ESTIMATION OF THE LOWER AND UPPER QUANTILES OF GUMBEL DISTRIBUTION: AN APPLICATION TO WIND SPEED DATA

AYDIN, D.

Department of Statistics, Sinop University, Sinop, Turkey
e-mail: daydin@sinop.edu.tr

(Received 29th Jul 2017; accepted 21st Nov 2017)

Abstract. In this paper, we consider different estimators of the quantiles of two-parameter Gumbel distribution. We use methodologies known as maximum likelihood, modified maximum likelihood and probability weighted moment to obtain the estimators of the quantiles. We compare the performances of the estimators with respect to bias and mean square error criteria via Monte Carlo simulation study. Their robustness properties are also examined in the presence of data anomalies. In the real data analysis part of the study, the seasonal maximum daily wind speed data from Sinop station (Turkey) in 2015 is considered. It is modelled by using two-parameter Gumbel distribution and analysed to compare the performances of the methodology presented in the study. All in all, the results of simulations and the real data application show that the maximum likelihood and modified maximum likelihood estimators, which have similar performance, provide better performance than the probability weighted moment estimator does in both obtaining estimates of the quantiles of Gumbel distribution and modelling of the data for almost all cases.
Keywords: *Gumbel distribution, quantile, modelling extreme events, efficiency, robustness*

Introduction

Estimation of the quantiles of any distribution is very important in real life problems. As examples due to Modarres et al. (2002) "*Estimates of the upper quantiles of the distribution of a risk factor or an exposure index are commonly used to assess the risk to human health as a result of exposure to chemicals and microbes in the environment, or to determine if concentration levels of contaminants exceed specified limits*" and Goel et al. (2004) "*Extreme wind quantiles are needed to calculate design values of wind load effect on structures*". Therefore, in literature, various different distributions have been considered by many authors in the context of extreme value analysis, for example Gumbel distribution, Wakeby distribution, Generalized Pareto distribution, Generalized Extreme-Value distribution, Log-normal, Log-logistic and Log-double exponential distributions and Frechet distribution (Landwehr and Matalas, 1979a; Landwehr and Matalas, 1979b; Hosking and Wallis, 1987; Martins and Stedinger, 2000; Modarres et al., 2002; Koutsoyiannis, 2004).

The Gumbel known as the Extreme Value type I distribution, first proposed by Gumbel (1941), is one of the most widely probabilistic models used in modelling the extreme events in many research studies, for example, total snowfall, maximum snow, air pollution and maximum daily flood discharges (Simiu et al., 2001; Koutsoyiannis, 2004; Graybeal and Leathers, 2006; Ercelebi and Toros, 2009; Aydin and Senoglu, 2015). On the other hand, in the literature, although the most widely used statistical distribution for modelling the wind speed data is Weibull, it may not provide better fitting for all wind regimes. For this reason, different distributions are used for modelling the wind speed data (Brano et al., 2011; Kantar and Usta, 2015; Alavi et al., 2016; Jung et al., 2017). For example, Gumbel distribution has also been used to both

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

estimate extreme wind speed required for the determination of the wind turbine class in the wind power industry and evaluate the wind energy potential required designing a wind turbine (Hong et al., 2013; Kang et al., 2015). Additionally, Lee et al. (2012) reported that the Gumbel distribution is more reliable than the Weibull distribution in modelling the extreme wind speeds. Martin et al. (2014) showed that the Gumbel distribution estimates wind speed more accurately than the Weibull distribution does.

Aim of this paper is to obtain the estimators of the lower and the upper quantiles of the Gumbel distribution. The estimators of the quantiles are obtained by using the well-known and widely used maximum likelihood (*ML*) methodology. The likelihood equations, however, do not have explicit solutions. Therefore, we use two different approaches to solve them. The first approach is iterative and other one is non-iterative which is called as modified maximum likelihood (*MML*). We also use, the probability weighted moment (*PWM*), which is very popular methodology in hydrology and climatology. The reason of using *PWM* is its conceptual simplicity, implementation and good performance. Furthermore, wind speed data obtained from the Turkish State Meteorological Service is modelled by Gumbel distribution and analysed to show the performance of the considered estimation methods.

Materials and methods

The seasonal wind speed data

In this study, the seasonal wind speed data recorded at the heights of 10 m in maximum daily basis in 2015 in Sinop station (Turkey) is analysed. Geographical coordinates for this station are given as

Station	Region in Turkey	Latitude (N)	Longitude (E)	Altitude (m)
Sinop	North	42°01'44"	35°09'19"	32

In *Table 1*, descriptive statistics which are mean, minimum (*Min*), maximum (*Max*), median, standard deviation (*SD*) and range for seasonal maximum daily wind speed data (m/s) are given.

***Table 1**. Summary of the descriptive statistics for the seasonal maximum daily wind speed data*

Season	Mean	*Min*	*Max*	Median	*SD*	Range	*n*
Winter	10.0189	3.2000	24.6000	9.8000	4.2694	21.4000	90
Spring	9.5087	4.6000	18.3000	9.2000	2.9934	13.7000	92
Summer	7.8228	4.4000	17.3000	7.6000	2.0648	12.9000	92
Autumn	8.9560	4.5000	19.5000	8.3000	2.9676	15.0000	91

According to the results given in *Table 1*, range (which is defined as the difference between the highest and the lowest value) is the largest in winter (December-January), and is the smallest in summer (June-August) as expected. Similar comments can also be done for *SD* which is another measure of variability.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

Gumbel distribution

Probability density function (*pdf*) $f(x)$ and the cumulative density function (*cdf*) $F(x)$ of the Gumbel distribution are given by *Equation 1*

$$f(x) = \frac{1}{\delta} e^{-((x-\theta)/\delta + e^{-(x-\theta)/\delta})},\ x \epsilon \mathbb{R} \tag{Eq. 1}$$

and *Equation 2*

$$F(x) = e^{-e^{-(x-\theta)/\delta}}, \tag{Eq. 2}$$

respectively. Here, $\theta \in \mathbb{R}$ is location parameter and $\delta > 0$ is scale parameter. The location parameter θ is also the mode of the distribution. Inverse of the *cdf* in *Equation 2*, i.e. $x(F)$, is obtained as follows (*Eq. 3*)

$$x(F) = \theta - \delta \ln(-\ln F). \tag{Eq. 3}$$

The moment generating function of Gumbel distribution is given by *Equation 4:*

$$M(t) = e^{\theta t} \Gamma(1 - \delta t),\ t < 1/\delta. \tag{Eq. 4}$$

Mean ($E(X)$), variance ($Var(X)$), skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) values of Gumbel distribution are given as follows:

