Water Resour Manage (2018) 32:3519–3534
https://doi.org/10.1007/s11269-018-2005-6

The Application of Mixture Distribution for the Estimation of Extreme Floods in Controlled Catchment Basins

Wiesław Szulczewski[1] · Wojciech Jakubowski[1]

Received: 19 December 2017 / Accepted: 30 April 2018 /
Published online: 24 May 2018

Abstract In the estimation of distribution of annual maximum flows it is a generally accepted assumption that the sequence of observations originates from a homogeneous population. This assumption, however, is rarely met. The observed annual maximum flow are only in part generated by flood events. The remaining ones are the result of the effect of other hydrological processes that do not have that character. For this reason, a new solution to this problem is proposed in the paper. It is assumed that the sought distribution is a mixture of two probability distributions: a three-parameter GEV distribution, describing flows generated by events with flood character, and a two-parameter gamma distribution, accounting for maximum annual flows that do not have such a character. The paper presents both the method of estimation of the mixture distribution and its application for gauging stations selected so as to take into account possible the most diverse conditions of meteorological, hydrological and geomorphological character. The area with such a high diversification, selected for the study, is the catchment basin of upper and central river Odra (South-West Poland). In the studied water gauge profiles the proposed mixture distribution indicates correct fit. Its advantages and limitations are presented through a comparative analysis with results obtained during estimation of distributions of maximum annual flows by means of standard methods.

Keywords Maximum annual flows estimation · Mixture distribution · GEV distribution · Gamma distribution

✉ Wiesław Szulczewski
wieslaw.szulczewski@upwr.edu.pl

Wojciech Jakubowski
wojciech.jakubowski@upwr.edu.pl

[1] Wrocław University of Environmental and Life Sciences, ul. Norwida 25, 50-375
Wrocław, Poland

Nomenclature

$\Gamma(x,\alpha,\beta)=\int_0^x \frac{1}{\Gamma(\beta)}\alpha^\beta t^{\beta-1}e^{-\alpha t}dt, \quad x>0$ – two-parameter gamma distribution with density function: $\gamma(x,\alpha,\beta)$;

$$\mathrm{GEV}(x,\mu,\sigma,\xi)=\begin{cases}\exp\left[-(1+\xi(x-\mu)/\sigma)^{-\frac{1}{\xi}}\right], \\ 1+\xi(x-\mu)/\sigma>0, \quad \xi\neq 0 \\ \exp\left[-\exp(-(x-\mu)/\sigma)\right], \quad \xi=0\end{cases}$$

three parameter distribution of generalised extreme values with density function: gev(x, μ, σ, ξ); for $\xi=0$ the GEV distribution is transformed into Gumbel distribution (EV1);
$F_{\mathrm{MIX}}(x,\theta)=p\Gamma(x)+(1-p)\mathrm{GEV}(x)$ – mixture of gamma and GEV distributions with density function f_{MIX}; $p\in[0,1]$ – mixture parameter; $\theta=(p,\alpha,\beta\,\mu,\sigma,\xi)$ – vector of distribution parameters;
$L(\theta)$ – likelihood function defined for mixture distribution F_{MIX};
$l(\theta)=\ln L(\theta)$ – log-likelihood function.

1 Introduction

Observed flood flows have for years constituted the basis for the probability estimation of high flood quantiles or exceedance probabilities, values used in the design of hydrotechnical structures, and also in the protection of river-side areas against floods. In the literature one can find numerous studies devoted to that problem. Starting from the nineties of the last century, from moment of publication of the Hosking (1990) paper on probability weighted moments, the number of publications in which distributions of maximum flows were estimated grew rapidly. Their extensive analysis, supported with numerous application examples, is described in papers from the turn of the 20th and 21st centuries. Vogel and Wilson (1996) demonstrated the application of probability weighted moments at nearly 1500 catchment basins in the United States. Katz et al. (2002) used the method of the maximum likelihood to demonstrate that trends in extreme hydrological conditions can be routinely included in analyses of extreme values, with predicted intensification of hydrological cycle within the framework of global climate change. Subsequent studies devoted mainly to non-stationarity of annual maximum flows also include significant introductions describing the applied methods of estimation of maximum flows Xiong et al. (2015).

