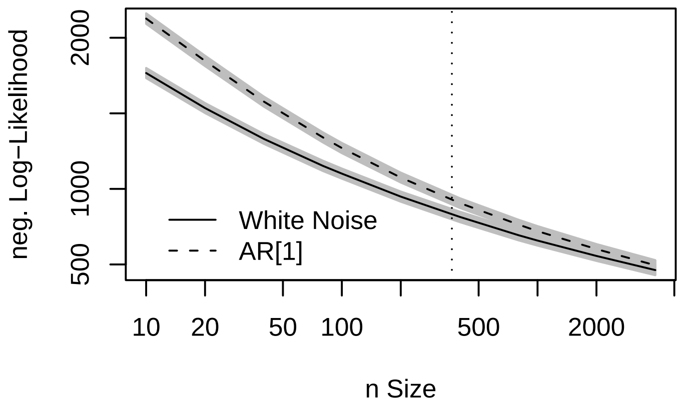

Fig. 11.1. Quality of approximation of a Gumbel fit to 2 000 maxima of realizations of a white noise and an AR[1] process for different block sizes n. The lines connect the means of 1 000 realizations, the shadows mark the mean plus/minus one standard deviation. The vertical line marks a block size of $n = 365$.

for all n the approximation is better for the uncorrelated series than for the AR[1] series. This finding is consistent with dependency reducing the effective number of data points [11.62], which in this case translates into a reduction of effective block size. The difference in approximation between the correlated and the uncorrelated case vanishes with increasing n.

11.2.2 GEV Parameter Estimation

To fully specify the model for the extremes, we estimate the GEV parameters from the data. Estimates can be obtained in several ways: probability weighted moments [11.30, 11.31], maximum likelihood (ML) [11.12, 11.53] or Bayesian methods [11.12, 11.14, 11.54]. These different approaches have advantages and drawbacks which are discussed in e.g., [11.13,11.31] and [11.55]. In the following we focus on ML estimation as the most general method. Within this framework models can be easily extended, for example to non-stationary distributions [11.34].

Let $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})$ be the maximum-likelihood estimate (cf. App. 11.10.2) for the location (μ), scale (σ), and form (ξ) parameter of the GEV. For large block sizes n approximate $(1-\alpha)100\%$ confidence intervals for these estimates can be obtained from the Fisher information matrix I_E as $\hat{\theta}_j \pm z_{\frac{\alpha}{2}}\sqrt{\beta_{j,j}}$; with $\beta_{j,k}$ denoting the elements of the inverse of I_E and $z_{\frac{\alpha}{2}}$ the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution (cf. App. 11.10.2).

The m-year return level can be calculated straight forwardly once the location, scale and shape parameter are estimated. In case of the Gumbel distribution the equation reads

$$\hat{r}_m = \hat{\mu} - \hat{\sigma}\log(y), \tag{11.2}$$

with $y = -\log(1 - \frac{1}{m})$. An approximated confidence interval for $\hat{r}_m$ can be obtained using the delta method described in App. 11.10.2 [11.12].

For maximum-likelihood estimation of the GEV parameters, we use the package `evd` [11.56] for the open source statistical language environment R [11.50][1].

11.3 Effects of Dependence on Confidence Intervals

Annual maxima from river run-off frequently appear uncorrelated from an investigation of the empirical ACF. The left panel in Fig. 11.2 shows the ACF of the annual maxima series from the gauge Vilsbiburg (solid) and of a simulated record (dotted). Both records contain 62 values and their ACF estimates basically do not exceed the 95% significance level for white noise. If a longer series was available, as it is the case for the simulated record, significant autocorrelation of the annual maxima are revealed by the ACF, Fig. 11.2 (right). This implies that considering annual maxima from run-off records *a priori* as uncorrelated can be misleading.

The ML-estimator relies on the assumption of independent observations and is thus, strictly speaking, not correct for dependent observations. The main effect is that standard errors are underestimated if obtained from the Fisher information matrix [11.15]. In the following we illustrate this effect by a Monte-Carlo (MC) simulation study using realizations of a long-range[2] dependent process (FAR[1,d], cf. Sect. 11.4.3) with Hurst exponent $H = 0.75$ (or, equivalently, fractional differencing parameter $d = H - 0.5 = 0.25$). To ameliorate resemblance to a daily run-off series, we transform the Gaussian

[1] Both are freely available from `http://cran.r-project.org`

[2] A process is long-range dependent, if its autocorrelation function is not summable, cf. 11.4.3.

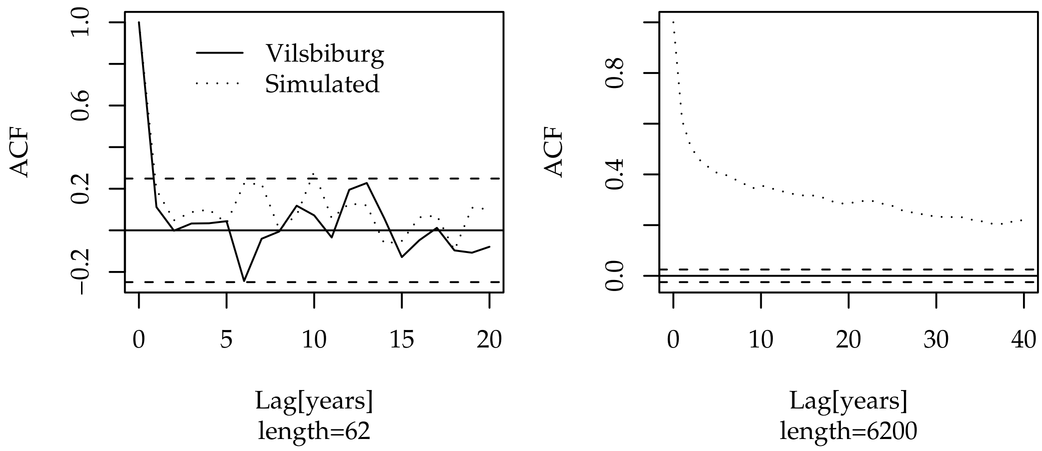

Fig. 11.2. Autocorrelation of the empirical maxima series and a section of same length cut out of the simulated series (left). The right panel shows the Acf of the full simulated maxima series (length=6 200). The 95% significance levels are marked as dashed lines.

series X_t with an exponential function. The resulting record $Z_t = \exp(X_t)$ is then log-normally distributed.

Considering Z_t as 100 years of daily run-off ($N = 36\,500$), we perform an extreme value analysis, i.e. we model the annual maxima series by means of a Gev. Since the marginal distribution is by construction log-normal, we restrict the extreme value analysis to a Gumbel distribution which is the proper limiting distribution in this case. Exemplarily, a 100-year return level is estimated using the Mle (cf. Sect. 11.2.2). Repeating this for 10 000 realizations of Z_t yields a frequency distribution representing the variability of the return level estimator for the Far$[1, d]$ process, shown as histogram in Fig. 11.3 (left panel). An analogous simulation experiment has been carried out for an uncorrelated series with a log-normal distribution (Fig. 11.3, right panel). Both histograms (grey) are compared with the limiting distribution (solid line) of the Mle (Eq. (11.19)) evaluated for an ensemble member with return level estimate close to the ensemble mean. For the uncorrelated series the limiting distribution provides a reasonable approximation in the sense that it roughly recovers the variability of the estimator. In the presence of correlation, the estimators variability is underestimated. This indicates that confidence intervals derived from the Mle's limiting distribution are not appropriate here.

Alternatively, confidence intervals can be obtained using the profile likelihood which is frequently more accurate [11.12]. Table 11.1 compares the upper limits of two-sided confidence intervals for three α-levels obtained using profile likelihood to the limiting distribution and the Monte Carlo simulation. The limits from the profile likelihood are indeed for the correlated and the uncorrelated process closer to the limits of the Monte Carlo ensemble. For the

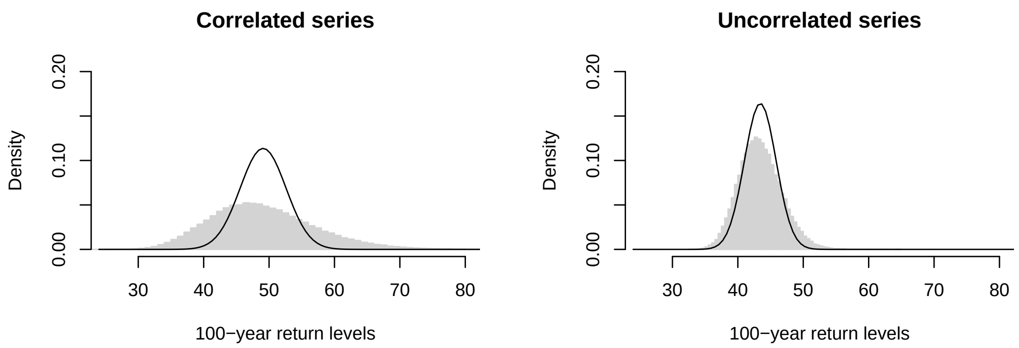

Fig. 11.3. Histogram (grey) of the estimated 100-year return levels from the MC ensemble of 10 000 realization of the FAR$[1, d]$ process with fractional difference parameter $d = 0.25$ (or equivalently $H = 0.75$) and AR parameter $\phi_1 = 0.9$ (left). The right panel shows the result for a white noise process. The realizations contain $N = 36\,500$ data points. The solid line shows a Gaussian density function representing the limiting distribution of the 100-year return level estimator derived from the Fisher information matrix for one ensemble member.