$E(X)$	$Var(X)$	$\sqrt{\beta_1}$	β_2
$\theta + \delta\gamma$	$\pi^2/6\,\delta^2$	1.14	5.4

where γ is Euler's constant defined by *Equation 5:*

$$\gamma = -\int_0^\infty \ln x\, e^{-x} dx. \tag{Eq. 5}$$

Gumbel distribution is related to the Weibull distribution. In particular, if Y has a Weibull distribution with shape parameter ϕ and scale parameter λ, then (*Eq. 6*)

$$X = -log(Y) \tag{Eq. 6}$$

has a Gumbel distribution with the location parameter $\theta = -log(\lambda)$ and the scale parameter $\delta = 1/\phi$.

The graphs of the *pdf* of the Gumbel distribution for some selected values of the location parameter θ and the scale parameter δ are given in *Fig. 1*. It is clear from *Fig. 1* that Gumbel distribution is unimodal and skewed to the right.

Estimation of quantiles

Let X_q be *q*-th quantile of the Gumbel random variable X. It is defined as (*Eq. 7*)

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

$$X_q = \theta - \delta \ln(-\ln q), 0 < q < 1, \quad \text{(Eq. 7)}$$

see *Equation 3*. Estimator of the quantile X_q, i.e. $\hat{X}_q$, is obtained by substituting the estimators of the parameters θ and δ in *Equation 7*.

In the following subsections, we briefly describe the estimation techniques mentioned before for estimating the quantiles of the Gumbel distribution.

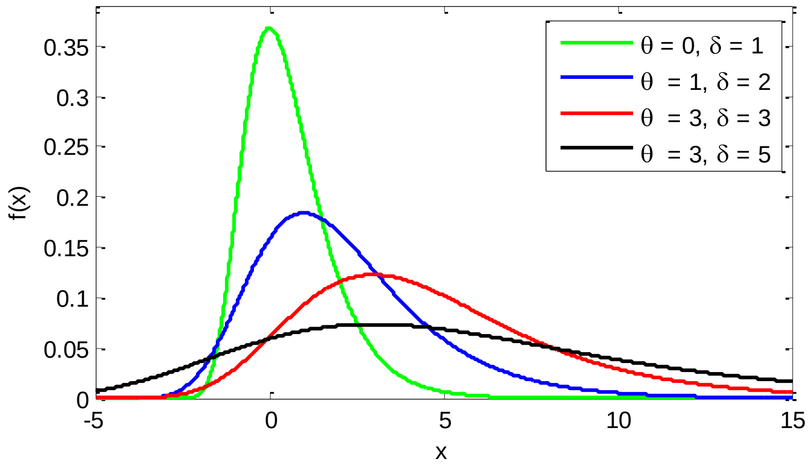

***Figure 1**. Plots of the Gumbel distribution for some selected and values*

The method of maximum likelihood

The *ML* estimators $\hat{\theta}$ and $\hat{\delta}$ of the parameters θ and δ are the solutions of the following likelihood equations (*Eqs. 8 and 9*)

$$\frac{\partial \ln L(\theta,\delta)}{\partial \theta} = \sum_{i=1}^{n} \frac{1}{\delta} - \frac{1}{\delta} \sum_{i=1}^{n} g(z_i) = 0 \quad \text{(Eq. 8)}$$

$$\frac{\partial \ln L(\theta,\delta)}{\partial \delta} = -\frac{n}{\delta} + \frac{1}{\delta} \sum_{i=1}^{n} z_i - \frac{1}{\delta} \sum_{i=1}^{n} z_i g(z_i) = 0 \quad \text{(Eq. 9)}$$

where $g(z_i) = e^{-z_i}$ and $z_i = (x_i - \theta)/\delta$. It is obvious that explicit solutions of the likelihood equations cannot be obtained because of the nonlinear term $g(.)$. Therefore, we can use two different approaches to solve the likelihood equations. One is iterative and the other one is non-iterative given in the next subsection.

The method of modified maximum likelihood

The *MML* estimators of parameters θ and δ are obtained by linearizing the non-linear term $g(z_i)$ in the likelihood equations in (*Eq. 8*) and (*Eq. 9*). We linearize the likelihood equations by using the first two terms of Taylor series expansion around the expected values of the standardized order statistics, i.e. $t_{(i)} = E(z_{(i)})$ and $z_{(i)} = (x_{(i)} - \theta)/\delta$, (Tiku, 1967; Tiku, 1968). Solutions of these modified likelihood equations are the following *MML* estimators (*Eq. 10*):

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

$$\hat{\theta}_{MML} = K + L\hat{\delta}_{MML} \text{ and } \hat{\delta}_{MML} = \frac{-B+\sqrt{B^2-4nC}}{2\sqrt{n(n-1)}}, \quad \text{(Eq. 10)}$$

where $K = \frac{1}{m}\sum_{i=1}^{n}\beta_i x_{(i)}$, $L = \frac{\Delta}{m}$, $\Delta = \sum_{i=1}^{n}\Delta_i$, $\Delta_i = (\alpha_i - 1)$, $m = \sum_{i=1}^{n}\beta_i$, $\mathrm{B} = \sum_{i=1}^{n}\Delta_i\left(x_{(i)} - \hat{\theta}_{MML}\right)$, $C = \sum_{i=1}^{n}\beta_i\left(x_{(i)} - \hat{\theta}_{MML}\right)^2$, $\alpha_i = e^{-t_{(i)}} + t_{(i)}e^{-t_{(i)}}$, $\beta_i = -e^{-t_{(i)}}$ and $t_{(i)} = -\ln\left(-\ln\left(\frac{i}{n+1}\right)\right)$, $i = 1,2,\cdots,n$.

The *MML* estimators are asymptotically equivalent to the *ML* estimators. Therefore, they are asymptotically fully efficient under the regularity conditions. They have high efficiencies even for small sample sizes. They are also robust to plausible deviations from the assumed distribution and also to the presence of the outliers in the data set (Tiku and Suresh, 1992; Vaughan and Tiku, 2000).

