Advances in Water Resources 83 (2015) 1–9

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Flood frequency analysis: Confidence interval estimation by test inversion bootstrapping

Thomas Schendel[a], Rossukon Thongwichian[b,*]

[a] Oranienburger Straße 33, 16775 Gransee, Germany
[b] Department of Biotechnology, Faculty of Science and Technology, Thammasat University, Pathumthani 12121, Thailand

ARTICLE INFO

Article history:
Received 20 December 2014
Revised 3 May 2015
Accepted 10 May 2015
Available online 21 May 2015

Keywords:
Flood frequency analysis
Confidence interval
Test inversion bootstrapping

ABSTRACT

A common approach to estimate extreme flood events is the annual block maxima approach, where for each year the peak streamflow is determined and a distribution (usually the generalized extreme value distribution (GEV)) is fitted to this series of maxima. Eventually this distribution is used to estimate the return level for a defined return period. However, due to the finite sample size, the estimated return levels are associated with a range of uncertainity, usually expressed via confidence intervals. Previous publications have shown that existing bootstrapping methods for estimating the confidence intervals of the GEV yield too narrow estimates of these uncertainty ranges. Therefore, we present in this article a novel approach based on the less known test inversion bootstrapping, which we adapted especially for complex quantities like the return level. The reliability of this approach is studied and its performance is compared to other bootstrapping methods as well as the Profile Likelihood technique. It is shown that the new approach improves significantly the coverage of confidence intervals compared to other bootstrapping methods and for small sample sizes should even be favoured over the Profile Likelihood.

1. Introduction

The analysis of extreme events in hydrology is without doubt of crucial importance for engineering practice regarding water resources design and management. For example, hydraulic constructions such as dams are usually required to hold back extreme floods with a defined return period. One common approach to estimate such flood events with a certain return period (the return level), is the annual block maxima approach, where for every observed year the maximal streamflow is determined and then a distribution to this series of annual maxima is fitted. From the obtained distribution, the streamflow of extreme events can be estimated.

The generalized extreme value distribution (GEV) is generally used for fitting the annual series of maxima and has been employed in many applications in meteorology and hydrology. Examples include areas such as precipitation [17,20,25,31,43], floods [5,22,32,33,36,41,46], air temperature [26,29], wind speed extremes [49], and extreme sea levels [25,48,50]. The particular usefulness of the GEV is partly due to the Fisher–Tippett-theorem [16,19], stating that the GEV is the limiting distribution of maxima of a series of independent and identically distributed observations.

However, it is important to bear in mind that return levels (for a defined return period) estimated from the observed time series will not be identical with the true return level of the (unknown) underlying true distribution. After all, this is the fate of every finite sample. Therefore the return level estimated from the observed time series may over- or underestimate the true return level. One common way to represent this uncertainty is the estimation of confidence intervals (CIs).

In order to estimate these CIs, the bootstrap approach is widely employed (for example in [2,10,27,28,38,39,41,43,47], for the theoretical background of the bootstrap approach, see [8,9,11–14,35,42], a good overview over these methods is given for example in [4,7]). The characteristic of the bootstrapping method is that from the given sample, new samples are generated by using either the non-parametric (via resampling with replacement) or parametric approach (via fitting a parametric distribution to the observed sample and then randomly drawing new samples from this distribution). In case of estimating the CI for large quantiles of extreme events, an extensive study by Kyselý [30] has convincingly shown that for the heavy-tailed GEV, the parametric bootstrapping methods are in general superior to the non-parametric ones yielding more realistic ranges for double sided CI. However, this study also revealed that even the best parametric bootstrap method resulted in CIs of return levels that were systematically too narrow. For example, in case of the 95% CI regarding the 100-year return level estimated from a sample

* Corresponding author. Tel.: +66 917809579.
E-mail address: rossukont@tu.ac.th (R. Thongwichian).

http://dx.doi.org/10.1016/j.advwatres.2015.05.004

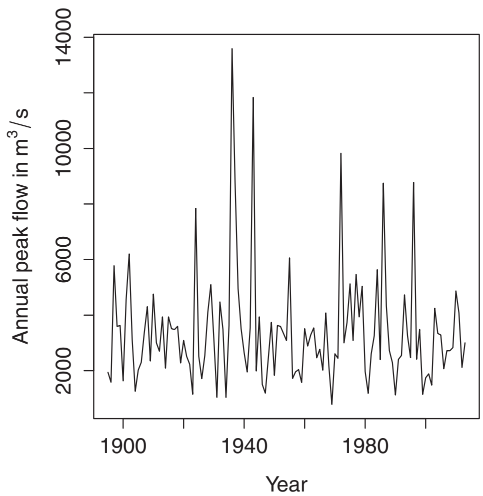

Fig. 1. Annual peak flow at Point of Rocks (Potomac River) from 1895–2013.

