QUANTITATIVE METHODS IN ECONOMICS
Vol. XII, No. 1, 2011, pp. 37– 52

THE QUANTILE ESTIMATION OF THE MAXIMA OF SEA LEVELS

Marcin Dudziński, Konrad Furmańczyk
Department of Applied Mathematics
Warsaw University of Life Sciences –SGGW
E-mails: marcin_dudzinski@sggw.pl; konrad_furmanczyk@sggw.pl

Abstract. The hydrological modeling has become an intensively studied subject in recent years. One of the most significant problems concerning this issue is to provide the mathematical and statistical tools, which allow to forecast extreme hydrological events, such as severe sea or river floodings. The extreme events on water have huge social and economic impact on the affected areas. Due to these reasons, each country has to protect itself against the flood danger, and consequently, the designing of reliable flood defences is of great importance to the safety of the region. For example, the sea dikes along the Dutch coastline are designed to withstand floods, which may occur once every 10 000 years. It means that the height of the dike is determined in such a way that the probability of the event that there is a flood in a given year equals 10^{-4}. The computation of such the height level requires the estimation of the corresponding quantiles of the distributions of certain maxima of sea levels. In our paper, we present the procedures, which lead to the estimation of such the quantiles. We are mainly concerned with the interval estimation; in this context, we present the frequentistic and Bayesian approaches in constructing the desired confidence intervals.

Key words: quantile estimation, frequentistic confidence interval, Bayesian confidence interval, peaks over threshold (POT)

INTRODUCTION AND PRELIMINARIES

The purpose of our work is to give the estimation procedures for the $(1-\alpha)$-quantile of the distribution of the maxima of the North Sea levels at Hook of Holland, the Netherlands. In our investigations, we apply the informations on sea levels from the period 1945-1995, available on the website http://live.waterbase.nl/.

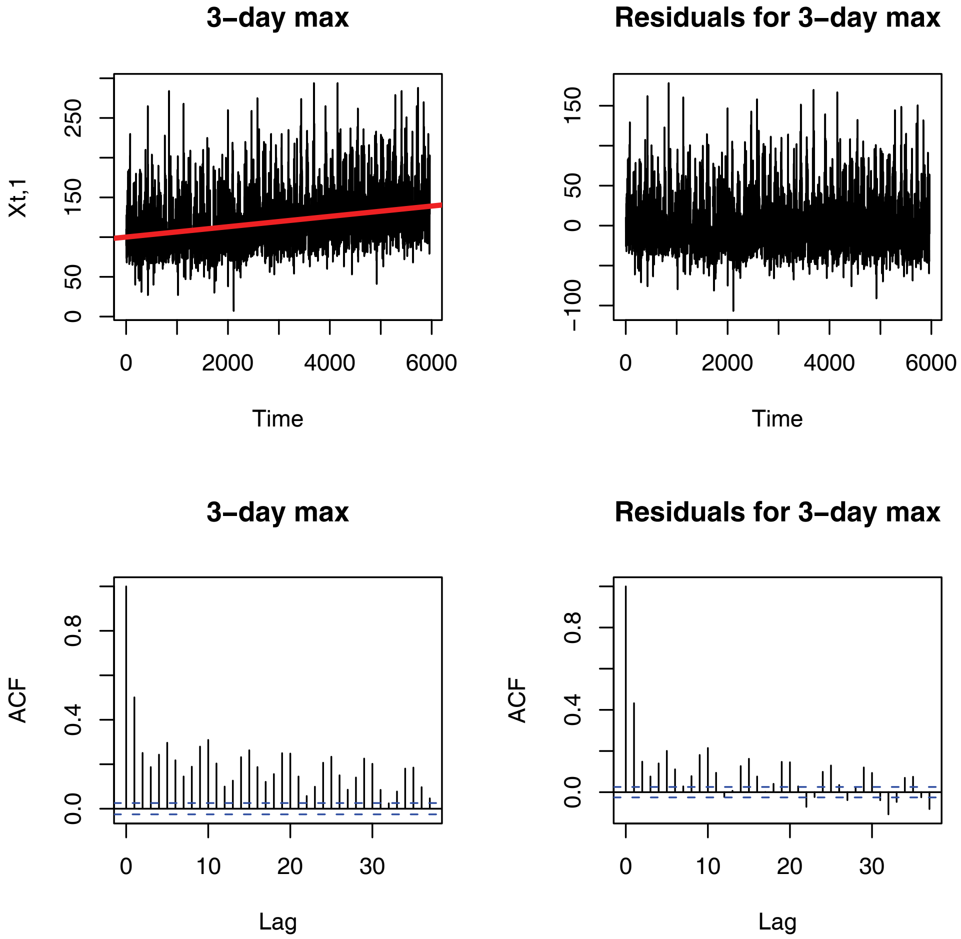
The problem of the estimation of such a quantile was earlier considered in the papers of [Van Gelder 1996] and [Van Gelder et al. 1995]. The mentioned articles attracted our attention towards the issue of the quantile estimation of extreme hydrological events and encouraged us to undertake some research in this field, although our approach differs from the methods proposed by Van Gelder et al. There are some other valuable publications devoted to the modeling of extreme water events. We cite in this context the papers of [Katz et al. 2002], [Knox and Kundzewicz 1997] and [Shiau 2003] among others.