The method of probability weighted moment

The *PWM* estimators of θ and δ are obtained as (*Eq. 11*)

$$\hat{\theta}_{PWM} = \widehat{M}_{(0)} - \gamma\hat{\delta}_{PWM} \text{ and } \hat{\delta}_{PWM} = \frac{\widehat{M}_{(0)}-2\widehat{M}_{(1)}}{\ln 2}, \quad \text{(Eq. 11)}$$

respectively (Greenwood et al., 1979; Landwehr et al., 1979a). Here, γ is Euler's constant and $\widehat{M}_{(k)}$ is an unbiased estimate of $M_{(k)}$ (*Eq. 12*):

$$\widehat{M}_{(k)} = \frac{1}{n}\sum_{i=1}^{n} x_{(i)}\frac{(n-i)!(n-1-k)!}{(n-1)!(n-i-k)!} \quad \text{(Eq. 12)}$$

where $x_{(i)}$ are *i*-th ordered observations and $M_{(k)} = M_{1,0,k}$ is calculated from the following *PWM*s equality for $i,j,k \in \mathbb{R}$ (*Eq. 13*):

$$M_{i,j,k} = E\left(X^i\left(F(X)\right)^j\left(1-F(X)\right)^k\right) = \int_0^1\left(x(F)\right)^i F^j(1-F)^k dF. \quad \text{(Eq. 13)}$$

Here, $F(X)$ is the *cdf* of the random variable X and $x(F)$ is the corresponding inverse distribution function.

Simulation study

To compare the performances of *ML*, *MML* and *PWM* estimators of the *q*-th quantile of the Gumbel distribution X_q, an extensive Monte Carlo simulation study is designed and conducted with respect to their biases and mean squared error (*MSE*) for different sample sizes and quantile values. Bias and *MSE* for $\hat{X}_q$ are calculated as (*Eqs. 14* and *15*):

$$Bias\left(\hat{X}_q\right) = 1/n\sum_{i=1}^{n}\left(X_q - \hat{X}_{q_i}\right) \quad \text{(Eq. 14)}$$

$$MSE\left(\hat{X}_q\right) = 1/n\sum_{i=1}^{n}\left(X_q - \hat{X}_{q_i}\right)^2 \quad \text{(Eq. 15)}$$

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

respectively. Here, n is the number of replication and $\hat{X}_{q_i}$ is the estimate of X_q in i-th replication. We also calculate the relative efficiencies (*RE*) of the *ML* estimator with respect to the *MM*L and *PWM* estimators of X_q, i.e. (*Eq. 16*),

$$RE = \left(MSE(\hat{X}_q)/MSE(\hat{X}_{q,ML})\right) \times 100. \quad \text{(Eq. 16)}$$

We consider the sample sizes, $n = 5$, 10, 50, 100 and 1000 and quantile values, $q = 0.01$, 0.05, 0.10, 0.90, 0.95 and 0.99. Bias and *MSE* values of the estimators are computed based on $[\![100{,}000/n]\!]$ replications. Here, $[\![.]\!]$ indicates the greatest integer value. Without loss of generality, it is assumed that the location parameter $\theta = 0$ and the scale parameter $\delta = 1$.

Here, the quantile estimates $\hat{X}_q$ are computed by substituting the estimates of the parameters θ and δ in *Equation 7*, i.e. (*Eq. 17*),

$$\hat{X}_q = \hat{\theta} - \hat{\delta}\ln(-\ln q),\ 0 < q < 1. \quad \text{(Eq. 17)}$$

Robustness properties of the estimators

To compare the robustness properties of estimators mentioned above, the efficiencies of the *ML*, *MML* and *PWM* estimators of the quantiles of the Gumbel distribution are examined via Monte-Carlo simulation study when there exist data anomalies, such as misspecification of the model and presence of the outliers in the data set. For this purpose, Gumbel with location parameter $\theta = 0$ and scale parameter $\delta = 1$, i.e., $G(\theta = 0, \delta = 1)$ is assumed as true model, and consider the following alternative models:

(i) *Model I*: Misspecified model: $G(\theta = 0, \delta = 2)$,
(ii) *Model II*: Misspecified model: $G(\theta = 1, \delta = 1)$,
(iii) *Model III*: Contamination model: $0.90G(\theta = 0, \delta = 1) + 0.10U(-3{,}3)$,
(iv) *Model IV*: Mixture model: $0.90G(\theta = 0, \delta = 1) + 0.10G(\theta = 0, \delta = 2)$,
(v) *Model V*: Dixon's outlier model:
$$(n-r)G(\theta = 0, \delta = 1) + rG(\theta = 0, \delta = 2),\ r = [\![0.1n + 0.5]\!].$$

Model evaluation

The suitability of estimates of Gumbel distribution in fitting the wind speed data can be evaluated by numerical methods. For this purpose, the root mean square error (*RMSE*) and coefficient of determination (R^2) are used and they are calculated by using the following formulas (*Eqs. 18* and *19*)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{F}(X_{(i)}) - u_i\right)^2} \quad \text{(Eq. 18)}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{F}(X_{(i)}) - u_i\right)^2}{\sum_{i=1}^{n}\left(\hat{F}(X_{(i)}) - \bar{\hat{F}}(X_{(i)})\right)^2}, \quad \text{(Eq. 19)}$$

respectively (Nash, 1970; Barrett, 1974; Jöreskog and Sörbom, 1981; Willmott, 1982). Here, $\hat{F}(X_{(i)})$ is the estimated value of the *cdf* for the i-th order statistics. u_i is the

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

expected value of $\hat{F}(X_{(i)})$ and is equivalent to $i/(n+1)$. $\bar{\hat{F}}$ is the mean of the estimated *cdf*s $\hat{F}$, i.e., $\bar{\hat{F}} = \frac{1}{n}\sum_{i=1}^{n} \hat{F}$. It should be noted that lower value of $RMSE$ and the higher the R^2 indicate better fitting to the data.

Results

Simulation results

To compare the performances of the methods presented in the previous section, results of some simulation studies are presented in *Table 2*. All the computations were performed by using MATLAB R2010a. It should be noted that *Table 2* gives the bias and *MSE* values of $\hat{X}_q$ for both the lower (i.e., $X_{0.01}$, $X_{0.05}$ and $X_{0.10}$) and the upper (i.e., $X_{0.90}$, $X_{0.95}$ and $X_{0.95}$) tail quantiles. It is observed that the *PWM* estimator of X_q shows better performance than the other estimators do with respect to bias criterion for all values of *q* even for small sample sizes (Landwehr et al., 1979a). As the sample size *n* increases, all the estimators show more or less the same performance.