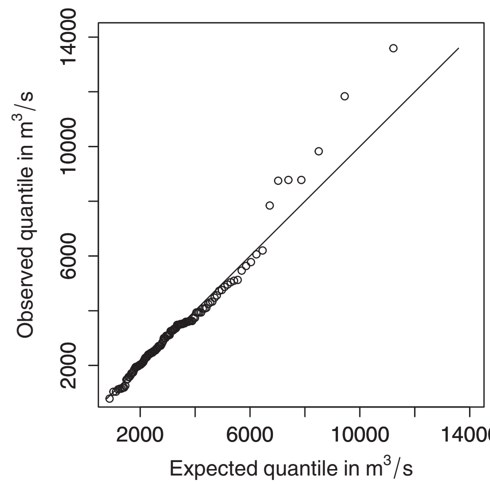

Fig. 2. Q–Q plot for fit of the generalized extreme value distribution to the annual peak flow at Point of Rocks (Potomac River). The line of quality indicates perfect fit.

size of $N = 100$, all parametric bootstrap approaches studied in [30] led to CIs where 7.7% or more of the samples drawn from a parent distribution (the GEV in this case) did not include the true 100-year return level. For even smaller sample sizes, the coverage error became even worse. These findings are in terms of a realistic risk assessment not satisfying.

Therefore, an alternative method to the bootstrap for estimating the CIs was proposed: the Profile Likelihood [6,15,51], which is based on the likelihood ratio test. While applications to hydrology are not widespread, a recent publication [37] suggests that the Profile Likelihood is superior to the bootstrap.

However, one parametric bootstrapping approach not considered by Kyselý [30] is the test inversion bootstrap (TIB), which exploits the duality between hypothesis testing and confidence intervals (for an introduction to the TIB, see [3,4]). This method is less known and applications (to our best knowledge absent in the field of hydrology) focused only on CI of single parameters of the distribution itself [3,4], but not on quantities like the return level, which is an explicit function of all three parameters of the GEV.

Therefore, in this article we aim to present a novel method particularly designed for estimating the confidence intervals of return levels of the GEV. This approach is based on test inversion bootstrapping combined with the use of the likelihood function. To our best knowledge, the application of the test inversion regarding the bootstrap to a quantity (in this case the return level) which is an explicit function of the three parameters of the distribution is new. The accuracy of this method is studied on both, single sided and double sided CIs, and compared to previously published results of other bootstrapping methods as well as the performance of the Profile Likelihood.

2. Methods

In order to demonstrate the novel approach of estimating confidence intervals (CI), the annual peak flow at Potomac River at Point of Rocks, MD, USA (source: US Geological Survey http://water.usgs.gov/nwis/peak) from 1895–2013 (water year from October to September) was studied. The respective data are presented in Fig. 1.

In the following, we use the generalized extreme value distribution (GEV) to fit to this annual peak flow time series. We use the following notation of the cumulative GEV distribution (with the parameters a, b, c and the peak flow x):

$$F_{a,b,c}(x) = e^{-\left(1+a\frac{x-c}{b}\right)^{-\frac{1}{a}}};\ a \neq 0;\ 1 + a\frac{x-c}{b} > 0$$
$$F_{0,b,c}(x) = e^{-e^{-\frac{x-c}{b}}};\ a = 0. \quad (1)$$

To avoid confusion, we would like to mention that an alternative parameterisation of the GEV with $-a$ instead of a is often used in hydrology. Anyway, for $a \geq 0$ the GEV has no upper boundary, and for $a > 0$ the GEV is heavily tailed (since the moments greater than $\frac{1}{a}$ are infinite) in contrast to $a \leq 0$, where the distribution has a finite upper bound. The special case $a = 0$ is called the Gumbel distribution, which is unbound but all of its moments are defined.

For the given example, we fitted the GEV distribution to the annual peak flow time series using the maximum likelihood estimation (MLE) [6,45]. While other methods such as probability weighted moments (PWM) [21,23]) and combinations of MLE and PWM [36] perform better for small samples, this effect will not play much of a role for the example presented here (which contains 119 values). Moreover, the MLE approach does not rely on the assumption that a finite mean exists.

For the given example of the annual peak flow at Point of Rocks (Potomac River), we obtained the shape parameter to $\hat{a} = 0.176$, which provides some evidence that the distribution is heavy tailed (P-value $P \approx 0.002$ for likelihood ratio test of $a = 0$). We would like to mention that this gauge was fitted to the GEV distribution by Smith and Katz and coworkers [44] and [25], where the record ended in 1986 and 2000, respectively. The shape parameter was estimated to 0.42 and 0.191. For the fit performed here, a Q–Q plot (presented in Fig. 2) shows a fairly good fit for lower and middle annual peak flow, however for very large annual peak flows the fit underestimates the observed values. If the 1000-year return level for the Potomac River at Point of Rocks is estimated using the GEV distribution, we arrive at a value of approximately 18,200 $\frac{m^3}{s}$. However, this estimation of the 1000-year return level is based on a finite sample. Therefore we need to address this uncertainty by estimating the CI.

The approach we present here to estimate the CI is based on test inversion bootstrapping (TIB). An introduction to TIB can be found in [3,4]. We would like to introduce here a less technical and more intuitive approach to motivate the test inversion approach. As example we consider the two-sided 95% CI. In order to estimate this interval, we need to answer the following question: if the 1000-year return level estimated from the observed sample is within a probability range of non-exceedance of 2.5–97.5% of the underlying (but unknown) “true distribution” , in which interval would the “true” 1000-year return level be located? This view highlights the fact that the observed time series is just a sample drawn from the “true” distribution and that therefore the 1000-year return level estimated from this sample has a certain (yet unknown) probability of non-exceedance regarding the 1000-year return level estimated from the