Table 11.1. Upper limits of two-sided confidence intervals for various confidence levels obtained from the asymptotic distribution (Asympt.), the profile likelihood approach (Profile) and the Monte Carlo ensemble (MC).

Uncorrelated Series

Level	Asympt.	Profile	MC
0.68	45.97	46.10	46.69
0.95	48.33	48.90	50.42
0.99	49.84	50.87	53.10

Correlated Series

Level	Asympt.	Profile	MC
0.68	52.72	53.00	56.91
0.95	56.21	57.19	67.26
0.99	58.43	60.12	75.34

correlated process this improvement is not satisfying, since the difference to the Monte Carlo limit is still about 20% of the estimated return level.

To facilitate the presentation in the following, we compare the results from the bootstrap approaches to the confidence intervals obtained using the asymptotic distribution.

11.4 Bootstrapping the Estimators Variance

We discuss non-asymptotic strategies to more reliably assess the variability of the return level estimator. These strategies are based on the bootstrap, i.e. the generation of an ensemble of artificial maxima series. These series are simulated using a model which has been motivated by the data [11.17]. In the given setting we estimate the return levels for each ensemble member and study the resulting frequency distribution.

There are various strategies to generate ensembles. Four of them will be briefly introduced and, as far as necessary, described in the following.

***bootstrap*$_{\mathrm{cl}}$.** The first approach is a classical bootstrap resampling of the maxima [11.17, 11.19], denoted in the following as *bootstrap*$_{\mathrm{cl}}$: one ensemble member is generated by sampling with replacement from the annual maxima series. Autocorrelation is not taken into account here.

***iaaft*$_{\mathrm{d}}$.** We denote the second strategy as *iaaft*$_{\mathrm{d}}$, it makes use of the daily observations. Ensemble members are obtained using the iterative amplitude adjusted Fourier transform (IAAFT) – a surrogate method described in Sect. 11.4.4. The IAAFT generates artificial series (so-called surrogates) preserving the distribution and the correlation structure of the observed daily record. Subsequently, we extract the maxima series to obtain an ensemble member. Linear correlation is thus accounted for in this case.

***bootstrap*$_{\mathrm{fp}}$.** The third strategy is a full parametric bootstrap approach denoted as *bootstrap*$_{\mathrm{fp}}$. It is based on parametric models for the distribution and the autocorrelation function of the yearly maxima. This approach operates on the annual maxima in order to exploit the Fisher-Tippett theorem motivating a parametric model for the maxima distribution.

***bootstrap*$_{\mathrm{sp}}$.** The fourth strategy is a semi-parametric approach, which we call *bootstrap*$_{\mathrm{sp}}$. It similarly uses a parametric model for the ACF of the maxima series, but instead of the GEV we choose a non-parametric model for the distribution.

While the first two strategies are common tools in time series analysis and are well described elsewhere [11.17, 11.52], we focus on describing only the full-parametric and semi-parametric bootstrap strategy.

11.4.1 Motivation of the Central Idea

A return level estimate is derived from an empirical maxima series which can be regarded as a realization of a stochastic process. The uncertainty of a return level estimate depends on the variability among different realizations of this process. As a measure of this variability, we consider the deviation of a realization's empirical distribution function from the true distribution function. If this variability is low, it is more likely to have obtained a good representative for the true maxima distribution from one sample. For a high variability instead, it is harder to get a representative picture of the underlying distribution from one sample. We illustrate this effect using long realizations ($N = 10\,000$) from the correlated and the uncorrelated process introduced in Sect. 11.3. We compare the difference between the distributions $\hat{F}_s(x)$ of a short section ($N = 100$) of a realization and the entire realization's distribution $\hat{F}_0(x)$ by means of the Kolmogorov-Smirnov distance $D = \max_x |\hat{F}_s(x) - \hat{F}_0(x)|$ [11.18]. Smaller distances D indicate a larger similarity between $\hat{F}_s$ and $\hat{F}_0$. Figure 11.4 shows the cumulative distribution function $\hat{F}(D)$ of these distances D for an uncorrelated (circles) and a correlated process (triangles). For the correlated

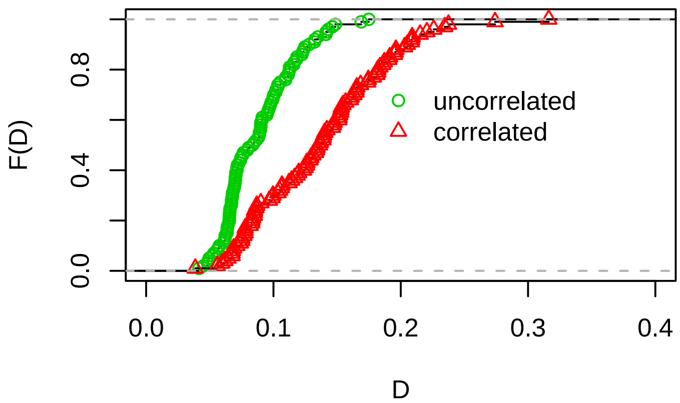

Fig. 11.4. Empirical cumulative distribution function $\hat{F}(D)$ of the Kolmogorov-Smirnov distances D between sections of a 100 year annual maxima series and the entire 10 000 year annual maxima series. For the uncorrelated record (◯) distances D are located at smaller values than for the correlated record (△).

process, we find a distribution of distances D located at larger values. This implies that the sections are more diverse in their distribution. Thus for correlated processes the variability in short realization's maxima distribution is larger than for the uncorrelated process. Realizations of correlated processes are therefore not as likely to yield as representative results for the underlying distribution as a comparable sample of an uncorrelated process.

Since the variability of the return level estimator is a result of the variability of realization's maxima distribution, we employ this illustrative example and study the estimator's variability among sections of a long record. Ideally, the properties of this long record should be close to the underlying properties of the process under consideration. This requires a satisfying model for the maxima series' distribution and the autocorrelation function. In the approach pursued here, we initially provide two separate models for the two characteristics. Realizations of these two models are then combined to obtain one realization satisfying both, the desired distribution and the Acf.

11.4.2 Modelling the distribution

The aim of modelling the distribution is to provide means for generating realizations used in a later step of the bootstrap procedure.

For the semi-parametric approach, the distribution of the maxima is modelled by the empirical cumulative distribution function from the observed series. This means realizations from this model can be obtained simply by sampling with replacement from the observed maxima series [11.17].

The full parametric approach exploits the Fisher-Tippett theorem for extreme values (Sect. 11.2.1). It uses the parametric GEV family as a model for the maxima distribution. Realizations are then obtained directly by sampling from the parametric model fitted to the empirical maxima series.

11.4.3 Modelling the ACF

At this point, we are mainly interested in an adequate representation of the empirical ACF. Such a representation can be achieved by modelling the series under investigation with a flexible class of linear time series models. We do not claim that these models are universally suitable for river run-off series. However, together with a static non-linear transformation function and a deterministic description of the seasonal cycle, they capture the most dominant features regarding the stochastic variability. Especially, an adequate representation of the ACF can be achieved, which is the objective of this undertaking. These models are obviously not adequate for studying, e.g., the effect of changing precipitation patterns on run-off or other external influences. This is, however, not in the focus of this paper.

ARMA Processes . A simple and prevailing correlated stochastic process is the autoregressive process of first order (AR[1]) process (or red noise) frequently used in various geophysical contexts, e.g., [11.26, 11.41, 11.63]. For a random variable X_t, it is a simple and intuitive way to describe a correlation with a predecessor in time by

$$X_t = \phi_1 X_{t-1} + \sigma_\eta \eta_t, \tag{11.3}$$

with η_t being a Gaussian white noise process ($\eta_t \sim W\mathcal{N}(0,1)$) and ϕ_1 the lag-one autocorrelation coefficient. This approach can be extended straightforwardly to include regressors X_{t-k} with lags $1 \leq k \leq p$ leading to AR[p] processes allowing for more complex correlation structures, including oscillations. Likewise lagged instances of the white noise process $\psi_l \eta_{t-l}$ (moving average component) with lags $1 \leq l \leq q$ can be added leading to ARMA[p,q] processes, a flexible family of models for the ACF [11.9, 11.10]. There are numerous applications of ARMA models, also in the context of river-runoff, e.g., [11.7, 11.27, 11.59].

FARIMA Processes . Since ARMA processes are short-range dependent, i.e. having a summable ACF [11.4], long-range dependence, which is frequently postulated for river run-off [11.40, 11.43, 11.47], cannot be accounted for. It is thus desirable to use a class of processes able to model this phenomenon. Granger [11.24] and Hosking [11.29] introduced fractional differencing to the concept of linear stochastic models and therewith extended the ARMA family to fractional autoregressive integrated moving average (FARIMA) processes. A convenient formulation of such a long-range dependent FARIMA process X_t is given by

$$\Phi(B)(1-B)^d X_t = \Psi(B)\eta_t, \tag{11.4}$$

with B denoting the back-shift operator ($BX_t = X_{t-1}$), $\eta_t \sim W\mathcal{N}(0, \sigma_\eta)$ a white noise process, and $d \in \mathbb{R}$ the fractional difference parameter. The latter is related to the Hurst exponent, which is frequently used in hydrology by $H = d + 0.5$ [11.4]. The autoregressive and moving average components are described by polynomials of order p and q

$$\Phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i, \quad \Psi(z) = 1 + \sum_{j=1}^{q} \psi_j z^j, \tag{11.5}$$

respectively. In practice, the fractional difference operator $(1-B)^d$ has to be expanded in a power series, cf. [11.24]. A FARIMA$[p, d, q]$ process is stationary if $d < 0.5$ and all solutions of $\Phi(z) = 0$ in the complex plane lie outside the unit circle. It exhibits long-range dependence or long-memory for $0 < d < 0.5$. Processes with $d < 0$ are said to possess intermediate memory, in practice this case is rarely encountered, it is rather a result of 'over-differencing' [11.4]. Recently, the FARIMA model class has been used as a stochastic model for river run-off, e.g., [11.4, 11.33, 11.40, 11.43, 11.44]. Also self-similar processes are sometimes used in this context. The latter provide a simple model class, but are not as flexible as FARIMA processes and are thus appropriate only in some specific cases. FARIMA models, however, can be used to model a larger class of natural processes including self-similar processes. (For a comprehensive overview of long-range dependence and FARIMA processes refer to [11.4, 11.46] and references therein.)