The *ML* estimator outperforms the other estimators almost in all cases in terms of the *MSE* criterion. It should be noted that both *MSE* and *Bias* decrease while the sample size *n* increases which signifies that all of these estimators are consistent. Especially for $n > 5$, *MSE* values of *ML* and *MML* estimators are quite close to one another as expected. Also, the *MSE*s of lower tail quantiles are smaller than *MSE*s of upper tail quantiles since the Gumbel distribution is skewed to the right, see *Table 2*.

Table 2*. Simulated Bias, MSE and RE values of* $\hat{X}_q$

		$q = 0.01, X_q = -1.53$			$q = 0.05, X_q = -1.10$			$q = 0.10, X_q = -0.83$		
n	**Method**	***Bias***	***MSE***	***RE***	***Bias***	***MSE***	***RE***	***Bias***	***MSE***	***RE***
5	*ML*	0.3274	0.4978	100.0	0.2595	0.3614	100.0	0.2106	0.2922	100.0
	MML	0.2339	0.6746	135.5	0.2259	0.4022	111.2	0.2094	0.3224	110.3
	PWM	0.0068	0.5969	119.9	0.0077	0.4025	111.3	-0.0008	0.3138	107.3
10	*ML*	0.1528	0.2093	100.0	0.1210	0.1526	100.0	0.0931	0.1277	100.0
	MML	0.1301	0.2077	99.2	0.1212	0.1535	100.5	0.1077	0.1303	102.0
	PWM	-0.0012	0.2591	123.7	-0.0046	0.1794	117.5	-0.0097	0.1432	112.1
50	*ML*	0.0452	0.0356	100.0	0.0265	0.0281	100.0	0.0190	0.0215	100.0
	MML	0.0468	0.0358	100.5	0.0315	0.0284	101.0	0.0261	0.0219	101.8
	PWM	0.0146	0.0452	126.9	0.0039	0.0342	121.7	-0.0019	0.0255	118.6
100	*ML*	0.0187	0.0186	100.0	0.0175	0.0130	100.0	0.0080	0.0108	100.0
	MML	0.0212	0.0187	100.5	0.0206	0.0131	100.7	0.0120	0.0109	100.9
	PWM	0.0001	0.0233	125.2	0.0075	0.0154	118.4	-0.0023	0.0128	118.5
1000	*ML*	0.0020	0.0018	100.0	0.0021	0.0012	100.0	-0.0059	0.0010	100.0
	MML	0.0026	0.0018	100.0	0.0026	0.0012	100.0	-0.0053	0.0010	100.0
	PWM	-0.0003	0.0022	122.2	0.0022	0.0015	125.0	-0.0084	0.0014	140.0

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

***Table 2**. (Continued)*

		$q = 0.90, X_q = 2.25$			$q = 0.95, X_q = 2.97$			$q = 0.99, X_q = 4.60$		
n	**Method**	***Bias***	***MSE***	***RE***	***Bias***	***MSE***	***RE***	***Bias***	***MSE***	***RE***
5	*ML*	-0.2794	1.0948	100.0	-0.3892	1.6518	100.0	0.6387	3.4784	100.0
	MML	0.1212	1.7614	160.8	0.1196	3.0598	185.2	-0.0938	7.9007	227.1
	PWM	-0.0037	1.3347	121.9	-0.0059	2.0758	125.6	-0.0119	4.5372	130.4
10	*ML*	-0.1378	0.5360	100.0	-0.1894	0.8217	100.0	0.3291	1.6893	100.0
	MML	0.0441	0.5872	109.5	0.0328	0.8950	108.9	0.0183	1.7976	106.4
	PWM	-0.0054	0.6220	116.0	0.0029	0.9923	120.7	0.0031	2.0887	123.6
50	*ML*	-0.0235	0.1033	100.0	-0.0363	0.1693	100.0	0.0617	0.3339	100.0
	MML	0.0068	0.1053	101.9	-0.0001	0.1714	101.2	0.0127	0.3383	101.3
	PWM	0.0021	0.1240	120.0	0.0019	0.2013	118.9	0.0062	0.4175	125.0
100	*ML*	-0.0202	0.0511	100.0	-0.0241	0.0796	100.0	0.0269	0.1589	100.0
	MML	-0.0066	0.0518	101.3	-0.0080	0.0804	101.0	0.0046	0.1616	101.6
	PWM	-0.0058	0.0577	112.9	-0.0015	0.0945	118.7	-0.0013	0.1944	122.3
1000	*ML*	0.0070	0.0052	100.0	-0.0040	0.0087	100.0	-0.0202	0.0160	100.0
	MML	0.0071	0.0053	101.9	-0.0020	0.0087	100.0	-0.0233	0.0160	100.0
	PWM	0.0111	0.0061	117.3	-0.0033	0.0103	118.3	-0.0157	0.0194	120.7

Robustness results

To assess the robustness properties of the methods mentioned earlier, results of some simulation studies are given in *Table 3*. It should be noted that different values of n are used in the simulation study, however, here the results are just reproduced for $n = 50$ as an illustration.

For $q \leq 0.10$, the *MML* estimator is the best among the others for models I, IV and V, the *PWM* estimator is more efficient than the others for models II-III. For $q > 0.10$, the *ML* outperforms the other methods for almost all alternative models (except for models II and III) with respect to the *MSE* criterion. The *PWM* estimator is the best for model III and the *MML* estimator performs better than the other estimators do for model I. However, all the estimators have substantial bias for all the alternative models.

Model evaluation results

In this study, to illustrate the practical use of the considered estimation methods in the previous section, we use the seasonal maximum daily wind speed modelled by the Gumbel distribution. Before analysing the data set, we evaluated the suitability of Gumbel distribution to fit the wind speed data by using Q–Q plots (which is the graphical technique) and Kolmogorov–Smirnov (*KS*) test, see *Table 4*.