The methodology of estimation of maximum flows can be divided into two parts (Kidson and Richards 2005):

1. Flood Frequency Analysis (FFA) – only one observation of maximum flow is taken from each hydrological year. On the basis of observations from a multi-year period, assuming their homogeneity, distributions of probability of maximum flows are estimated. Well fitted distributions allow the determination of high quantiles of non-exceedance, treated as n-year waters – flows occurring on average once per n years.
2. Peak Over Threshold (POT) – the method proposed in 1970 by Teodorović and Zelenhasić (1970) (see also Hosking and Wallis 1987 or Mudelsee 2010) requires the determination of all flood events above an adopted threshold. More about threshold level selection can be found in the next manuscripts (Gharib et al. 2017) or (Roth et al. 2015). In this method the result of the analysis can be either the volume of the flood event or the highest flow value in the flood. The POT method requires the determination

of parameters that exclude dependent flood events or those of small volume, insignificant from the viewpoint of maximum flow analysis (Teodorović and Zelenhasić 1970). The determination of quantiles of annual maximum flows requires also the estimation of the distribution of the number of flood events in a year.

A detailed review of the application of the above methods is given in reference (Khaliq et al. 2006).

Both methods require the determination of the class of theoretical distributions that correctly describe the maximum flows. More information on various kinds of uncertainties that appear during the estimation of extreme flow distributions can be found in reference (Yen 2002). In the case of the POT method the choice of distribution is limited to either a generalised extreme value distribution GEV or a generalised Pareto distribution (Wang 1991). Both of those distributions are related with the probabilistic theory of extreme distributions (Coles 2001). In the case of the first of those models, the list of the distributions used is more extensive (Kidson and Richards 2005; Maidment 1993; Cassalho et al. 2018) and includes about 10 distributions of various types, beginning with the shifted gamma distribution (Pearson type III), and ending with GEV type distributions. The choice of the class of distributions depends largely on the local conditions. In the USA, the basic distribution in use is the log-Pearson distribution (Bobee 1975), in Poland the set of allowable distributions includes the following: Pearson type III, log-normal, Gumbel (Fisher-Tippett type I) and GEV (Węglarczyk 2015). The choice of a distribution for further use in e.g. designing a hydrotechnical structure is, unfortunately, a subjective one, and in ambiguous cases additional measures of fit are applied, e.g. the Akaike criterion (Akaike 1974).

In standard procedure the choice of distribution is always preceded by the assumption of sample homogeneity. So far it has been assumed that observed maximum flows originate from a simple sample, and if not – the factor affecting maximum flows is primarily climate change (Xiong et al. 2015; Kundzewicz et al. 2005; Yang and Hill 2012; Cannon 2010). The determination of existing trends allows their filtering and processing the data as homogeneous data. This means that it is assumed that all kinds of trends related with changes in the supply of a water course with precipitation and underground waters, changes in the land use of the catchment basin, long-term fluctuations related with climate change, and in the case of POT method also seasonal changes, have no effect on the observed maximum flows. However, the observed annual maximum flows include distinct freshets generated by various processes occurring in the catchment basin. There appears the problem of genetic heterogeneity. To take it into account in a probabilistic model, it is assumed that the phenomenon of a maximum flow is caused by two different processes. This means that the sought distribution is either a mixture of two probability distributions (Hess et al. 2005) or their product (Frances 1998). The above methodology is recently more and more frequently used in statistical modelling of population distributions in various branches of science, e.g. in economics – (Engle and Lunde 2003), in environmental sciences – (Kollu et al. 2012), or in ecology – (Żyromski et al. 2016) or (Szulczewski et al. 2018). In this study we limited ourselves to the analysis of mixture distributions of the form:

$$F(x) = pF_1(x) + (1 - p)F_2(x) \qquad (1)$$

estimating them for annual maximum flows obtained with the FFA method; F_1, F_2 are component probability distributions, while $p \in [0, 1]$ is a mixture parameter.