Parameter Estimation . The model parameters $(\phi_1, ..., \phi_p, d, \psi_1, ..., \psi_q)$ are estimated using Whittle's approximation to the ML-estimator [11.4]. It operates in Fourier space – with the spectrum being an equivalent representation of the autocorrelation function [11.48] – and is computationally very efficient due to the use of the fast Fourier transform. Therefore, the Whittle estimator is especially useful for long records where exact MLE is not feasible due to computational limits.

The model orders p and q are *a priori* unknown and can be determined using the Hannan-Quinn Information Criterion HIC which is advocated for FARIMA processes:

$$\text{HIC} = N \log \hat{\sigma}_\eta^2 + 2c \log\log N(p+q+1), \tag{11.6}$$

with $c > 1$, $\hat{\sigma}_\eta^2$ the ML estimate of the variance of the driving noise η_t and $p+q+1$ being the number of parameters [11.5, 11.6, 11.25]. We choose the model order p and q such that the HIC takes a minimum.

Indirect Modelling of the Maxima Series' ACF. Modelling the ACF of a run-off maxima series is usually hampered by the shortness of the records. For a short time series it is often difficult to reject the hypothesis of independence, cf. Sec. 11.3. To circumvent this problem, we model the daily series and assume that the resulting process adequately represents the daily series' ACF. It is used to generate long records whose extracted maxima series are again modelled with a FARIMA$[p,d,q]$ process. These models are then considered as adequate representatives of the empirical maxima series ACF. This indirect approach of modelling the maxima series' ACF relies on the strong assumption that the model for the daily series and also it's extrapolation to larger time scales is adequate.

Modelling a Seasonal Cycle . The seasonal cycle found in a daily river run-off series has a fixed periodicity of one year. It can be modelled as a deterministic cycle $C(t)$ which is periodic with period T (1 year): $C(t+T) = C(t)$. Combined with the stochastic model $X(t)$ this yields the following description:

$$Y(t) = C(t) + X(t). \tag{11.7}$$

In the investigated case studies $C(t)$ is estimated by the average yearly cycle obtained by averaging the run-off $Q(t)$ of a specific day over all M years, i.e. $\hat{C}(t) = 1/M \sum_j Q(t+jT), t \in [1,T]$, cf. [11.28].

Including a Static Non-Linear Transformation Function. River run-off is a strictly positive quantity and the marginal distribution is in general positively skewed. This suggests to include an appropriate static non-linear transformation function in the model. Let $Z = T(Y)$ denote the Box-Cox transformation (cf. 11.10.3, [11.8]) of the random variable Y (Eq. 11.7). Then Z is a positively skewed and strictly positive variable, suitable to model river run-off. One can think of this static transformation function as a change of the scale of measurement. This transformation has been suggested for the modelling of river run-off by Hipel and McLeod [11.28].

The full model for the ACF can be written as

$$\Phi(B)(1-B)^d X_t = \Psi(B)\eta_t \tag{11.8}$$
$$Y(t) = C(t) + X(t) \tag{11.9}$$
$$Z(t) = T(Y). \tag{11.10}$$

Simulation of FARIMA Processes. Several algorithms are known to simulate data from a FARIMA process (for an overview refer to [11.1]). Here, we use a method based on the inverse Fourier transform described in [11.60]. It was originally proposed for simulating self-similar processes but can be straightforwardly extended to Farima processes.[3]

11.4.4 Combining Distribution and Autocorrelation

Having a model for the distribution and for the Acf we can generate realizations, i.e. a sample $\{W_i\}_{i=1,\dots,N}$ from the distribution model and a series $\{Z_i\}_{i=1,\dots,N}$ from the Farima model including the Box-Cox transformation and, if appropriate, the seasonal cycle. To obtain a time series $\{Q_i\}_{i=1,\dots,N}$ with distribution equal to the one of $\{W_i\}_{i=1,\dots,N}$ and Acf comparable to that of $\{Z_i\}_{i=1,\dots,N}$, we employ the iterative amplitude adjusted Fourier transform (Iaaft).

The Iaaft was developed by Schreiber and Schmitz [11.51] to generate surrogate time series used in tests for nonlinearity [11.58]. The surrogates are generated such that they retain the linear part of the dynamics of the original time series including a possible non-linear static transfer function. This implies that the power spectrum (or Acf, equivalently) and the frequency distribution of values are conserved. The algorithm basically changes the order of the elements of a record in a way that the periodogram stays close to a desired one [11.52].

Besides using the Iaaft on the daily series to generate an ensemble of surrogates (denoted as $iaaft_{\mathrm{d}}$), we employ this algorithm also to create records $\{Q_i\}_{i=1,\dots,N}$ with a periodogram prescribed by a series $\{Z_i\}_{i=1,\dots,N}$ and a frequency distribution coming from $\{W_i\}_{i=1,\dots,N}$.

11.4.5 Generating Bootstrap Ensembles

With the described methods we are now able to build a model for the distribution and Acf of empirical maxima series and combine them to obtain long records. This provides the basis of the full parametric bootstrap ensembles $bootstrap_{\mathrm{fp}}$ and the semi-parametric ensemble $bootstrap_{\mathrm{sp}}$.

In detail, the strategies to obtain the ensembles $bootstrap_{\mathrm{fp}}$ and $bootstrap_{\mathrm{sp}}$ can be outlined as follows:

1. Model the correlation structure of the maxima
   a) If necessary, transform the daily run-off data to follow approximately a Gaussian distribution using a log or Box-Cox transform [11.8, 11.28], cf. 11.10.3.
   b) Remove periodic cycles (e.g., annual, weekly).

[3] An R package with the algorithms for the Farima parameter estimation (based on the code from Beran [11.4]), the model selection (Hic) and the simulation algorithm can be obtained from the author.

c) Model the correlation structure using a FARIMA$[p, d, q]$ process, select the model orders p and q with HIC (Sect. 11.4.3).
d) Generate a long series from this model ($N_{\text{long}} \gtrsim 100 N_{\text{data}}$).
e) Add the periodic cycles from 1a). The result is a long series sharing the spectral characteristics, especially the seasonality with the empirical record.
f) Extract the annual maxima series.
g) Model the correlation structure of the simulated maxima series using a FARIMA$[p_{\max}, d, q_{\max}]$ process, with orders $p_{\max}$ and $q_{\max}$ selected with HIC.

2. Model the distribution of the maxima according to the approach used
   bootstrap$_{\text{fp}}$: Estimate the parameters of a GEV model from the empirical maxima series using MLE (Sect. 11.2.2).
   bootstrap$_{\text{sp}}$: Use the empirical maxima distribution as model.
3. Generate an ensemble of size N_{ensemble} of maxima series with length $N_{\max}$ with correlation structure and value distribution from the models built in 1 and 2:
   a) Generate a series $\{Z_i\}$ with the FARIMA$[p_{\max}, d, q_{\max}]$ model from step 1f) of length $N_{\text{ensemble}} N_{\max}$ back-transform according to 1a)[4].
   b) Generate a sample $\{W_i\}$ with length $N_{\text{ensemble}} N_{\max}$
      bootstrap$_{\text{fp}}$ from the GEV model specified in step 2a).
      bootstrap$_{\text{sp}}$ from sampling with replacement from the empirical maxima series.
   c) By means of IAAFT $\{W_i\}$ is reordered such that its correlation structure is similar to that of $\{Z_i\}$. This yields the run-off surrogates $\{Q_i\}_{i=1,\ldots,N_{\text{ensemble}} N_{\max}}$.
   d) Splitting $\{Q_i\}_{i=1,\ldots,N_{\text{ensemble}} N_{\max}}$ into blocks of size $N_{\max}$ yields the desired ensemble.

Estimating the desired return level from each ensemble member as described in Sect. 11.2.2 yields a frequency distribution of return level estimates which can be used to assess the variability of this estimator.

11.5 Comparison of the Bootstrap Approaches

A comparison of the four different bootstrap approaches *bootstrap*$_{\text{cl}}$, *iaaft*$_{\text{d}}$, *bootstrap*$_{\text{fp}}$, and *bootstrap*$_{\text{sp}}$, is carried out on the basis of a simulation study. We start with a realization of a known process, chosen such that its correlation structure as well as its value distribution are plausible in the context of river

[4] Instead of back-transforming here, one can Box-Cox transform the outcome of step 3b) combine the results with IAAFT as in step 3c) and back-transform afterwards. This procedure turned out to be numerically more stable.

run-off. Here, this is a Farima process similar to those used in [11.43] with a subsequent exponential transformation to obtain a log-normal distribution. This process is used to generate a Monte Carlo ensemble of simulated daily series. For each realization we extract the maxima series and estimate a 100-year return level. We obtain a distribution of 100-year return level estimates which represent the estimators variability for this process. In the following, this distribution is used as a reference to measure the performance of the bootstrap approaches. A useful strategy should reproduce the distribution of the return level estimator reasonably well.