Table 4 shows that computed values of the *KS* test given by the *ML*, *MML* and the *PWM* of Gumbel distribution are less than the theoretical values (which are $KS_{0.05,90} = 0.1434$, $KS_{0.05,91} = 0.1426$ and $KS_{0.05,92} = 0.1418$). Therefore, the results of the *KS* test and Q–Q plots are showed that the Gumbel distribution provides a plausible model for the data, see *Fig. 2*.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

***Table 3**. Simulated Bias, MSE and RE values of $\ddot{X}_q$ for the alternative models when $n = 50$*

	$q = 0.01, X_q = -1.53$			$q = 0.05, X_q = -1.10$			$q = 0.10, X_q = -0.83$		
	Model I								
Method	***Bias***	***MSE***	***RE***	***Bias***	***MSE***	***RE***	***Bias***	***MSE***	***RE***
ML	1.4590	2.2803	100.0	1.0569	1.2143	100.0	0.7775	0.6903	100.0
MML	0.7675	0.7154	31.3	0.4928	0.3289	27.0	0.2963	0.1676	24.2
PWM	1.5202	2.5148	110.2	1.0999	1.3342	109.8	0.8177	0.7714	111.7
	Model II								
ML	-1.0305	1.0985	100.0	-1.0263	1.0793	100.0	-1.0204	1.0639	100.0
MML	-1.8477	3.4365	312.8	-1.8036	3.2730	303.2	-1.7724	3.1635	297.3
PWM	-0.9986	1.0451	95.1	-1.0010	1.0349	95.8	-0.9987	1.0242	96.2
	Model III								
ML	0.3116	0.1984	100.0	0.2517	0.1341	100.0	0.2129	0.0977	100.0
MML	0.2881	0.1781	89.7	0.2263	0.1170	87.2	0.1863	0.0830	84.9
PWM	0.1800	0.0966	48.6	0.1518	0.0729	54.3	0.1387	0.0585	59.8
	Model IV								
ML	0.1840	0.1194	100.0	0.1328	0.0775	100.0	0.1134	0.0592	100.0
MML	0.1669	0.1013	84.8	0.1142	0.0644	83.1	0.0941	0.0495	83.5
PWM	0.1823	0.1022	85.5	0.1319	0.0666	86.0	0.1146	0.0506	85.5
	Model V								
ML	0.1885	0.1181	100.0	0.1365	0.0753	100.0	0.1045	0.0565	100.0
MML	0.1699	0.0982	83.1	0.1176	0.0626	83.1	0.0845	0.0464	82.1
PWM	0.1880	0.0989	83.7	0.1358	0.0636	84.4	0.1075	0.0473	83.7
	$q = 0.90, X_q = 2.25$			$q = 0.95, X_q = 2.97$			$q = 0.99, X_q = 4.60$		
	Model I								
Method	***Bias***	***MSE***	***RE***	***Bias***	***MSE***	***RE***	***Bias***	***MSE***	***RE***
ML	-2.2027	5.3105	100.0	-2.8807	8.9102	100.0	-4.4939	21.4414	100.0
MML	-1.7585	3.5304	66.4	-2.2435	5.6380	63.2	-3.3523	12.4176	57.9
PWM	-2.2650	5.6816	106.9	-2.9417	9.4012	105.5	-4.6243	22.9100	106.8
	Model II								
ML	-0.9698	1.0465	100.0	-0.9664	1.0909	100.0	-0.9671	1.2431	100.0
MML	-1.4403	2.2183	211.9	-1.3730	2.0825	190.8	-1.2232	1.8460	148.4
PWM	-0.9947	1.1133	106.3	-1.0040	1.1985	109.8	-1.0202	1.4323	115.2
	Model III								
ML	-0.2171	0.1857	100.0	-0.3361	0.3271	100.0	-0.5201	0.7733	100.0
MML	-0.2573	0.2131	114.7	-0.3806	0.3708	113.3	-0.5715	0.8483	109.6
PWM	-0.0351	0.1184	63.7	-0.0889	0.1945	59.4	-0.1578	0.4155	53.7
	Model IV								
ML	-0.2665	0.2374	100.0	-0.3518	0.3866	100.0	-0.5252	0.8493	100.0
MML	-0.2956	0.2541	107.0	-0.3837	0.4081	105.5	-0.5628	0.8738	102.8
PWM	-0.2577	0.2584	108.8	-0.3432	0.4109	106.3	-0.5089	0.8909	104.9
	Model V								
ML	-0.2543	0.2169	100.0	-0.3516	0.3819	100.0	-0.5543	0.8337	100.0
MML	-0.2840	0.2338	107.7	-0.3815	0.4003	104.8	-0.5900	0.8575	102.8
PWM	-0.2503	0.2384	109.8	-0.3373	0.4097	107.2	-0.5256	0.8639	103.6

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

***Table 4**. Computed values of KS test using the ML, MML and the PWM of Gumbel distribution for each season*

	Winter		**Spring**		**Summer**		**Autumn**	
Method	***KS***	***p*-value**	***KS***	***p*-value**	***KS***	***p*-value**	***KS***	***p*-value**
ML	0.0632	0.6089	0.0474	0.6147	0.0569	0.6112	0.0682	0.6071
MML	0.0643	0.6085	0.0465	0.6150	0.0595	0.6102	0.0699	0.6064
PWM	0.0575	0.6110	0.0509	0.6134	0.0487	0.6142	0.0618	0.6094