In the literature, for the estimation of maximum flows with the use of formula (1) various sets of component distributions are used. However, every time it is assumed that both of them describe events with their full variation. One of the first studies using a mixture distribution was presented by Hawkins (1974). In that study it was assumed that the components have a normal distribution. In subsequent studies the assumption on the form of the distributions was changed. In 2002, Alila and Mtiraoui (2002) assumed that both distributions in the mixture were log-normal. Their fitting to empirical data from Arizona (USA) was compared by fitting to standard distributions (Pearson type III, log-normal, GEV, and five-parameter Wakeby). In 2007, in two publications, Escalante Sandoval estimated distributions of maximum annual flows by a mixture of two Weibull distributions – (Escalante-Sandoval 2007a), and using EV1 and GEV – (Escalante-Sandoval 2007b). The mixture of EV1 and GEV was estimated with the maximum likelihood method, using the Rosenbrock algorithm. In 2009, in a study by Calenda et al. (2009), a mixture of normal distribution and EV1 was used for the estimation of maximum flows in river Tiber. And finally, in a publication of 2017 (Stojković et al. 2017), the authors proposed a set of mixtures of two distributions of the same type with each other. They studied mixtures of GEV, EV1, Pearson type III, and log-Pearson distributions.

The basic problem in the estimation of mixture distribution is the method of estimation of the parameters. In the case of the FFA method the number of observations is limited, while the number of estimated parameters of distribution (1) is doubled, relative to the classic methods. The literature provides examples of the application of various methods, beginning with the method of moments (Hawkins 1974), through the maximum likelihood method, making use of the Rosenbrock algorithm for the estimation of parameter p of the mixture, to Expectation Maximisation algorithm (Vaidyanathan and Vani Lakshmi 2017) in the estimation of distribution (1) as a mixture of two gamma distributions.

In this paper we propose a new method for statistical description of maximum flows with the FFA method. They are modelled with the use of distribution (1), constructed so as to comply additionally with a condition of genetic character. It is postulated that maximum annual flows are the result of operation of two hydrological processes with significantly different, not necessarily separable, intervals of flow variation.

2 Material and Method

The proposed probabilistic model of maximum annual flows has been subjected to verification in catchment basins of upper and central river Odra (Fig. 1). The area is particularly interesting (Migoń 2010). The left-hand tributaries flowing out of the massif of the Sudetes are mountain rivers, and in their lower section – submontane rivers. The right-hand tributaries, on the other hand, are lowland rivers. In addition, above the city of Wrocław, river Odra itself is partially regulated. This causes that in the area of that section of river Odra there is a frequent occurrence of extreme hydrological events that are the result of the effect of variable natural and anthropogenic environment. Some of them take place in the higher parts of the Sudetes, and thus they have a limited range, but other ones affect developed and densely populated regions. The catalogue of extreme events includes various components of the hydrological cycle. The extreme meteorological events include torrential rains, or rainfalls of moderate intensity but long duration. The catchment basin of upper and central river Odra is also known for its rapid disappearance of snow cover in the winter period, caused by the foehn effect.

In consequence there appear extreme hydrological phenomena, and under certain circumstances also geomorphological ones. Freshets on the mountain tributaries of river Odra, caused by precipitations or thaws, are frequent and often cause serious damage in riverside