We now take a representative realization out of this ensemble and consider it as a record, we possibly could have observed. On the basis of this "observed" series. From this record, we generate the four bootstrap ensembles according to the approaches presented in Sect. 11.4.5. The resulting four frequency distributions of the 100-year return level estimates are then compared to the distribution of the reference ensemble and to the asymptotic distribution of the Ml-estimator.

11.5.1 Monte Carlo Reference Ensemble

We simulate the series for the reference ensemble with a Farima$[1, d, 0]$ process with parameters $d = 0.25$ (or $H = 0.75$), $\phi_1 = 0.9$, variance $\sigma^2 \approx 1.35$, Gaussian driving noise η, and length $N = 36\,500$ (100 years of daily observations). The skewness typically found for river run-off is achieved by subsequently transforming the records to a log-normal distribution. To resemble the procedure of estimating a 100-year return level we extract the annual maxima and estimate a 0.99 quantile using a Gumbel distribution as parametric model. From an ensemble of 100 000 runs, we obtain a distribution of 100-year return levels (0.99 quantiles) serving as a reference for the bootstrap procedures.

11.5.2 The Bootstrap ensembles

Taking the "observed" series, we generate the four bootstrap ensembles according to Sect. 11.4.5. Since *iaaft*$_{\mathrm{d}}$ and *bootstrap*$_{\mathrm{cl}}$ are well described in the literature, we focus on the semi-parametric and full-parametric approach.

Following the outline in Sect. 11.4.5, we start modelling the correlation structure of the "observed" series using a Farima process as described in Sect. 11.4.3. As a transformation, we choose a log-transform as a special case of the Box-Cox. We treat the process underlying the sample as unknown and fit Farima$[p, d, q]$ models with $0 \leq q < p \leq 4$. Figure 11.5 shows the result of the model selection criterion Hic (Eq. (11.6)) for these fits. The Farima$[1, d, 0]$ with parameters and asymptotic standard deviation: $d = 0.250 \pm 0.008, \phi = 0.900 \pm 0.005, \sigma_\eta^2 = 0.0462$ yields the smallest Hic and is chosen to model the record. Thus, the proper model structure is recovered and step 1b) is completed.

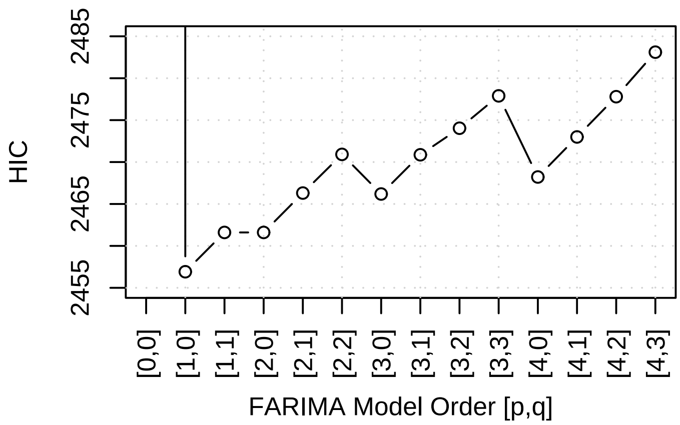

Fig. 11.5. Comparison of the HIC of different FARIMA$[p, d, q]$ models. Smaller values indicate a better model. On the abscissa different model orders $[p, q]$ are plotted ($0 \leq q \leq 3$, $0 \leq p \leq 4$ and $q \leq p$). The model orders are discrete, the lines connecting the discrete orders $[p, q]$ are drawn to enhance clarity.

With this model we generate an artificial series longer than the original one ($N_{\text{long}} = 100 N_{\text{data}}$) according to step 1c). The extracted annual maxima series contains $N_{\max} = 10\,000$ data points. Since we do not expect the ACF of this maxima series to require a more complex model (in terms of the number of parameters) than the daily data, we use orders $p \leq 1$ and $q = 0$, leaving only two models. The FARIMA$[0, d, 0]$ has a slightly smaller HIC value (HIC$_{[0,d,0]}$=21792.06) than FARIMA$[1, d, 0]$ (HIC$_{[1,d,0]}$=21792.40) and will thus be chosen in the following, step 1f). The resulting parameters are $d = 0.205 \pm 0.008$ and $\sigma_\eta^2 = 0.535$.

Having modelled the correlation structure we now need a representation for the distribution. For the full-parametric bootstrap ensemble *bootstrap*$_{\text{fp}}$, we get back to the “observed” series and model the annual maxima with a parametric Gumbel distribution, step 2a). This results in ML-estimates for the location and scale parameters: $\mu = 10.86, \sigma = 8.35$. Since the semi-parametric approach *bootstrap*$_{\text{sp}}$ does not need a model but uses a classical bootstrap resampling from the empirical annual maxima series, we now can generate the desired bootstrap ensembles *bootstrap*$_{\text{fp}}$ and *bootstrap*$_{\text{sp}}$ both with 1 000 members according to step 3.

Figure 11.6 compares the frequency distributions of estimated return levels from the four bootstrap ensembles to the reference distribution (grey filled) and to the asymptotic distribution of the ML-estimator (dotted). The left plot shows the result of the *bootstrap*$_{\text{cl}}$ (solid) and the *iaaft*$_{\text{d}}$ (dashed) ensembles. While *bootstrap*$_{\text{cl}}$ accounts for more variability than the asymptotic distribution, *iaaft*$_{\text{d}}$ exhibits less variability, although it takes autocorrelation of the daily data into account. This might be due to the fact that the records in the *iaaft*$_{\text{d}}$ ensemble consists of exactly the same daily run-off values arranged in a different order. While this allows for some variability on the daily scale, an annual maxima series extracted from such a record is limited to a much smaller set of possible values. Since the temporal order of the maxima series does not influence the return level estimation, the variability of the estimates is reduced.

The right panel in Fig. 11.6 shows the result from the *bootstrap*$_{\text{sp}}$ (solid) and the *bootstrap*$_{\text{fp}}$ (dashed) ensembles. The latter strategy is slightly better than the nonparametric bootstrap resampling but still yields a too narrow distribution. In contrast, the result from the *bootstrap*$_{\text{sp}}$ ensemble gets very close to the reference ensemble. Thus, this approach is a promising strategy to improve the uncertainty analysis of return level estimates and is studied in more detail in the following section.

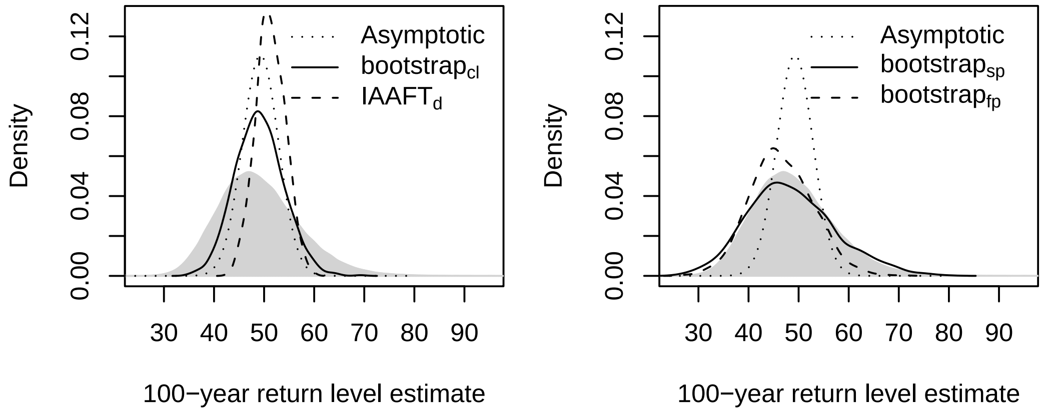

Fig. 11.6. Comparison of the result for different bootstrap ensembles to the MC reference ensemble (grey area) and the asymptotic distribution of the ML-estimator (dotted). The bootstrap ensembles consist each of 1 000 members. The left plot shows the results of the nonparametric bootstrap resampling and the daily IAAFT surrogates. The full parametric and semi-parametric bootstrap strategies are shown in the right plot.

11.5.3 Ensemble Variability and Dependence on Ensemble Size

We investigate the potential of the semi-parametric bootstrap approach by studying its inter-ensemble variability and the dependence on ensemble size. This can be achieved by performing an extensive simulation study, i.e. generating different sets of $bootstrap_{\mathrm{sp}}$ ensembles, each set containing 100 ensembles of a fixed size. We are interested in the variability of the ensemble runs within one set of fixed size and, as well as in the effect of the ensemble size. The ensemble size varies between $N_{\mathrm{ensemble}} = 50$ and $N_{\mathrm{ensemble}} = 6\,000$. To facilitate the representation of the result, we do not consider the entire distribution of the return level estimates for each ensemble, but rather 5 selected quantiles with relative frequencies of 5%, 25%, 50%, 75%, and 95%. Figure 11.7 shows these quantiles estimated from the ensembles for different ensemble sizes as grey dots. Due to the large number of grey dots, they cannot be perceived as individual dots but rather as grey clouds.