The set of our empirical data consists of the maximum sea levels along the Dutch coast in the area of Hook of Holland in the periods of 3 days. We will denote such the 3-day maxima by $X_1, X_2, ..., X_T$; $T = 5963$ (i.e., X_t will stand for the maximum sea level in the t-th 3-day period; we had 5963 such the 3-day periods in the years 1945-1995). The reason for such a choice of data is that the storms on the North Sea never last longer than 72 hours (actually, an average storm there lasts from 9 to 12 hours).

We observed some autocorrelations and the increasing linear trend in the sequence of the 3-day maxima of sea levels at Hook of Holland in the period 1945-1995 (see Figure 1). By using the least squares method, we obtained the model of the form $\hat{X}_t = 100.1 + 0.006507 \cdot t$ (both the parameters turned out to be statistically significant; the corresponding p-value $= 2e\text{-}16$). It means that, on average, the 3-day maxima were increasing by 0.006507 cm from one 3-day subperiod to the next 3-day subperiod, and consequently, they increased by around 39 cm within 50 years.

Therefore, we will consider the following model for the 3-day maxima of sea levels at Hook of Holland: $X_t = a + bt + \varepsilon_t$, where a, b are some constants and the random errors ε_t (all with a common marginal d.f. F_ε) may be correlated (see Figure 1), but in turn, the error terms exceeding a certain, sufficiently large threshold u are independent (this independence is a consequence of the fact that an average storm on the North Sea does not last longer than 3 days).

Figure 1. The 3-day maxima of sea levels, the residuals $x_t - \hat{x}_t$ and the corresponding autocorrelation functions



Source: own preparation

We will assume that for sufficiently large threshold u:

(A1) $\overline{F}_\varepsilon(x) = P(\varepsilon > x) \approx \exp\left(-\frac{x}{\beta}\right)$, for some $\beta > 0$, if $x > u$,

(A2) the r.v.'s $Y_i = (\varepsilon_i - u \mid \varepsilon_i > u)$ (where $i = 1,2,\ldots,K$; $K = \sum_{t=1}^{T} I(\varepsilon_t > u)$) are independent.

The assumption in (A1) means that the distribution of the model errors is (approximately) of the thin-tailed, exponential type. The validity of this condition for the errors of our model will be verified in the further stages of our study. The assumption in (A2) has its clear physical explanation since, as we have already mentioned, the storms on the North Sea never last longer than 3 days.

Our main aim is to construct the confidence intervals for $F_{X_T}^{-1}(1-\alpha)$ - the $(1-\alpha)$-quantile of the distribution of X_T. The $(1-\gamma)\cdot 100\%$ confidence interval $(L(T), R(T))$ for $F_{X_T}^{-1}(1-\alpha)$ is defined by

$$P\left(L(T) \leq F_{X_T}^{-1}(1-\alpha) \leq R(T)\right) = 1-\gamma.$$

Let us denote by $F_\varepsilon^{-1}(1-\alpha)$ the $(1-\alpha)$-quantile of the distribution of the model errors ε_t. Furthermore, let $l(T)$, $r(T)$ satisfy

$$P\left(l(T) \leq F_\varepsilon^{-1}(1-\alpha) \leq r(T)\right) = 1-\gamma.$$

Obviously, we have: $L(T) = a + bT + l(T)$, $R(T) = a + bT + r(T)$, where a, b are the parameters of the linear model for X_t. Thus, in order to construct the confidence interval for $F_{X_T}^{-1}(1-\alpha)$, it is sufficient to establish the confidence interval for $F_\varepsilon^{-1}(1-\alpha)$.

The remainder of the paper is organized as follows. In Section "The approximated quantile of the distribution of the model errors", we will derive the approximated formula for $F_\varepsilon^{-1}(1-\alpha)$. In Section "Interval estimation of the quantile ..." we will construct the approximated frequentistic and Bayesian confidence intervals for $F_\varepsilon^{-1}(1-\alpha)$ and $F_{X_T}^{-1}(1-\alpha)$. We will apply these formulas later in Section "The confidence intervals for the quantiles of the maxima of sea levels ..." to compute the realizations of the confidence intervals for the quantiles of the distribution of the 3-day maxima of the North Sea levels at Hook of Holland. In Section "The correctness and accuracy of the obtained estimation procedures ...", we will assess the quality of our estimation procedures. Finally, in Section "Final conclusions", we will conclude and summarize our study.

THE APPROXIMATED QUANTILE OF THE DISTRIBUTION OF THE MODEL ERRORS

Suppose that a threshold u is such as in the assumptions (A1), (A2) from the previous section. Let, for $y > 0$, $F_u(y) := P(\varepsilon - u \leq y \mid \varepsilon > u)$. We have

$$F_u(x-u) = P(\varepsilon - u \leq x-u \mid \varepsilon > u) = P(\varepsilon \leq x \mid \varepsilon > u) = \frac{F_\varepsilon(x) - F_\varepsilon(u)}{1-F_\varepsilon(u)}, \tag{1}$$

if $x-u>0$. By (1) and the assumption (A1), we obtain that

$$F_u(x-u)=1-\frac{\overline{F}_\varepsilon(x)}{\overline{F}_\varepsilon(u)}\approx 1-\frac{\exp(-x/\beta)}{\exp(-u/\beta)},\ \text{for sufficiently large } u \text{ and } x>u,$$

where $\beta=\beta(u)$ is a certain parameter, which depends on u. Therefore,

$$P(\varepsilon-u\le x-u\mid\varepsilon>u)=F_u(x-u)\approx 1-\exp\left(-\frac{x-u}{\beta}\right),\ \text{if } x>u. \tag{2}$$