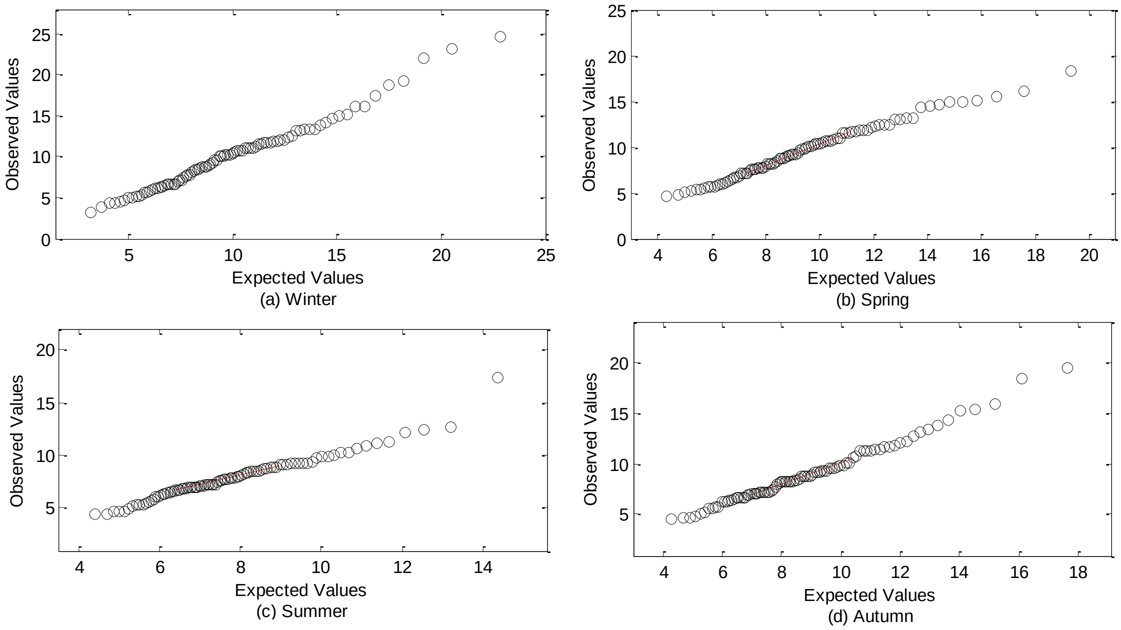

***Figure 2**. Q–Q plots of the seasonal maximum daily wind speed data for Gumbel distribution*

Then, it is purposed to determine a distribution providing better fit to wind speed data among Gumbel distribution based on the *MML*, *ML* or *PVM*. For this aim, the *ML*, *MML* and *PWM* estimates of the parameters and also R^2 and *RMSE* values of Gumbel distribution based on the estimators are calculated for each season. *Table 5* shows that the Gumbel distribution based on *MML* estimates provides the best fit to the spring and the summer, Gumbel distribution based on *PWM* estimates gives a better fit than the others for winter, Gumbel distribution based on *ML* estimates fit best for autumn, since the *RMSE* and R^2 values corresponding to these estimates are the lowest and the highest respectively, among the others.

Furthermore, in order to identify the distribution providing better fit to wind speed data by visual, histograms and fitted Gumbel probability plots for seasonal maximum daily wind speeds are used and results of analyses are presented in *Fig. 3*. It shows that the Gumbel distribution based on both *ML* and *MML* estimates also provides a better fit to the seasonal maximum daily wind speed data (except for winter) since curves of Gumbel probability plots of *ML* and *MML* estimates are almost superimposed. It should be noted that the results in *Table 4* are also consistent with graphs of the frequency histograms and fitted Gumbel probability plots based on the estimates in *Fig. 3*.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

***Table 5**. Estimates of the parameters and computed values of R^2 and RMSE corresponding the ML, MML and the PWM of Gumbel distribution for each season*

	Winter				Spring			
Method	$\hat{\theta}$	$\hat{\delta}$	***RMSE***	R^2	$\hat{\theta}$	$\hat{\delta}$	***RMSE***	R^2
ML	8.1074	3.2645	0.0262	0.9921	8.0938	2.4860	0.0273	0.9916
MML	8.1271	3.2842	0.0257	0.9924	8.1119	2.4981	0.0263	0.9923
PWM	8.0860	3.3487	0.0243	0.9930	8.0914	2.4554	0.0292	0.9905
	Summer				**Autumn**			
Method	$\hat{\theta}$	$\hat{\delta}$	***RMSE***	R^2	$\hat{\theta}$	$\hat{\delta}$	***RMSE***	R^2
ML	6.8945	1.6556	0.0269	0.9907	7.6448	2.2138	0.0222	0.9942
MML	6.9100	1.6601	0.0260	0.9914	7.6580	2.2257	0.0228	0.9939
PWM	6.9020	1.5954	0.0267	0.9912	7.6231	2.3094	0.0222	0.9939

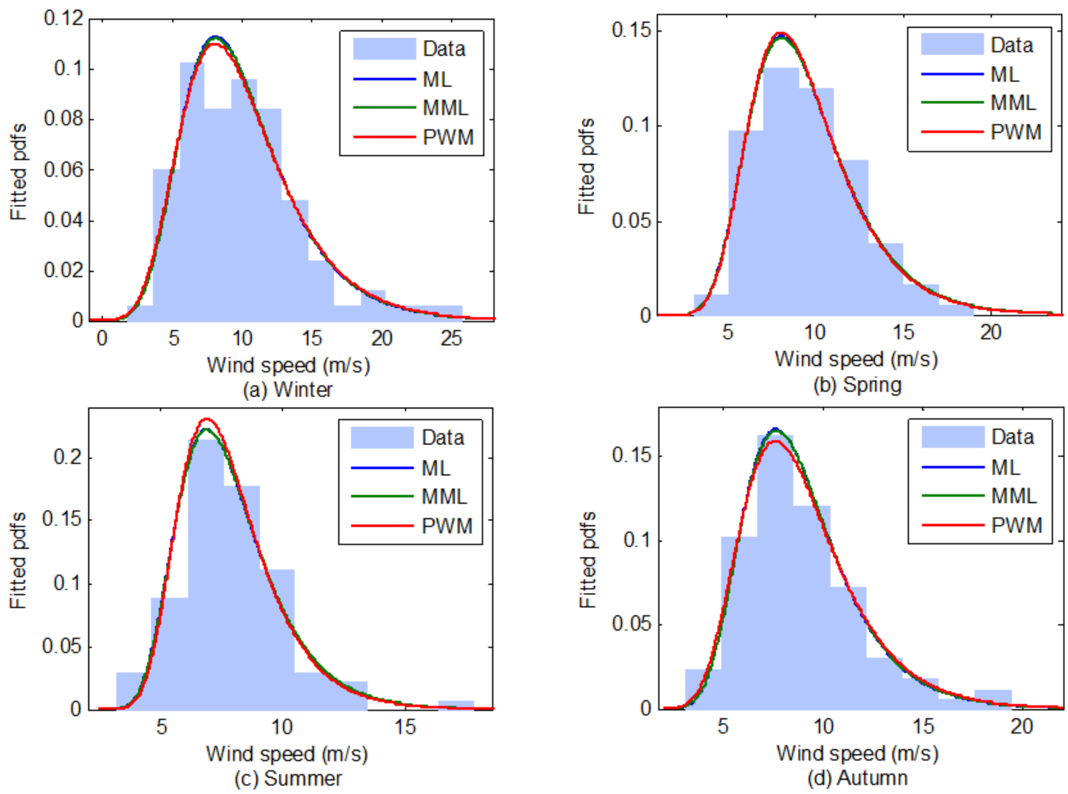

***Figure 3**. Histograms and fitted Gumbel probability plots based on ML, MML and PWM estimates superimposed for seasonal maximum daily wind speeds*

Results of quantiles estimates for wind speed data

In this part, performances of the estimators of quantiles are examined by using the considered methodologies for wind speed data recorded in Sinop. For this purpose, estimates of quantiles and their bootstrap standard deviations (*BSD*) are calculated for the values of *q* (i.e., 0.01, 0.05, 0.10, 0.90, 0.95 and 0.99) for each season, see *Table 6*.