The variability of the quantile estimates decreases with increasing ensem-

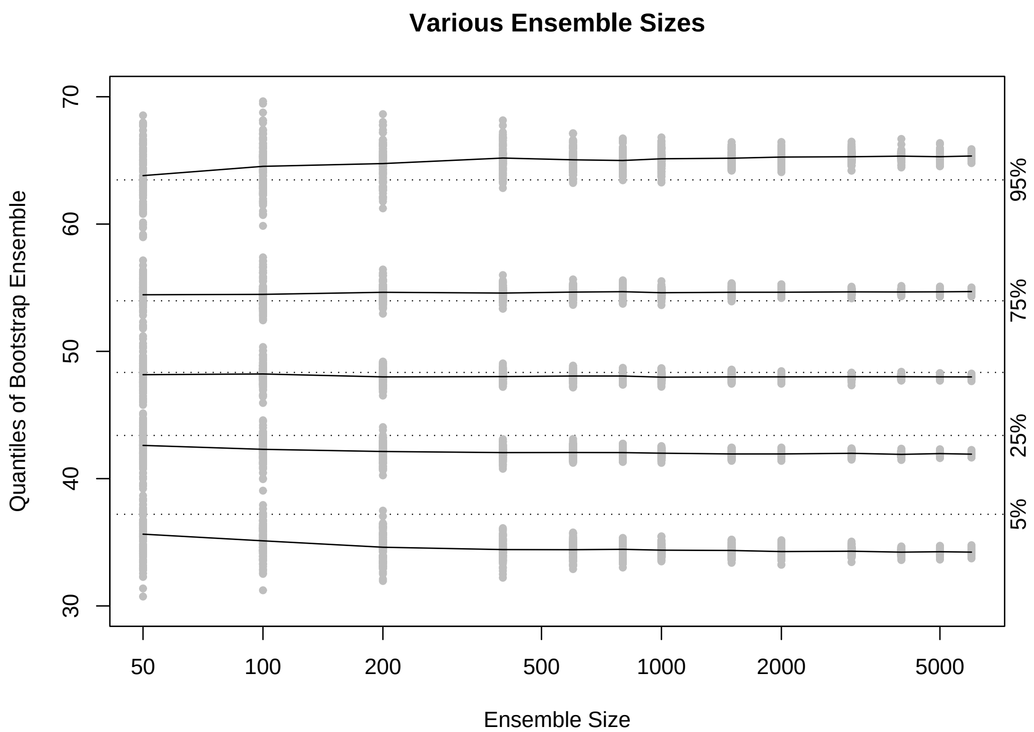

Fig. 11.7. The quantiles of the semi-parametric bootstrap ensembles of different size. 100 ensembles of the same size are grouped in a set. The 5%, 25% 50% 75% and 95% quantiles of each ensemble in a set is marked with a grey dot. This results in grey areas representing the variability within a set. The solid lines connect the sets' mean values for each quantile. The quantiles from the reference ensemble are represented as a dotted line.

ble size indicated by the convergence of the grey clouds for each of the five quantiles. Consequently, the ensemble size should be chosen according to the accuracy needed. For small sizes, the means of the selected quantiles are close to the values from the reference ensemble, especially for the three upper quantiles. With an increasing size, difference to the reference ensemble increases for the extreme quantiles until they stagnate for ensembles with more than about 2000 members. For the 5% and 95% quantiles, the difference between the bootstrap and the Monte Carlo is less than 6% of the return-levels estimate.

11.6 Case Study

To demonstrate the applicability of the suggested semi-parametric bootstrap approach (*bootstrap*$_{\mathrm{sp}}$), we exemplify the strategy with a case study. We consider the run-off record from the gauge Vilsbiburg at the river Große Vils in the Danube River catchment. Vilsbiburg is located in the south-east of Germany about 80km north-east of Munich. The total catchment area of this gauge extends to 320km^2. The mean daily run-off has been recorded from 01/11/1939 to 07/01/2002 and thus comprises $N_{\mathrm{years}} = 62$ full years or $N = 22\,714$ days. The run-off averaged over the whole observation period is about 2.67 m^3/s.

Extreme Value Analysis . First, we perform an extreme value analysis as described in Sect. 11.2.2, i.e. extracting the annual maxima and determining the parameters of a GEV distribution by means of ML-estimation. In order to test whether the estimated shape parameter $\hat{\xi} = 0.04$ is significantly different from zero, we compare the result to a Gumbel fit using the likelihood-ratio test [11.12]. With a p-value of $p = 0.74$ we cannot reject Gumbel distribution as a suitable model on any reasonable level. The resulting location and scale parameters with asymptotic standard deviation are $\mu = (28.5 \pm 2.0)\mathrm{m}^3/\mathrm{s}$ and $\sigma = (15.0 \pm 1.5)\mathrm{m}^3/\mathrm{s}$. The associated quantile and return level plots are shown in Fig. 11.8 together with their 95% asymptotic confidence limits.

According to Eq. (11.2) we calculate a 100-year return level ($m = 100$) and use the delta method (11.21) to approximate a standard deviation under the hypothesis of independent observations: $r_{100} = 97.7 \pm 7.9$.

Modelling the ACF of the Daily Series. In the second step, we model the correlation structure which requires preprocessing of the run-off series: To get closer to a Gaussian distribution, a Box-Cox transformation (sect. 11.10.3) is applied to the daily run-off. The parameter $\lambda = -0.588$ is chosen such that the unconditional Gaussian likelihood is maximised. Subsequent to this static transformation, we calculate the average annual cycle in the mean and the variance, cf. Sect. 11.4.3. The cycle of in the mean is subtracted from the respective days and the result is accordingly divided by the square root of

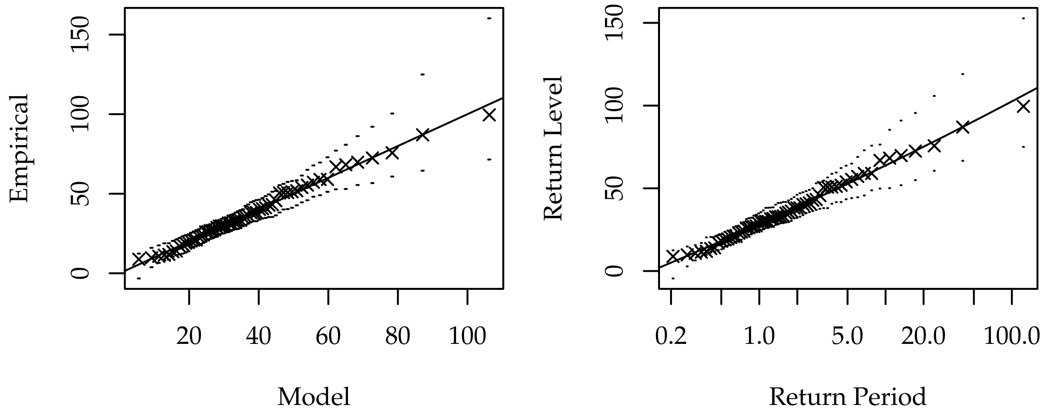

Fig. 11.8. Result of the ML-estimation of the Gumbel parameters for the Vilsbiburg yearly maxima series compared to the empirical maxima series in a quantile plot (left panel) and return level plot (right panel).

the variance cycle. As we find also indication for a weekly component in the periodogram, we subtract this component analogously. The Box-Cox transformation as well as the seasonal filters have been suggested by Hipel and McLeod [11.28] for hydrological time series. Although the proposed estimates of the periodic components in mean and variance are consistent, especially the estimates of the annual cycle exhibits a large variance. Several techniques such as STL [11.11] are advocated to obtain smoother estimates. Studying those periodic components is, however, not in the focus of this paper.

Figure 11.9 shows a comparison of the transformed and mean adjusted daily run-off record to a Gaussian distribution in form of a density plot (left panel) and a plot of empirical versus theoretical quantiles (right panel). The transformed distribution is much less skewed than the original one. Differences to a Gaussian are mainly found in the tails.

We now fit FARIMA$[p, d, q]$ models of various orders with $0 \leq q \leq 5$, $0 \leq p \leq 6$ and $q \leq p$ and compare the HIC of the different models in Fig. 11.10(left). The smallest value for the HIC is obtained for the FARIMA$[3, d, 0]$ process, which is thus chosen to model the autocorrelation of the daily run-off. The parameters estimated for this process with their asymptotic standard deviation are: $d = 0.439 \pm 0.016, \phi_1 = 0.415 \pm 0.017, \phi_2 = -0.043 \pm 0.007, \phi_3 = 0.028 \pm 0.008$ and $\sigma_\eta^2 = 0.2205$. Using the goodness-of-fit test proposed by Beran [11.4], we obtain a p-value of $p = 0.015$. The model thus cannot be rejected on a 1% level of significance. The result of this fit is shown in the spectral domain in Fig. 11.10(right).