The relation in (2) means that the conditional distribution of $\varepsilon-u$ given the event that ε exceeds a large threshold u is approximately the exponential distribution $Exp(1/\beta)$. Obviously, due to (1), we have $F_\varepsilon(x)=(1-F_\varepsilon(u))F_u(x-u)+F_\varepsilon(u)$, for $x>u$. This and the approximation in (2) imply

$$F_\varepsilon(x)\approx(1-F_\varepsilon(u))\left(1-\exp\left(-\frac{x-u}{\beta}\right)\right)+F_\varepsilon(u),\ \text{if } x>u.$$

Let us denote by $q_{1-\alpha}^{(\varepsilon)}>u$ the $(1-\alpha)$-quantile of F_ε. Then, $F_\varepsilon(q_{1-\alpha}^{(\varepsilon)})=1-\alpha$ and, by substituting $q_{1-\alpha}^{(\varepsilon)}$ for x into the relation above, we get

$$1-\alpha\approx(1-F_\varepsilon(u))\left(1-\exp\left(-\frac{q_{1-\alpha}^{(\varepsilon)}-u}{\beta}\right)\right)+F_\varepsilon(u).$$

Hence, the approximated $(1-\alpha)$-quantile of F_ε is given by

$$F_\varepsilon^{-1}(1-\alpha)=q_{1-\alpha}^{(\varepsilon)}\approx u+\beta\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right). \tag{3}$$

The purpose of our next study is to construct the confidence intervals for the quantile $F_\varepsilon^{-1}(1-\alpha)$, and consequently, for the quantile $F_{X_T}^{-1}(1-\alpha)$ as well. We will conduct our investigations under the condition that $K=k$ (as, due to the estimated model for X_T, the number of residuals exceeding u is known). We will consider two approaches of interval estimation: the first is based on the frequentistic analysis, while the ideas of the second approach derive from the Bayesian analysis.

INTERVAL ESTIMATION OF THE QUANTILE - THE PROPOSED ESTIMATION PROCEDURES

The frequentistic approach

Let $Y_i:=(\varepsilon_i-u\mid\varepsilon_i>u)$, where $i=1,2,\ldots,k$ and ε_i, u, k are such as in the previous sections. By the conditions in (A1), (A2), we have that the $Y_i^{'}s$ are independent and have approximately the $Exp(1/\beta(u))$ distribution (see (2)). Hence, the maximum likelihood estimate of the parameter $\beta(u)$ is of the form

$$\hat{\beta} = \sum_{i=1}^{k} Y_i / k. \tag{4}$$

Due to (3), (4), we have the following estimate for the quantile $q_{1-\alpha}^{(\varepsilon)} = F_\varepsilon^{-1}(1-\alpha)$

$$\hat{q}_{1-\alpha}^{(\varepsilon)} = u + \hat{\beta} \ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right). \tag{5}$$

Let us construct the frequentistic confidence interval for $q_{1-\alpha}^{(\varepsilon)} - \hat{q}_{1-\alpha}^{(\varepsilon)}$. In order to obtain it, we need to find v_γ, w_γ, satisfying the following relation

$$P\left(v_\gamma \le q_{1-\alpha}^{(\varepsilon)} - \hat{q}_{1-\alpha}^{(\varepsilon)} \le w_\gamma\right) = 1-\gamma. \tag{6}$$

By substituting the expressions on the r.h.s. of (3) and (5) for $q_{1-\alpha}^{(\varepsilon)}$ and $\hat{q}_{1-\alpha}^{(\varepsilon)}$ into (6), we obtain $P\left(v_\gamma \le (\beta - \hat{\beta}) \ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right) \le w_\gamma\right) = 1-\gamma$, which (since $\frac{1-F_\varepsilon(u)}{\alpha} > 1$, as $q_{1-\alpha}^{(\varepsilon)} > u$) yields

$$P\left(\frac{-w_\gamma}{\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)} \le \hat{\beta} - \beta \le \frac{-v_\gamma}{\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)}\right) = 1-\gamma.$$

By using (4), we have

$$P\left(\left(\frac{-w_\gamma}{\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)} + \beta\right) k \le \sum_{i=1}^{k} Y_i \le \left(\frac{-v_\gamma}{\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)} + \beta\right) k\right) = 1-\gamma.$$