According to the results presented in *Table 6*, in terms of the *BSD*, the *ML* estimate of X_q is the best with respect to *BSD* for summer (all values of *q*) and winter (values of *q*, i.e., $q \geq 0.90$) seasons. The *MML* estimate of X_q outperforms for autumn (all values

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

of *q*), winter (values of *q*, i.e., $q \leq 0.10$) and spring (values of *q*, i.e., $q \leq 0.10$). The *PWM* estimate of X_q has the best performance for spring (values of *q*, i.e., $q \geq 0.90$). Additionally, its *BSD* values of *MML* are quite close to *BSD* values of *ML* because of the asymptotic equivalence of the *ML* and the *MML* estimators (Bhattacharyya, 1985; Vaughan and Tiku, 2000; Senoglu and Tiku, 2002). This result is consistent with the simulation results presented in *Table 2*.

***Table 6**. Estimates of X_q and their BSD values for summer maximum daily wind speed data for all seasons*

	Winter						**Spring**					
	$q = 0.01$		$q = 0.05$		$q = 0.10$		$q = 0.01$		$q = 0.05$		$q = 0.10$	
Method	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***
ML	3.1562	0.3934	4.5790	0.3419	5.4186	0.3246	4.3358	0.2865	5.3875	0.2636	6.0368	0.2429
MML	3.1677	0.3834	4.5959	0.3351	5.4392	0.3208	4.3427	0.2755	5.4007	0.2547	6.0536	0.2370
PWM	3.0339	0.5303	4.4979	0.4381	5.3535	0.3973	4.3979	0.3640	5.4324	0.3274	6.0747	0.2905
	$q = 0.90$		$q = 0.95$		$q = 0.99$		$q = 0.90$		$q = 0.95$		$q = 0.99$	
Method	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***
ML	15.4433	0.8058	17.7021	0.9960	23.0077	1.3086	13.6726	0.5419	15.4400	0.6309	19.4613	0.9043
MML	15.5050	0.8149	17.7738	1.0049	23.1056	1.3192	13.7380	0.5638	15.5142	0.6521	19.5591	0.9375
PWM	15.5897	0.8757	17.8900	1.1216	23.2691	1.5150	13.5754	0.4964	15.3165	0.5962	19.2744	0.8461
	Summer						**Autumn**					
	$q = 0.01$		$q = 0.05$		$q = 0.10$		$q = 0.01$		$q = 0.05$		$q = 0.10$	
Method	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***
ML	4.3919	0.2280	5.0981	0.2064	5.5290	0.1866	4.2899	0.2823	5.2380	0.2382	5.8290	0.2249
MML	4.3855	0.2303	5.1023	0.2080	5.5392	0.1887	4.3066	0.2796	5.2551	0.2366	5.8458	0.2237
PWM	4.5056	0.2830	5.1776	0.2425	5.5934	0.2134	4.1419	0.3472	5.1282	0.2819	5.7450	0.2546
	$q = 0.90$		$q = 0.95$		$q = 0.99$		$q = 0.90$		$q = 0.95$		$q = 0.99$	
Method	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***	$\ddot{X}_q$	***BSD***
ML	10.5956	0.3675	11.7843	0.4359	14.4837	0.6554	12.5904	0.5502	14.2155	0.6879	17.7625	0.9858
MML	10.6788	0.3708	11.8871	0.4380	14.6203	0.6586	12.6113	0.5473	14.2369	0.6842	17.7867	0.9784
PWM	10.4600	0.4253	11.5965	0.5245	14.2128	0.8190	12.7651	0.6023	14.4534	0.7716	18.1356	1.1553

Discussion and conclusions

In this paper, we investigate the performances of different methods for estimating the several specified quantiles of the Gumbel distribution. Robustness of the estimators is also investigated. Their performances are compared via Monte Carlo simulation study with respect to the bias and *MSE* criteria.

Simulation results show that the *PWM* method outperforms the other methods even for small sample sizes with respect to the bias criterion. In terms of the *MSE*, the *ML* method has the best performance for all sample sizes and all values of *q*. The *MSE* values of the *MML* and *ML* estimates, however, are very close especially for $n > 5$.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

In the presence of outliers, the *ML* estimator is found to be robust to the data anomalies (except for models I and III) as expected. Also, all the estimators have substantial bias in almost all cases.

In application, seasonal maximum daily wind speed data taken from Sinop station in Turkey is modelled by using Gumbel distribution based on the *ML*, *MML* and *PWM* estimates. The results of the analyses demonstrate that the fitted densities corresponding to the *ML* and *MML* estimates provide better fit than the fitted densities corresponding to the *PWM* estimate for almost all seasons (except for winter season), see *Table 5* and *Fig. 3*. Also note that *ML* and *MML* estimators provide the best performance based on *BSD* for almost all seasons except for several *q* values of spring as shown in the *Table 6*.

On the other hand, extreme value data generally demonstrate excess kurtosis and/or heavy right tails (Pinheiro and Ferrari, 2016). Gumbel distribution is non-heavy-tailed and characterized by constant skewness and kurtosis, although it is commonly used in modelling environmental data. In this study, it provides quite well modelling in the seasonal maximum daily wind speed data according to the results of *KS* tests, *Q-Q* plots and the histograms and fitted densities superimposed. Additionally, the result of analyses of the real data shows that the *ML* and *MML* estimators provided better results than *PWM* estimator does in both modelling Gumbel distribution to the wind speed data and estimating the lower and upper quantiles of Gumbel distribution for many cases. The *MML* estimators are also numerically very close to the *ML* estimates since they are asymptotically equivalent (Tiku and Akkaya, 2004).