Modelling the ACF of the Maxima Series. Using this model, a long series is generated with $N_{\mathrm{long}} = 100 N_{\mathrm{data}}$. This simulated series is partially back-transformed: the overall mean and the seasonal cycles in mean and vari-

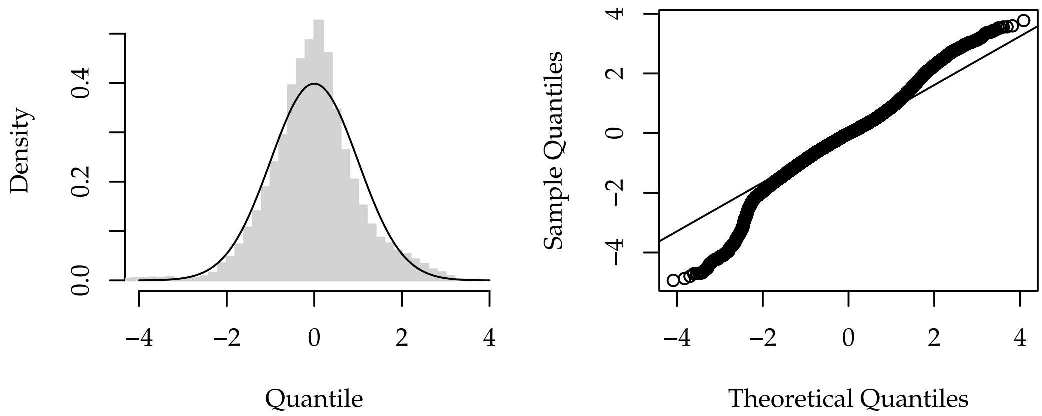

Fig. 11.9. Comparison of the Box-Cox transformed, deseasonalised and mean adjusted run-off values to a Gaussian distribution. Histogram (left, grey) with a density estimate (solid line) and quantile plot (right).

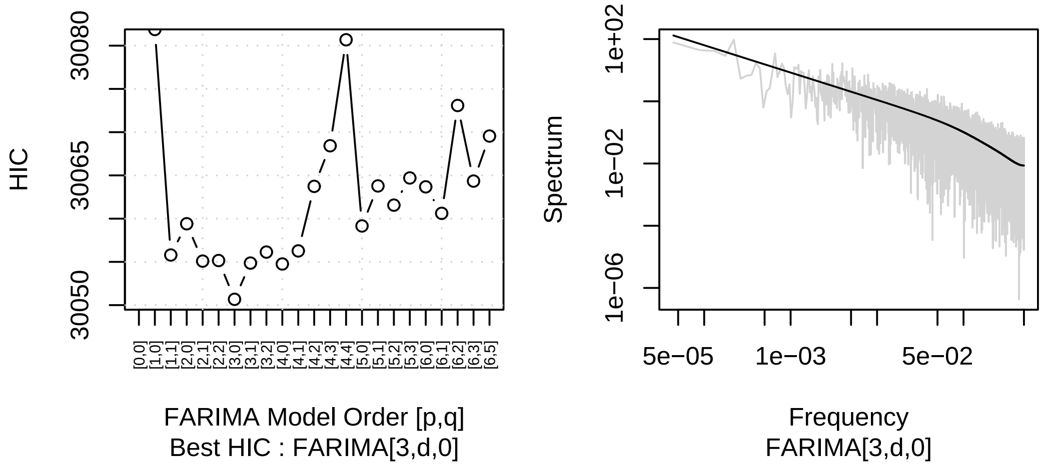

Fig. 11.10. Model selection for the Box-Cox transformed and deseasonalised daily run-off with mean removed. The spectral density of the model with smallest Hic is shown in double logarithmic representation together with the periodogram (grey) of the empirical series (right panel).

ance are added. Note, that the Box-Cox transform is not inverted in this step. From the resulting record, we extract the annual maxima series. Figure 11.2 (left panel) shows the Acf of the original maxima series (solid) and compares it to the Acf of a section of the same length cut out of the maxima series gained from the simulated run (dotted). The original series has been Box-Cox transformed as well to achieve a comparable situation. The autocorrelation basically fluctuates within the 95% significance level (dashed) for a white noise. However, the full maxima series from the simulated run of 6 200 data points exhibits prominent autocorrelation (Fig. 11.2, right panel). This indicates that, although an existing autocorrelation structure is not necessarily visible in a short sample of a process, it might still be present. Accounting for this dependence improves the estimation of confidence limits.

In the next step, we model the correlation structure of the maxima series with a FARIMA process. Again, we do not expect this series being more adequately modelled by a more complex process than the daily data. We use HIC to choose a model among orders $p_{\max} \leq p = 3, q_{\max} \leq q = 1$. The smallest values for the HIC is attained for a FARIMA$[0, d, 0]$ model with $d = 0.398 \pm 0.010$. The goodness-of-fit yields a p-value of $p = 0.319$ indicating a suitable model.

Combining Distribution and ACF. Having the model for the ACF of the maxima series, we are now able to generate a bootstrap ensemble of artificial data sets according to the semi-parametric strategy *bootstrap*$_{\text{sp}}$ as described in Sect. 11.4.5. We use IAAFT to combine the results of the FARIMA simulation with the Box-Cox transformed resampled empirical maxima series. In the last step the ensemble is restored to the original scale of measurement by inverting the Box-Cox transform we started with.

Calculating the Confidence Limit. Subsequently, the 100-year return level $\hat{r}^*$ (or 0.99 quantile) is estimated for each ensemble member yielding the desired distribution for the 100-year return level estimator shown in Fig. 11.11. From this distribution we obtain an estimate for a $(1-\alpha)\%$ one-sided upper confidence limit r^α using order statistics. r^α is calculated from order statistics as $r^\alpha = \hat{r}_{100} - (\hat{r}^*_{(N+1)\alpha} - \hat{r}_{100})$ [11.17], where the $\hat{r}^*_i$ are sorted in ascending order. With an ensemble size $N_{\text{ensemble}} = 9\,999$ we ensure $(N+1)\alpha$ being an integer for common choices of α.

To facilitate the comparison of the 95% confidence limits obtained from the bootstrap ensemble ($r^{0.95}_{\text{boot}} \approx 148\text{m}^3/\text{s}$) and the asymptotic distribution ($r^{0.95}_{\text{asymp}} \approx 110\text{m}^3/\text{s}$) they are marked as vertical lines in Fig. 11.11. The bootstrap 95% confidence level $r^{0.95}_{\text{boot}}$ clearly exceeds the quantile expected from the asymptotic distribution confirming the substantial increase in uncertainty due to dependence. Furthermore, the tails of the bootstrap ensemble decay slower than the tails of the asymptotic distribution. The interpretation of such a confidence level is the following: In 95% of 100-year return level estimates the expected ("true") 100-year return level will not exceed the 95% confidence limit.

11.7 Discussion

The approach is presented in the framework of GEV modelling of annual maxima using maximum likelihood. The concept can also be applied in the context of other models for the maxima distribution (e.g., log-normal) or also different parameter estimation strategies (e.g., probability weighted moments). Further-

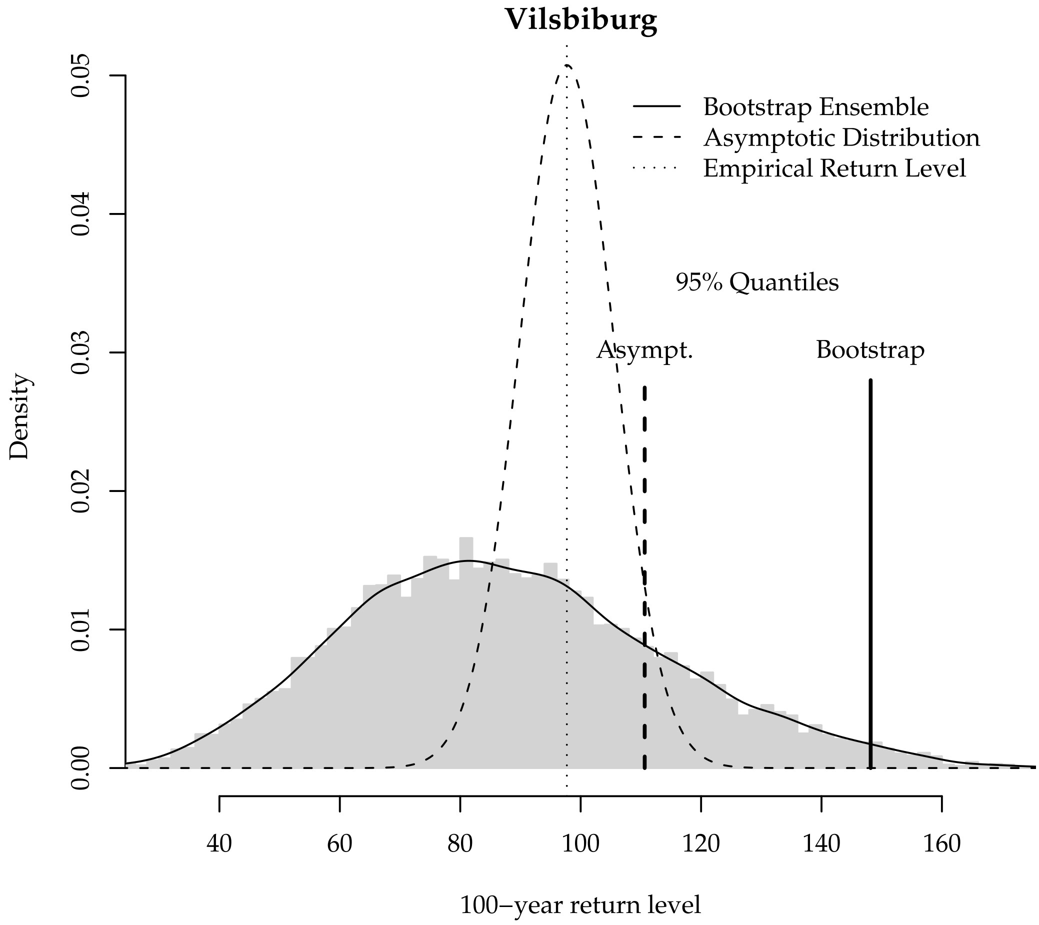

Fig. 11.11. Frequency distribution of the 100-year return level estimates from the *bootstrap*$_{\text{sp}}$ ensemble with 9 999 members for Vilsbiburg as histogram (grey) and density estimate (solid line) compared to the asymptotic distribution of the ML estimator derived from the Fisher information matrix (dashed). The 100-year return level estimate from the empirical maxima series is marked as dotted vertical line. The 95% quantiles of the asymptotic and bootstrap distributions are indicated as dashed and solid vertical lines, respectively.

more, it is conceivable to extend the class of models describing the dependence to FARIMA models with dependent driving noise (FARIMA-GARCH [11.20]) or seasonal models [11.40, 11.44].