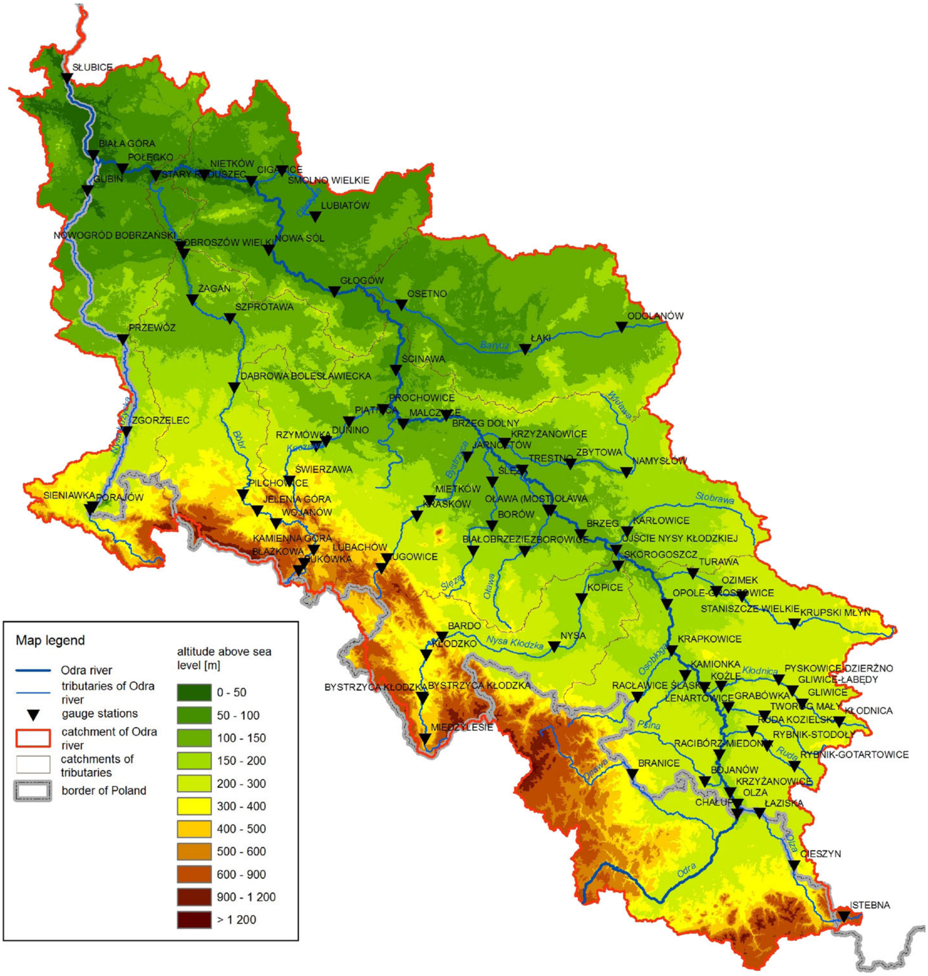

Fig. 1 Catchment basin of upper and central river Odra

localities and areas. Violent floods cause also events of geomorphological character, leading even to remodelling of the relief of the valley bottoms. Rivers of the lowland part of the catchment basin, flowing through agricultural areas, flood mainly in the spring, during the dynamic process of disappearance of snow cover over large areas.

The paper presents nine examples of specially selected rivers and water gauges, chosen so as to represent the diverse conditions of meteorological, hydrological and geomorphological character (Table 1).

The water gauges in Międzylesie, Wilkanów and Kamienna Góra are situated in mountain catchment basins in which in recent years violent floods were noted that were underestimated with the standard methods.

In this study it was assumed that the maximum annual flows obtained with the FFA method (input data for the model) are not homogeneous. It was assumed that their values are the result of two different kinds of hydrological processes. They can be classified as maxima, the value of which is related primarily with the long-term condition of the catchment

Table 1 Characteristics of the selected gauge stations

Water gauge River	Catchment basin	Catchment area [km^2]	Development in the catchment
Kamienna Góra Bóbr	mountain	190	Retention reservoir build in years 1978-1989
Międzylesie Nysa Kłodzka	mountain	50	None
Wilkanów Wilczka	mountain	35	Retention reservoir build in years 1906-1908
Łażany Strzegomka	highland	362	Retention reservoir build in the year 1986
Trestno Odra	lowland	20400	Regulated river; polders above the gauge
Ścinawa Odra	lowland	29600	Flow interfered with by the Wrocław Water Junction
Zbytowa Widawa	lowland	739	Flow restricted by overgrowing with bottom vegetation
Korzeńsko Orla	lowland	1225	None
Kłodnica Kłodnica	urban	770	Crossing the entire Silesian agglomeration