Since $\sum_{i=1}^{k} Y_i$ has approximately the gamma $\Gamma(k, 1/\beta)$ distribution (as Y_i are independent and, approximately, $Y_i \sim Exp(1/\beta)$) and a confidence level equals $1-\gamma$, we may write that:

$$\left(\frac{-w_\gamma}{\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)} + \beta\right) k = \Gamma^{-1}\left(\frac{\gamma}{2}, k, \frac{1}{\beta}\right),$$

$$\left(\frac{-v_\gamma}{\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)}+\beta\right)k=\Gamma^{-1}\left(1-\frac{\gamma}{2},k,\frac{1}{\beta}\right),$$

where $\Gamma^{-1}(\cdot,\cdot,\cdot)$ denote the corresponding quantiles of the gamma distribution. Therefore:

$$v_\gamma=-\left(\frac{\Gamma^{-1}\left(1-\frac{\gamma}{2},k,\frac{1}{\beta}\right)}{k}-\beta\right)\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right), \tag{7}$$

$$w_\gamma=-\left(\frac{\Gamma^{-1}\left(\frac{\gamma}{2},k,\frac{1}{\beta}\right)}{k}-\beta\right)\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right). \tag{8}$$

Obviously, the frequentistic confidence interval for $q_{1-\alpha}^{(\varepsilon)}$ is given by

$$P\left(v_\gamma+\hat{q}_{1-\alpha}^{(\varepsilon)}\le q_{1-\alpha}^{(\varepsilon)}\le w_\gamma+\hat{q}_{1-\alpha}^{(\varepsilon)}\right)=1-\gamma. \tag{9}$$

Thus, due to (9) and (5), (7), (8), we obtain the following left (lower) and right (upper) ends (limits) of the frequentistic confidence interval for $q_{1-\alpha}^{(\varepsilon)}=F_\varepsilon^{-1}(1-\alpha)$ - the $(1-\alpha)$-quantile of the distribution of the model errors:

$l_{fr}(T)$ - the left end of the frequentistic confidence interval for ε_T

$$=u+\hat{\beta}\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)-\left(\frac{\Gamma^{-1}\left(1-\frac{\gamma}{2},k,\frac{1}{\beta}\right)}{k}-\beta\right)\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right), \tag{10}$$

$r_{fr}(T)$ – the right end of the frequentistic confidence interval for ε_T

$$=u+\hat{\beta}\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)-\left(\frac{\Gamma^{-1}\left(\frac{\gamma}{2},k,\frac{1}{\beta}\right)}{k}-\beta\right)\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right). \tag{11}$$

As $F_\varepsilon(u)$ is unknown, we may use its empirical distribution

$$\hat{F}_\varepsilon(u)=1-k/T, \tag{12}$$

where (for recollection) k stands for the number of residuals exceeding u.

Thus, by substituting the estimates for β and $F_\varepsilon(u)$ (see (4) and (12), respectively) into the formulas (10), (11), we have:

$$\hat{l}_{fr}(T)=u+\left(2\hat{\beta}-\frac{\Gamma^{-1}\left(1-\frac{\gamma}{2},k,\frac{1}{\hat{\beta}}\right)}{k}\right)\ln\left(\frac{k}{T\alpha}\right), \tag{13}$$

$$\hat{r}_{fr}(T)=u+\left(2\hat{\beta}-\frac{\Gamma^{-1}\left(\frac{\gamma}{2},k,\frac{1}{\hat{\beta}}\right)}{k}\right)\ln\left(\frac{k}{T\alpha}\right). \tag{14}$$

The relations above determine the ends (limits) of the approximated frequentistic confidence interval for $q_{1-\alpha}^{(\varepsilon)}=F_\varepsilon^{-1}(1-\alpha)$. Consequently, the approximated frequentistic confidence interval for $q_{1-\alpha}=F_{X_T}^{-1}(1-\alpha)$ is given by:

$$\hat{L}_{fr}(T)=\hat{a}+\hat{b}T+\hat{l}_{fr}(T), \qquad \hat{R}_{fr}(T)=\hat{a}+\hat{b}T+\hat{r}_{fr}(T), \tag{15}$$

where $\hat{a}$, $\hat{b}$ stand for the parameter estimates of the model $X_t=a+bt+\varepsilon_t$.

The Bayesian approach

Assume that the prior distribution of the parameter β is the inverse gamma distribution $IG(c,d)$, where c and d are some constants, which may be chosen on subjective grounds by using our knowledge before any data set is available. It follows from (A1) and (A2) that the r.v.'s $Y_i=(\varepsilon_i-u\mid\varepsilon_i>u)$; $i=1,2,...,k$, where $k=\sum_{t=1}^{T}I(\varepsilon_t>u)$, are independent and (approximately) distributed according to the exponential distribution $Exp(1/\beta)$ (see the approximation in (2)). Let (y_i) denote the realizations of (Y_i). Then, the posterior distribution of $\beta\mid y_1,...,y_k$ is the inverse gamma distribution $IG(c+k,\left(d+\sum_{i=1}^{k}y_i\right)^{-1})$. Our goal now is to construct the Bayesian confidence interval for the quantile $q_{1-\alpha}^{(\varepsilon)}=F_\varepsilon^{-1}(1-\alpha)$. It means that, for a fixed confidence level $1-\gamma$, we need to find $l_{Ba}=l_{Ba}(T,c,d,k)$, $r_{Ba}=r_{Ba}(T,c,d,k)$, such that the following condition holds: $P(l_{Ba}\le q_{1-\alpha}^{(\varepsilon)}\le r_{Ba}\mid y_1,...,y_k)=1-\gamma$. This and the approximation in (3) imply $P\left(l_{Ba} \le u + \beta \ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right) \le r_{Ba} \mid y_1,\ldots,y_k\right) = 1-\gamma$. Since $\frac{1-F_\varepsilon(u)}{\alpha} > 1$ (as $q_{1-\alpha}^{(\varepsilon)} > u$), the last relation is equivalent to the following one: $P\left(\frac{l_{Ba}-u}{\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)} \le \beta \le \frac{r_{Ba}-u}{\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)} \mid y_1,\ldots,y_k\right) = 1-\gamma$. Due to the fact that $\beta \mid y_1,\ldots,y_k$ has the $IG\left(c+k,\left(d+\sum_{i=1}^k y_i\right)^{-1}\right)$ distribution, we get:

$$\frac{l_{Ba}-u}{\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)} = IG^{-1}\left(\frac{\gamma}{2}, c+k, \left(d+\sum_{i=1}^k y_i\right)^{-1}\right),$$

$$\frac{r_{Ba}-u}{\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)} = IG^{-1}\left(1-\frac{\gamma}{2}, c+k, \left(d+\sum_{i=1}^k y_i\right)^{-1}\right),$$

where $IG^{-1}(\cdot,\cdot,\cdot)$ denote the corresponding quantiles of the inverse gamma distribution. Hence, the Bayesian confidence interval for the quantile $q_{1-\alpha}$ is determined by:

$$l_{Ba} = u + IG^{-1}\left(\frac{\gamma}{2}, c+k, \left(d+\sum_{i=1}^k y_i\right)^{-1}\right)\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right),$$

$$r_{Ba} = u + IG^{-1}\left(1-\frac{\gamma}{2}, c+k, \left(d+\sum_{i=1}^k y_i\right)^{-1}\right)\ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right).$$

As $IG^{-1}(1-\alpha, c, d) = 1/\Gamma^{-1}(\alpha, c, 1/d)$, we may rewrite l_{Ba} and r_{Ba} by means of the quantiles of the gamma distribution as follows:

$$l_{Ba} = u + \ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)/\Gamma^{-1}\left(1-\frac{\gamma}{2}, c+k, d+\sum_{i=1}^k y_i\right),$$

$$r_{Ba} = u + \ln\left(\frac{1-F_\varepsilon(u)}{\alpha}\right)/\Gamma^{-1}\left(\frac{\gamma}{2}, c+k, d+\sum_{i=1}^k y_i\right).$$

Let us put: $c = 1/\beta_0 + 1$, $d = 1$, where $\beta_0 = \sum_{i=1}^k y_i / k$ is the initial estimate of β. Then, since $\beta \sim IG(c,d)$, we have $E\beta = d/(c-1) = \beta_0$.

Furthermore, as $F_\varepsilon(u)$ is unknown, we may replace it by its natural estimate in (12). Therefore:

$$\hat{l}_{Ba} = \hat{l}_{Ba}(T) = u + \ln\left(\frac{k}{T\alpha}\right) / \Gamma^{-1}\left(1 - \frac{\gamma}{2}, \frac{k}{\sum_{i=1}^{k} y_i} + 1 + k, 1 + \sum_{i=1}^{k} y_i\right), \tag{16}$$

$$\hat{r}_{Ba} = \hat{r}_{Ba}(T) = u + \ln\left(\frac{k}{T\alpha}\right) / \Gamma^{-1}\left(\frac{\gamma}{2}, \frac{k}{\sum_{i=1}^{k} y_i} + 1 + k, 1 + \sum_{i=1}^{k} y_i\right). \tag{17}$$

The relations above determine the ends (limits) of the approximated Bayesian confidence interval for $q_{1-\alpha}^{(\varepsilon)} = F_\varepsilon^{-1}(1-\alpha)$. Consequently, the approximated Bayesian confidence interval for $q_{1-\alpha} = F_{X_T}^{-1}(1-\alpha)$ is given by:

$$\hat{L}_{Ba}(T) = \hat{a} + \hat{b}T + \hat{l}_{Ba}(T), \qquad \hat{R}_{Ba}(T) = \hat{a} + \hat{b}T + \hat{r}_{Ba}(T), \tag{18}$$

where $\hat{a}$, $\hat{b}$ denote the parameter estimates of the model $X_t = a + bt + \varepsilon_t$.

In the following section, we will present the computed realizations of the confidence intervals for the quantiles of the distribution of the 3-day maxima of sea levels, obtained according to the derived formulas for confidence intervals.

THE CONFIDENCE INTERVALS FOR THE QUANTILES OF THE MAXIMA OF SEA LEVELS - SOME COMPUTATION RESULTS

As we have already mentioned, our data set consists of the 3-day maxima of the North Sea levels along the Dutch coast at Hook of Holland, collected in the period 1945-1995. It is graphically presented on Figure 1. The sample size equals $T = 5963$ observations and some autocorrelations, as well as the increasing linear trend can be seen in the sample of the 3-day maxima X_t. By using the least squares method, we obtained the estimation of the model $X_t = a + bt + \varepsilon_t$. It had the form $\hat{X}_t = 100.1 + 0.006507 \cdot t$ (both the parameters were statistically significant; p-value $= 2e\text{-}16$).