The modelling approach using FARIMA[p,d,q] models and a subsequent adjustment of the values has been investigated in more detail by [11.61]. Using simulation studies, it was demonstrated that the combination of FARIMA models and the IAAFT is able to reproduce also other characteristics of time series then the distribution and power spectrum. Also the increment distribution and structure functions for river run-off are reasonably well recovered.

In the approach described, we obtain a model for the ACF of the maxima series only with the help of a model of the daily series. The longer this daily series is the more reliable the model will be. Regarding the uncertainty of return level estimates, we might consider this approach with the assumption of long-range dependence as complementary to the assumption of independent maxima. The latter assumption yields the smallest uncertainty while, the assumption of long-range dependence yields larger confidence intervals which can be considered as an upper limit of uncertainty. The actual detection of long-range dependence for a given time series is by no means a trivial problem [11.41] but it is not in the focus of this paper.

It is also possible to include available annual maxima in the procedure for periods where daily series have not been recorded. This enhances the knowledge of the maxima distribution but cannot be used for the modelling of the Acf.

11.8 Conclusion

We consider the estimation of return levels from annual maxima series using the Gev as a parametric model and maximum likelihood (Ml) parameter estimation. Within this framework, we explicitly account for autocorrelation in the records which reveals a substantial increase in uncertainty of the flood return level estimates. In the standard uncertainty assessment, i.e. the asymptotic confidence intervals based on the Fisher information matrix or the profile likelihood, autocorrelations are not accounted for. For long-range dependent processes, this results in uncertainty limits being too small to reflect the actual variability of the estimator. On the way to fill this gap, we study and compare four bootstrap strategies for the estimation of confidence intervals in the case of correlated data. This semi-parametric bootstrap strategy outperforms the three other approaches. It showed promising results in the validation study using an exponentially transformed Farima[1,d,0] process. The main idea involves a resampling approach for the annual maxima and a parametric model (Farima) for their autocorrelation function. The combination of the resampling and the Farima model is realized with the iterative amplitude adjusted Fourier transform, a resampling method used in nonlinearity testing. The results of the semi-parametric bootstrap approach are substantially better than those based on the standard asymptotic approximation for Mle using the Fisher information matrix. Thus this approach might be of considerable value for flood risk assessment of water management authorities to avoid floods or misallocation of public funds. Furthermore, we expect the approach to be applicable also in other sectors where an extreme value analysis with dependent extremes has to be carried out.

The practicability is illustrated for the gauge Vilsbiburg at the River Vils in the Danube catchment in southern Germany. We derived a 95% confidence limit for the 100-year flood return level. This limit is about 38% larger than the one derived from the asymptotic distribution, a dimension worth being considered for planing options. To investigate to what extend this increase in uncertainty depend on catchment characteristics, we plan to systematically study other gauges. Furthermore, a refined model selection strategy and the accounting for instationarities due to climate change is subject of further work.

11.9 Acknowledgements

This work was supported by the German Federal Ministry of Education and Research (grant number 03330271) and the European Union Baltic Sea INTERREG III B neighbourhood program (ASTRA, grant number 0085). We are grateful to the Bavarian Authority for the Environment for the provision of data. We further wish to thank Dr. V. Venema, Dr. Boris Orlowsky, Dr. Douglas Maraun, Dr. P. Braun, Dr. J. Neumann and H. Weber for several fruitful discussions.

11.10 Appendix

11.10.1 General Extreme Value Distribution

Consider the maximum

$$M_n = \max\{X_1, \ldots, X_n\} \tag{11.11}$$

of a sequence of n independent and identically distributed (iid) variables $X_1, \ldots, X_n$ with common distribution function F. This can be, for example, daily measured run-off at a gauge; M_n then represents the maximum over n daily measurements, e.g., the annual maximum for $n = 365$. The three types theorem states that

$$\Pr\{(M_n - b_n)/a_n \leq z\} \to G(z), \text{ as } n \to \infty, \tag{11.12}$$

with a_n and b_n being normalisation constants and $G(z)$ a non-degenerate distribution function known as the General Extreme Value distribution (Gev)

$$G(z) = \exp\left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}. \tag{11.13}$$

z is defined on $\{z | 1 + \xi(z - \mu)/\sigma > 0\}$. The model has a location parameter μ, a scale parameter σ and a form parameter ξ. The latter decides whether the distribution is of type II (Fréchet, $\xi > 0$) or of type III (Weibull, $\xi < 0$). The type I or Gumbel family

$$G(z) = \exp\left[- \exp\left\{ - \left(\frac{z - \mu}{\sigma} \right) \right\} \right], \{z | -\infty < z < \infty\} \tag{11.14}$$

is obtained in the limit $\xi \to 0$ [11.12].

It is convenient to transform Eq. (11.12) into

$$\Pr\{M_n \leq z\} \approx G((z - b_n)/a_n) = G^*(z). \tag{11.15}$$

The resulting distribution $G^*(z)$ is also a member of the GEV family and allows the normalisation constants and the location, scale and shape parameter to be estimated simultaneously.

We consider an autocorrelated stationary series $\{X_1, X_2, \ldots\}$ and define a condition of near-independence: For all $i_1 < \ldots < i_p < j_1 < \ldots < j_q$ with $j_1 - i_p > l$,

$$|\Pr\{X_{i_1} \leq u_n, \ldots, X_{i_p} \leq u_n, X_{j_1} \leq u_n, \ldots, X_{j_q} \leq u_n\} - \tag{11.16}$$
$$\Pr\{X_{i_1} \leq u_n, \ldots, X_{i_p} \leq u_n\} \Pr\{X_{j_1} \leq u_n, \ldots, X_{j_q} \leq u_n\}| \leq \alpha(n, l), \tag{11.17}$$

where $\alpha(n, l_n) \to 0$ for some sequence l_n, with $l_n/n \to 0$ as $n \to \infty$. It can be shown that the three types theorem holds also for correlated processes satisfying this condition of near-independence [11.12, 11.39]. This remarkable result implies that the limiting distribution of the maxima of uncorrelated and (a wide class) of correlated series belongs to the GEV family.

11.10.2 Maximum-Likelihood Parameter Estimation of the GEV

Let $\{M_{n,1}, M_{n,2}, \ldots, M_{n,m}\}$ be a series of independent block maxima observations, where n denotes the block size and m the number of blocks available for estimation. We denote $M_{n,i}$ as z_i. The likelihood function now reads

$$L(\mu, \sigma, \xi) = \prod_{i=1}^{m} g(z_i; \mu, \sigma, \xi), \tag{11.18}$$

where $g(z) = dG(z)/dz$ is the probability density function of the GEV. In the following, we consider the negative log-likelihood function $l(\mu, \sigma, \xi | z_i) = -\log L(\mu, \sigma, \xi | z_i)$.

Minimising the log-likelihood with respect to $\theta = (\mu, \sigma, \xi)$ leads to the ML estimate $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})$ for the GEV. Under suitable regularity conditions – among them independent observations z_i – and in the limit of large block sizes $(n \to \infty)$ $\hat{\theta}$ is multivariate normally distributed:

$$\hat{\theta} \sim \mathrm{MVN}_d(\theta_0, I_E(\theta_0)^{-1}) \tag{11.19}$$

with $I_E(\theta)$ being the expected information matrix (or Fisher information matrix) measuring the curvature of the log-likelihood. Denoting the elements of the inverse of I_E evaluated at $\hat{\theta}$ as $\beta_{j,k}$ we can approximate an $(1 - \alpha)100\%$ confidence interval for each component j of $\hat{\theta}$ by

$$\hat{\theta}_j \pm z_{\frac{\alpha}{2}} \sqrt{\beta_{j,j}}, \tag{11.20}$$

with $z_{\frac{\alpha}{2}}$ being the $(1-\alpha/2)$ quantile of the standard normal distribution [11.12].

The m-year return level can be easily calculated as specified in equation (11.2). An approximated confidence interval for $\hat{r}_m$ can be obtained under the hypothesis of a normally distributed estimator $\hat{r}_m$ and making use of the standard deviation $\sigma_{\hat{r}_m}$. The latter can be calculated from the information matrix using the delta method [11.12]. For the Gumbel distribution we obtain

$$\sigma^2_{\hat{r}_m} = \beta_{11} - (\beta_{22} + \beta_{21}) \log(-\log(1 - \frac{1}{m})) + \beta_{22}(\log(-\log(1 - \frac{1}{m})))^2. \tag{11.21}$$

11.10.3 Box-Cox Transform

The Box-Cox transformation can be used to transform a record $\{x_i\}$ such that its distribution is closer to a Gaussian. For records $\{x_i\}$ with $x_i > 0$ for all i, it is defined as [11.8]

$$y = \begin{cases} \frac{(x^\lambda - 1)}{\lambda}, \ \lambda \neq 0 \\ \log(x), \ \lambda = 0 \end{cases}$$

We choose the parameter λ such that the unconditional Gaussian likelihood is maximised. Hipel also advocate the use of this transformation for river runoff [11.28].