basin. Such maximum flows usually fall within the ranges of medium water levels and do not have a flood character. The second class are actual flood flows, often related with water overflowing the river bed. The two classes are not separable, and in a majority of the water gauge transects it is difficult to indicate a threshold that would separate them univocally. In relation to the above, we sought a mixture distribution that would be universal enough to take into account two kinds of maximum annual flows, and flexible enough to comprise possibly the broadest spectrum of distributions used so far. A detailed analysis concerning the distributions of the mixture components was conducted. Due to the character of the analysed catchment basin it was assumed that the GEV distribution is one of the components, responsible for high flows. The second component of mixture (1) that was subjected to analysis included the following two-parameters distributions: normal, log-normal, gamma, and the Weibull distribution. The compatibility tests conducted demonstrated that only a mixture of gamma and GEV distributions has the best statistical fit. In the cases under analysis, 100% fit effectiveness was obtained for the mixture of gamma and GEV distributions, while the other mixtures did not exceed the level of 50%. In such a case, distribution (1) can be reduced to the form:

$$F_{\mathrm{MIX}}(x) = p\Gamma(x, \alpha, \beta) + (1 - p)\mathrm{GEV}(x, \mu, \sigma, \xi) \tag{2}$$

The proposed distribution (2), further on referred to as MIX, requires the estimation of 6 unknown parameters.

As it is easy to notice, the analysed distribution (2) determines a certain family of distributions. Depending on the values of the estimated parameters, it can be reduced to a two-parameter gamma distribution, a three-parameter GEV, or a two-parameter Gumbel distribution (Fig. 2). With the engineering postulate of choosing one type of distribution for the whole country (Węglarczyk 2015), the MIX distribution (2) appears to be a good proposal. Its flexibility, presented in Fig. 2, allows the estimation of distributions of maximum

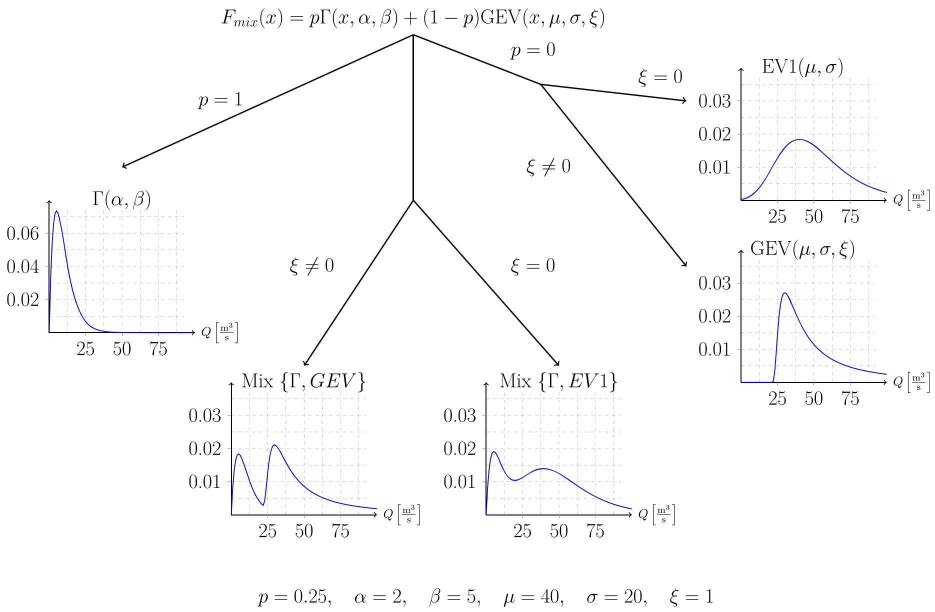
$F_{mix}(x) = p\Gamma(x,\alpha,\beta) + (1-p)\mathrm{GEV}(x,\mu,\sigma,\xi)$
$p = 1$
$p = 0$
$\xi = 0$
$\xi \neq 0$
$\xi \neq 0$
$\xi = 0$
$\Gamma(\alpha,\beta)$
EV1(μ,σ)
GEV(μ,σ,ξ)
Mix $\{\Gamma, GEV\}$
Mix $\{\Gamma, EV1\}$
0.06
0.04
0.02
0.03
0.01
25
50
75
$Q\left[\frac{m^3}{s}\right]$
$p = 0.25,\quad \alpha = 2,\quad \beta = 5,\quad \mu = 40,\quad \sigma = 20,\quad \xi = 1$

Fig. 2 Scheme of possible distributions realized by (2)

annual flows in mountain, submontane and lowland rivers, characterised by a high diversity of kinds of flows.