Let: (x_t) denote our empirical data, $\hat{x}_t = 100.1 + 0.006507 \cdot t$, $\hat{\varepsilon}_t = x_t - \hat{x}_t$; $t = 1,2,\ldots,5963$. We carried out the plot of the sample autocorrelation functions for the residuals $\hat{\varepsilon}_t$ (see Figure 1.), which denied the independence of the model errors. On the other hand, there was no trend in the sequence of residuals

$(\hat{\varepsilon}_t)$ (see also Figure 1.), which confirmed the stationarity of the error terms of the model. We also checked that the tail of the d.f. of the sequence $(\hat{\varepsilon}_t)$ declined exponentially (at least approximately). Thus, as we might assume that (approximately) the tail of the d.f. of the error term ε declined exponentially (i.e., $\overline{F}_\varepsilon(x) = P(\varepsilon > x) \approx \exp(-x/\beta)$ for some $\beta > 0$, if $x \to \infty$ - see the condition in (A1)), we obtained that, for each i and sufficiently large threshold u, $Y_i = (\varepsilon_i - u \mid \varepsilon_i > u)$ had approximately the exponential distribution $Exp(1/\beta(u))$ (see (2)). Next, we determined several sufficiently large thresholds u. For each threshold, we chose those of the values of residuals $\hat{\varepsilon}_t$, which exceeded u and constructed the sample $y_i = (\hat{\varepsilon}_i - u | \hat{\varepsilon}_i > u)$; $i = 1,2,...,k$, where $k = \sum_{t=1}^{5963} I(\hat{\varepsilon}_t > u)$. Next, we carried out the plot of the sample autocorrelation functions for y_i and concluded that it might be assumed that (Y_i) was a sequence of independent r.v.'s (see the condition in (A2)). Moreover, by using the Kolmogorov test, we also verified the accordance of $Y_i = (\varepsilon_i - u | \varepsilon_i > u)$ with the exponential distribution. Thus, we were in a position to apply the derived formulas for the approximated confidence intervals.

Below, we present the results of our computations. We considered the following thresholds u: 30, 40, 45, 50 and the confidence level $1-\gamma = 0.95$. It turned out that the shortest confidence interval for $F_\varepsilon^{-1}(1-\alpha) = q_{1-\alpha}^{(\varepsilon)}$ was reached for $u = 30$. The realizations of the frequentistic (fr) and Bayesian (Ba) confidence intervals for $F_\varepsilon^{-1}(1-\alpha) = q_{1-\alpha}^{(\varepsilon)}$, calculated according to (13), (14) and (16), (17), in the case where $u = 30$, are given in the table below:

Table 1. The frequentistic and Bayesian confidence intervals for $q_{1-\alpha}^{(\varepsilon)}$; $u = 30$

	$1-\alpha = 0.95$	$1-\alpha = 0.99$	$1-\alpha = 0.999$	$1-\alpha = 0.9999$
fr	$(54.8;58.4)$	$(93.9;103.3)$	$(149.9;167.4)$	$(205.8;231.6)$
Ba	$(54.9;58.5)$	$(94.1;103.5)$	$(150.3;167.9)$	$(206.5;232.2)$

Source: own calculations

Thus, due to (15), (18), by taking into account the computed correction of trend, which equaled $100.1 + 0.006507 \cdot 5963 = 138.9012$ cm, we obtained the following realizations of the confidence intervals for $F_{X_T}^{-1}(1-\alpha) = q_{1-\alpha}$, if $u = 30$:

Table 2. The frequentistic and Bayesian confidence intervals for $q_{1-\alpha}$; $u = 30$

	$1-\alpha = 0.95$	$1-\alpha = 0.99$	$1-\alpha = 0.999$	$1-\alpha = 0.9999$
fr	$(193.7;197.3)$	$(232.8;242.2)$	$(288.8;306.3)$	$(344.7;370.5)$
Ba	$(193.8;197.4)$	$(233.0;242.4)$	$(289.2;306.8)$	$(345.4;371.1)$

Source: own calculations

THE CORRECTNESS AND ACCURACY OF THE OBTAINED ESTIMATION PROCEDURES – A SIMULATION STUDY

Since the approximation in (2) is valid when F_ε is of the thin-tailed type d.f., we conducted our simulations assuming that the data come from one of the following distributions (the values in parentheses denote the threshold values, which had been chosen in such a manner that, for the given quantile rank $1-\alpha$, u satisfied $1-F_\varepsilon(u) > \alpha$, i.e., $\ln\frac{1-F_\varepsilon(u)}{\alpha} > 0$): i) the exponential distribution $Exp(1/100)$ ($u = 200$), ii) the gamma distribution $\Gamma(30,1/4)$ ($u = 150$), iii) the Gumbel distribution $Gumbel(100,50)$ ($u = 220$).