In conclusion, Gumbel distributions based on the *ML* and *MML* estimates can be proposed as an alternative distribution to Gumbel distribution based on the *PWM* estimate because of their superiority on modelling the peak of the wind speed distribution. Moreover, the *ML* and *MML* estimation methods can be recommended to be used in estimating the quantiles of Gumbel distribution for the data due to advantage of having the small *BSD* values.

REFERENCES

[1] Alavi, O., Mohammadi, K., Mostafaeipour, A. (2016): Evaluating the suitability of wind speed probability distribution models: a case of study of east and southeast parts of Iran. – Energy Conversion and Management 119: 101–108.

[2] Aydin, D., Senoglu, B. (2015): Monte Carlo comparison of the parameter estimation methods for the two-parameter Gumbel distribution. – Journal of Modern Applied Statistical Methods 14(2): 123–140.

[3] Barrett, J. P. (1974): The coefficient of determination-some limitations. – The American Statistician 2: 19–20.

[4] Bhattacharyya, G. K. (1985): The asymptotics of maximum likelihood and related estimators based on Type II censored data. – Journal of the American Statistical Association 80: 398–404.

[5] Brano, V. L., Orioli, A., Ciulla, G., Culotta, S. (2011): Quality of wind speed fitting distributions for the urban area of Palermo, Italy. – Renewable Energy 36: 1026–1039.

[6] Ercelebi, S. G., Toros, H. (2009): Extreme value analysis of Istanbul air pollution data. – Clean Soil Air Water 37: 122–131.

[7] Graybeal, D., Leathers, D. (2006): Snowmelt-related flood risk in Appalachia: First estimates from a historical snow climatology. – Journal of Applied Meteorology and Climatology 45(1): 178–193.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

[8] Greenwood, J. A., Landwehr, J. M., Matalas, N. C., Wallis, J. R. (1979): Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. – Water Resources Research 15(6): 1049–1054.
[9] Goel, N. K., Burnb, D. H., Pandeyb, M. D., An, Y. (2004): Wind quantile estimation using a pooled frequency analysis approach. – Journal of Wind Engineering and Industrial Aerodynamics 92: 509–528.
[10] Gumbel, E. J. (1941): The return period of flood flows. – The Annals of Mathematical Statistics 12(2): 163–190.
[11] Hong, H. P., Li, S. H., Mara, T. G. (2013): Performance of the generalized least-squares method for the Gumbel distribution and its application to annual maximum wind speeds. – Journal of Wind Engineering and Industrial Aerodynamics 119: 121–132.
[12] Hosking, J. R. M., Wallis, J. R. (1987): Parameter and quantile estimation for the generalized Pareto distribution. – Technometrics 29: 339–349.
[13] Jöreskog, K. G., Sörbom, D. (1981): LISREL V: Analysis of linear structural relationships by maximum likelihood and least squares methods. – Research Report 81−8, University of Uppsala, Department of Statistics, Uppsala, Sweden.
[14] Kang, D., Ko, K., Huh, J. (2015): Determination of extreme wind values using the Gumbel distribution. – Energy 8: 51–58.
[15] Kantar, Y. M., Usta, I. (2015): Analysis of the upper-truncated Weibull distribution for wind speed. – Energy Conversion and Management 96: 81–88.
[16] Koutsoyiannis, D. (2004): Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. – Hydrological Sciences Journal 49(4): 575–590.
[17] Landwehr, J. M., Matalas, N. C., Wallis, J. R. (1979a): Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. – Water Resources Research 15(6): 1055–1064.
[18] Landwehr, J. M., Matalas, N. C., Wallis J. R. (1979b): Estimation of parameters and quantiles of Wakeby distributions. – Water Resources Research 15: 1361–1379.
[19] Lee, B. H, Ahn, D. J., Kim, H. G., Ha, Y. C. (2012): An estimation of the extreme wind speed using the Korea wind map. – Renewable Energy 42: 4–10.
[20] Martin, D., Zhang, W., Chan, J., Lindley, J. (2014): A comparison of Gumbel and Weibull statistical models to estimate wind speed for wind power generation. – Power Engineering Conference (AUPEC), Australasian Universities, pp. 1–6, Sept. 28 2014–Oct. 1.
[21] Martins, E. S., Stedinger, J. R. (2000): Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. – Water Resources Research 36(3): 737–744.
[22] Modarres, R., Nayak, T. K., Gastwirth, J. L. (2002): Estimation of upper quantiles under model and parameter uncertainty. – Computational Statistics & Data Analysis 39: 529–554.
[23] Nash, J. E., Sutcliffe, J. V. (1970): River flow forecasting through conceptual models, 1, A discussion of principles. – Journal of Hydrology 10: 282–290.
[24] Pinheiro, E. C., Ferrari, S. L. P. (2016): A comparative review of generalizations of the Gumbel extreme value distribution with an application to wind speed data. – Journal of Statistical Computation and Simulation 86: 2241–2261.
[25] Senoglu, B., Tiku, M. L. (2002): Linear contrast in experimental design with nonidentical error distributions. – Biometrical Journal 44: 359–374.
[26] Simiu, E. I., Heckertb, N. A., Filliben, J. J., Johnson, S. K. (2001): Extreme wind load estimates based on Gumbel distribution of dynamic pressures: an assessment. – Structural Safety 23: 221–229.
[27] Tiku, M. L. (1967): Estimating the mean and standard deviation from censored normal samples. – Biometrika 54(1–2): 155–165.
[28] Tiku, M. L. (1968): Estimating the parameters of log-normal distribution from censored samples. – Journal of the American Statistical Association 63: 134–140.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015

[29] Tiku, M. L., Akkaya, A. D. (2004): Robust Estimation and Hypothesis Testing. – New Age International (P) Limited Publishers, New Delhi.
[30] Tiku, M. L., Suresh, R. P. (1992): A new method of estimation for location and scale parameters. – Journal of Statistical Planning and Inference 30: 281–292.
[31] Vaughan D. C., Tiku M. L. (2000): Estimation and hypothesis testing for non-normal bivariate distribution with applications. – Mathematical and Computer Modeling 32: 53–67.
[32] Willmott, C. J. (1982): Some comments on the evaluation of model performance. – American Meteorological Society 63(11): 1309–1313.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1601_001015