References

11.1 J.-M. Bardet, G. Lang, G. Oppenheim, A. Philippe, and M. S. Taqqu. Generators of long-range dependent processes: A survey. In P. Doukhan, G. Oppenheim, and M. S. Taqqu, editors, *Theory and Applications of Long-Range Dependence*, pages 579–623. Birkhäuser, Boston, 2003.

11.2 A. Bárdossy and F. Filiz. Identification of flood producing atmospheric circulation patterns. *Journal of Hydrology*, 313(1-2):48–57, 2005.

11.3 A. Bárdossy and S. Pakosch. Wahrscheinlichkeiten extremer Hochwasser unter sich ändernden Klimaverhältnissen. *Wasserwirtschaft*, 7-8:58–62, 2005.

11.4 J. Beran. *Statistics for Long-Memory Processes.* Monographs on Statistics and Applied Probability. Chapman & Hall, 1994.

11.5 J. Beran, R. J. Bhansali, and D. Ocker. On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrika*, 85(4):921–934, 1998.

11.6 L. Bisaglia. Model selection for long-memory models. *Quaderni di Statistica*, 4:33–49, 2002.

11.7 P. Bloomfield. Trends in global temperature. *Climatic Change*, pages 1–16, 1992.

11.8 G. E. P. Box and D. R. Cox. An analysis of transformations. *J. R. Statist. Soc.*, 62(2):211–252, 1964.

11.9 G. E. P. Box and G. M. Jenkins. *Time Series Analysis: forecasting and control.* Prentice Hall, New Jersey, 1976.

11.10 P. J. Brockwell and R. A. Davis. *Time series: Theory and Methods.* Springer Series in Statistics. Springer, Berlin, 1991.

11.11 R.B. Cleveland, W.S. Cleveland, J.E. McRae, and I. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess. *J. Official Statist.*, 6(1):3–73, 1990.

11.12 S. G. Coles. *An Introduction to Statistical Modelling of Extreme Values.* Springer, London, 2001.
11.13 S. G. Coles and M. J. Dixon. Likelihood-based inference for extreme value models. *Extremes*, 2(1):5–23, 1999.
11.14 S. G. Coles and J. A. Tawn. A Bayesian analysis of extreme rainfall data. *Appl. Stat.*, 45:463–478, 1996.
11.15 S. G. Coles and J. A. Tawn. Statistical methods for extreme values. Internet, 1999.
11.16 J. M. Cunderlik and T. B. Ouarda. Regional flood-duration-frequency modeling in the changing environment. *Journal of Hydrology*, 318(1-4):276–291, 2006.
11.17 A. C. Davison and D. B. Hinkley. *Bootstrap Methods and their Application.* Cambridge series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1997.
11.18 M. H. DeGroot. *Probability and Statistics.* Series in Statistics. Addison-Wesley, Reading, second edition, 1975.
11.19 B. Efron. Bootstrap methods: Another look at the jackknife. *Ann. Stat.*, 7(1):1–26, 1979.
11.20 P. Elek and L. Márkus. A long range dependent model with nonlinear innovations for simulating daily river flows. *Nat. Hazards Earth Syst. Sci.*, 4:227–283, 2004.
11.21 P. Embrechts, C. Klüppelberger, and T. Mikosch. *Modelling Extremal Events for Insurance and Fincance.* Springer, Berlin, 1997.
11.22 R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Phil. Soc.*, 24:180–190, 1928.
11.23 B. V. Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math. (2)*, 44:423–453, 1943.
11.24 C. W. Granger and R. Joyeux. An introduction to long-memory time series models and fractional differences. *J. Time Ser. Anal.*, 1(1):15–29, 1980.
11.25 E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *J. R. Statist. Soc.*, 41:190–195, 1979.
11.26 H. Held and T. Kleinen. Detection of climate system bifurcations by degenerate fingerprinting. *Geophys. Res. Lett.*, 31(L23207), 2004.
11.27 K. W. Hipel and A. I. McLeod. Preservation of the rescales adjusted range 2. Simulation studies using Box-Jenkins models. *Water Resour. Res.*, 14(3):509–516, 1978.
11.28 K.W. Hipel and A.I. McLeod. *Time Series Modelling of Water Resources and Environmental Systems.* Developement in Water Science. Elsevier, Amsterdam, 1994.
11.29 J. R. M. Hosking. Fractional differencing. *Biometrika*, 68:165–176, 1981.
11.30 J. R. M. Hosking. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. R. Statist. Soc.*, B:105–124, 1990.
11.31 J. R. M. Hosking, J. R. Wallis, and E. F. Wood. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261, 1985.
11.32 Climate change 2001: The scientific basis, contribution of working group I to the Third Assessment Report, 2001. Intergovernmental Panel on Climate Change (IPCC).
11.33 M. Kallache, H. W. Rust, and J. Kropp. Trend assessment: Applications for hydrology and climate. *Nonlin. Proc. Geoph.*, 2:201–210, 2005.
11.34 R. W. Katz, M. B. Parlange, and P. Naveau. Statistics of extremes in hydrology. *Advances in Water Resources*, 25:1287–1304, 2002.
11.35 D. Koutsoyiannis. Climate change, the Hurst phenomenon, and hydrological statistics. *Hydrol. Sci. J.*, 48(1), 2003.
11.36 J. P. Kropp, A. Block, F. Reusswig, K. Zickfeld, and H. J. Schellnhuber. Semiquantitative assessment of regional climate vulnerability: The North Rhine – Westphalia study. *Climatic Change*, 76(3-4):265–290, 2006.
11.37 Z. W. Kundzewicz. Is the frequency and intensity of flooding changing in Europe. In W. Kirch, B. Menne, and R. Bertollini, editors, *Extreme Weather Events and Public Health Responses.* Springer, 2005.
11.38 Z.W. Kundzewicz and H.-J. Schellnhuber. Floods in the IPCC TAR perspective. *Natural Hazards*, 31:111–128, 2004.
11.39 M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes.* Springer Series in Statistics. Springer, New York, 1983.

11.40 M. Lohre, P. Sibbertsen, and T. Könnig. Modeling water flow of the Rhine River using seasonal long memory. *Water Resour. Res.*, 39(5):1132, 2003.
11.41 D. Maraun, H. W. Rust, and J. Timmer. Tempting long-memory - on the interpretation of DFA results. *Nonlin. Proc. Geophys.*, 11:495–503, 2004.
11.42 B. Merz and A.H. Thieken. Separating natural and epistemic uncertainty in flood frequency analysis. *J. Hyd.*, 309:114–132, 2005.
11.43 A. Montanari, R. Rosso, and M. S. Taqqu. Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation and simulation. *Water Resour. Res.*, 33(5):1035, 1997.
11.44 A. Montanari, R. Rosso, and M. S. Taqqu. A seasonal fractional ARIMA model applied to the nile river monthly flows. *Water Resour. Res.*, 36(5):1249–1259, 2000.
11.45 P. Naveau, M. Nogaj, C. Ammann, P. Yiou, D. Cooley, and V. Jomelli. Statistical methods for the analysis of climate extremes. *C.R. Geoscience*, 377:1013–1022, 2005.
11.46 M. Ooms and J. Doornik. Inference and forecasting for fractional autoregressive integrated moving average models, with an application to us and uk inflation. Technical report, Econometric Institute, Erasmus University, Rotterdam, 1999. Econometric Institute Report 9947/A.
11.47 M. Ooms and P.H. Franses. A seasonal periodic long memory model for monthly river flows. *Environ. Model. & Software*, 16:559–569, 2001.
11.48 M. B. Priestley. *Spectral Analysis and Time Series.* Academic Press, London, 1992.
11.49 C. Prudhomme, D. Jakob, and C. Svensson. Uncertainty and climate change impact on the flood regime of small UK catchments. *J. Hyd.*, 277:1–23, 2003.
11.50 R Development Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2004. ISBN 3-900051-07-0.
11.51 T. Schreiber and A. Schmitz. Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.*, 77:635, 1996.
11.52 T. Schreiber and A. Schmitz. Surrogate time series. *Physica D*, 142:346–382, 2000.
11.53 R. L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, 1985.
11.54 R. L. Smith. Comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Appl.Stat.*, 36(3):358–369, 1987.
11.55 R. L. Smith. Environmental statistics. Internet, 2001. Version 5.0.
11.56 A. Stephenson and C. Ferro. *evd: Functions for extreme value distributions*, 2004. R package version 2.1-6.
11.57 P. A. Stott, D. A. Stone, and M. R. Allen. Human contribution to the European heatwave of 2003. *Nature*, 432, 2004.
11.58 J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer. Testing for nonlinearity in time series: The method of surrogate data. *Physica D*, 58(77), 1992.
11.59 J. Timmer. Modeling noisy time series: Physiological tremor. *Int. J. Bif. Chaos*, 8(7), 1998.
11.60 J. Timmer and M. König. On generating power law noise. *A&A*, 300(3):707–710, 1995.
11.61 V. Venema, S. Bachner, H. W. Rust, and C. Simmer. Statistical characteristics of surrogate data based on geophysical measurements. *Nonlin. Proc. Geophys.*, pages 449–466, 2006.
11.62 H. von Storch and F. W. Zwiers. *Statistical Analysis in Climate Research.* Cambridge University Press, Cambridge ; New York, 1999.
11.63 T. M. L. Wigley and P. D. Jones. Detecting CO2-induced climatic change. *Nature*, 292:205 – 208, 1981.

peer-reviewed by 2 reviewers