For the chosen d.f. F_ε and the threshold u, we carried out the simulations according to the following scheme: *step 1)* we simulated a sample of size 10000 from F_ε, *step 2)* for the chosen threshold u and the obtained sample $\varepsilon_1,...,\varepsilon_{10000}$, we created the sample $\varepsilon_1 - u,...,\varepsilon_{10000} - u$, *step 3)* we chose those of $\varepsilon_t - u$, which satisfied $\varepsilon_t - u > 0$, and denoted the obtained sample by $y_1,..., y_k$, *step 4)* we calculated an average of $y_1,..., y_k$ and this way, we obtained the value of the estimate of the parameter β (see (4)), *step 5)* by using the Kolmogorov test, we checked the accordance of the sample $y_1,..., y_k$ with the exponential distribution, *step 6)* we generated 10000 samples of size 10000 from F_ε; in this way, we obtained (by proceeding as in the steps 2)-4)) 10000 estimates of the parameter β, *step 7)* we calculated an average of 10000 estimates of β obtained in the step 6); we took this average for a real value of the parameter β, *step 8)* by using the value of β (calculated in the step 7)) and the formula in (3), we obtained the approximated quantiles of the ranks 0.95, 0.99, 0.999, 0.9999, *step 9)* we repeated 100 times the simulations from the steps 1)-4) and, by applying (13), (14), we obtained the realizations of the 95% frequentistic confidence intervals for the quantiles of appropriate ranks, *step 10)* we repeated 100 times the simulations from the steps 1)-4) and, by applying (16), (17), we obtained the realizations of the 95% Bayesian confidence intervals for the quantiles of appropriate ranks.

The results of our simulations in the case of the frequentistic approach are:

For the case in i):

The exact quantiles from the $Exp(1/100)$ distribution are as follows:

Quantile rank	0.95	0.99	0.999	0.9999
Exact quantile value	299.6	460.5	690.8	921.0

The value of β, obtained by the MC method (see the step 7)): $\beta = 100.03$

The approximated quantiles (see the step 8)) are as follows:

Quantile rank	0.95	0.99	0.999	0.9999
Approximated quantile	299.6	460.6	690.9	921.3

The estimate of β, obtained according to the step 3): $\hat{\beta} = 99.95$

The result of the Kolmogorov test on the accordance of the sample $y_1,...,y_k$ with the $Exp(1/100.03)$ distribution: p-value$= 0.9326$; it means that we may accept the approximation by the exponential distribution.

For the case in ii):

The exact quantiles from the $\Gamma(30,1/4)$ distribution are as follows:

Quantile rank	0.95	0.99	0.999	0.9999
Exact quantile value	158.2	176.8	199.2	219.0

The value of β, obtained by the MC method (see the step 7)): $\beta = 12.38$

The approximated quantiles (see the step 8)) are as follows:

Quantile rank	0.95	0.99	0.999	0.9999
Approximated quantile	157.5	177.5	206.0	234.5

The estimate of β, obtained according to the step 3): $\hat{\beta} = 11.94$

The result of the Kolmogorov test on the accordance of the sample $y_1,...,y_k$ with the $Exp(1/12.38)$ distribution: p-value$= 0.4749$; it means that we may accept the approximation by the exponential distribution.

For the case in iii):

The exact quantiles from the $Gumbel(100,50)$ distribution are as follows:

Quantile rank	0.95	0.99	0.999	0.9999
Exact quantile value	248.5	330.0	445.4	560.5

The value of β, obtained by the MC method (see the step 7)): $\beta = 51.15$

The approximated quantiles (see the step 8)) are as follows:

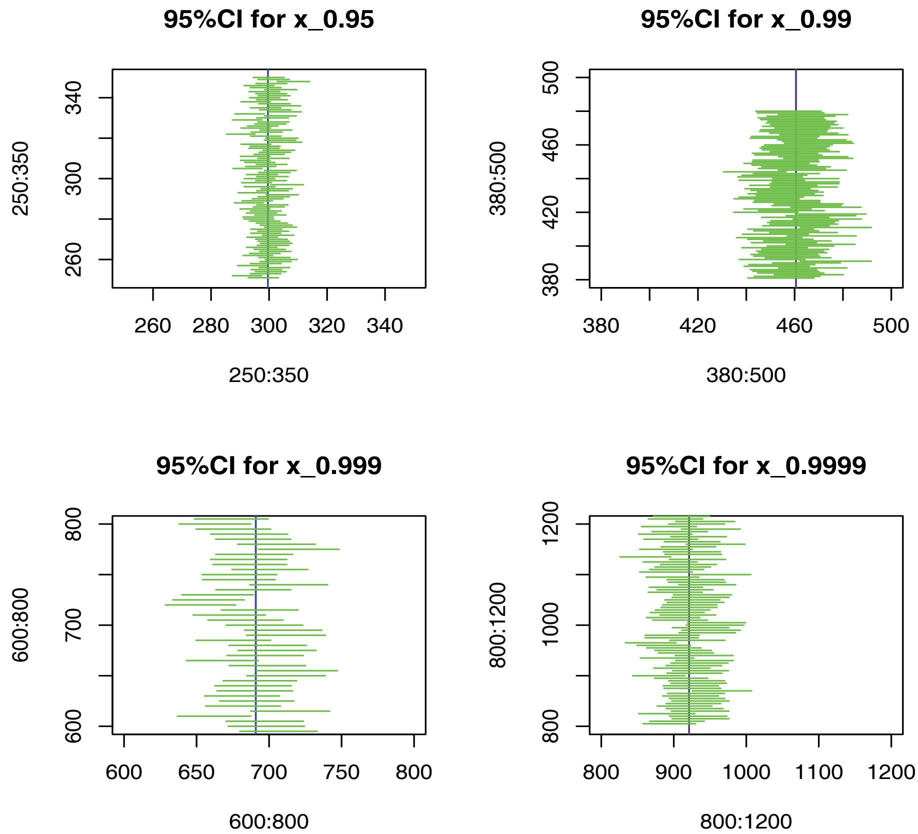
Quantile rank	0.95	0.99	0.999	0.9999
Approximated quantile	248.2	330.5	448.3	566.0

The estimate of β, obtained according to the step 3): $\hat{\beta} = 52.41$

The result of the Kolmogorov test on the accordance of the sample $y_1,\ldots,y_k$ with the $Exp(1/51.15)$ distribution: p-value$=0.3043$; it means that we may accept the approximation by the exponential distribution.

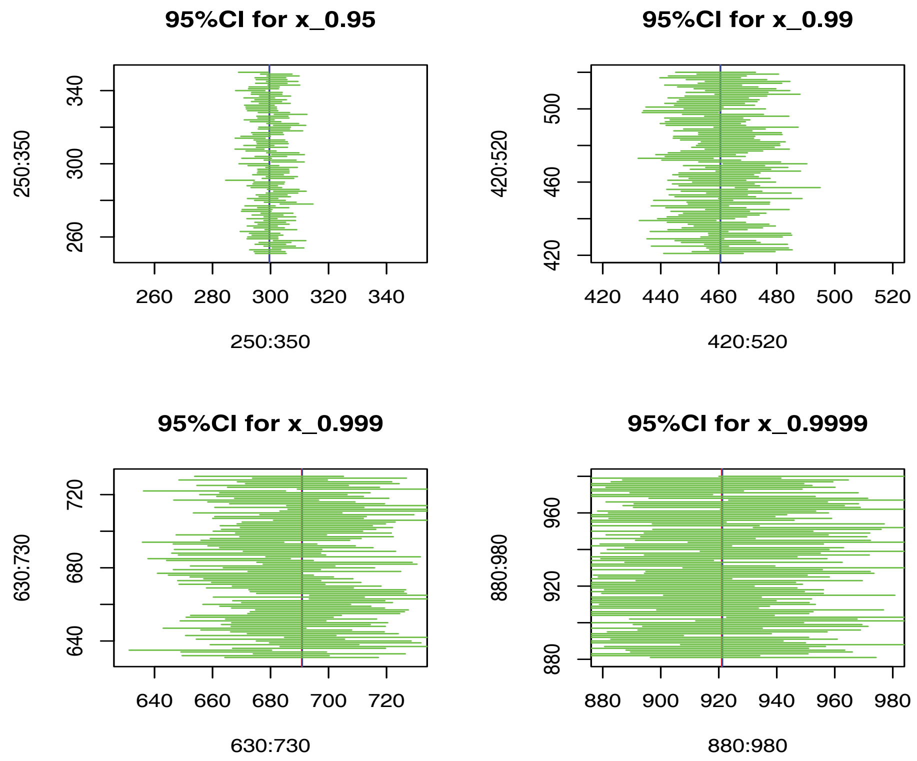
Below, we give the graphical presentation of the realizations of the frequentistic and Bayesian confidence intervals for the quantiles of appropriate ranks. For brevity, we only show the results for the case in i):

Figure 2. The realizations of the $(1-\gamma)\cdot 100\% = 95\%$ frequentistic confidence intervals for the quantiles of appropriate ranks in the case i)



Source: own preparation

Figure 3. The $(1-\gamma)\cdot 100\% = 95\%$ Bayesian confidence intervals for the quantiles $x_{0.95}$, $x_{0.99}$, $x_{0.999}$, $x_{0.9999}$ in the case i)



Source: own preparation

FINAL CONCLUSIONS

Conclusions concerning the obtained results for the frequentistic confidence intervals are as follows: a) the received confidence intervals estimate the corresponding quantiles reasonably well, b) in the case where the approximation by the exponential distribution is appropriate, the estimation errors are small, c) the following procedure can be proposed: choose (by applying the Kolmogorov test) the threshold u in such a manner that the sample $y_1,\dots,y_k$, where $y_i = \varepsilon_i - u \mid \varepsilon_i > u$ is i.i.d. and comes from the exponential distribution, and compute the approximated quantile and the realizations of the approximated confidence intervals for the quantile by means of the formulas derived in Section “Interval estimation of the quantile - the proposed estimation procedures”.

REFERENCES

Katz R. W., Parlange M. B. and Naveau P. (2002) Statistics of extremes in hydrology, Advances in water resources, 25, pp. 1287-1304.

Knox J. C and Kundzewicz Z. W. (1997) Extreme hydrological events, palaeo-information and climate change, Hydrological Sciences-Journal-des Sciences Hydrologiques, 42, pp. 765-779.

Shiau J. T. (2003) Return period of bivariate distributed extreme hydrological events, Stochastic environmental research and risk assessment, 17, pp. 42-57.

Van Gelder P. H. A. J. M. (1996) A Bayesian analysis of extreme water levels along the Dutch coast using flood historical data, Stochastic Hydraulics, 7, pp. 243-251.

Van Gelder P. H. A. J. M., Roos, A. and Tonneijck, M. R. (1995) On the probabilistic design of dikes in the Netherlands, Applications of Statistics and Probability, 3, pp. 1505-1508.