The estimation of the unknown parameters of distribution (2) was conducted with the maximum likelihood method. In this case, the likelihood function

$$L(\theta) = L(p, \alpha, \beta, \mu, \sigma, \xi) = \prod_{i=1}^{n} f_{\mathrm{MIX}}(x_i) \tag{3}$$

depends on six parameters, where n is the number of observations, and $x_i,\ i = 1, \ldots, n$ the observed maximum annual flows.

The global maximum of the likelihood function was determined using the genetic algorithm of searching for the global extreme of multivariate functions developed by Price et al. (2005). The very definition of distribution (2) indicates that the likelihood function can be multimodal, which was also demonstrated by the calculations performed. Figure 3 presents a two-dimensional graph obtained in the estimation of the distribution of maximum flows at water gauge Międzylesie (Nysa Kłodzka river).

The behaviour of the estimated mixture distribution (2) on the example of water gauge Trestno (Odra river) is presented in Fig. 4. Part (a) presents graphs of the density functions of the mixture distribution and its components. One can easily note the range of maximum flows in which the population is estimated by both components of the mixture. Part (b) presents the distribution functions of the mixture, its components, and the fit to observations.

The same example of water gauge Trestno was used to present the transition zone between flows generated by medium flows – estimated by a gamma distribution, and flood flows, estimated by a GEV distribution (Fig. 5). That interval (approximately in the range from 950 to 1100 m^3) is situated above the zone of medium states.

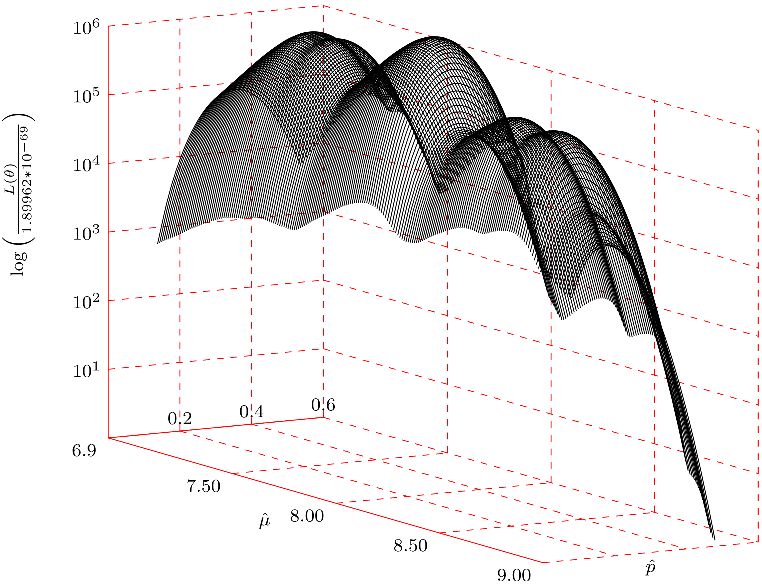

Fig. 3 Likelihood function (3) plotted with the assumption of constant values of estimators $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$, $\hat{\xi}$; water gauge Międzylesie, Nysa Kłodzka river

With regard to hydrological applications, it is interesting to determine the intervals of confidence for quantiles of exceedance. They were determined with the use of the delta method (Coles 2001). In accordance with that method, the estimator $\hat{x}_q$ of the quantile of maximum annual flow has normal distribution with a mean x_q and variance V. In the formal notation, for probability π we obtain:

$$\hat{x_q} = F_{\text{MIX}}^{-1}(\pi) \sim N(x_q, V). \tag{4}$$

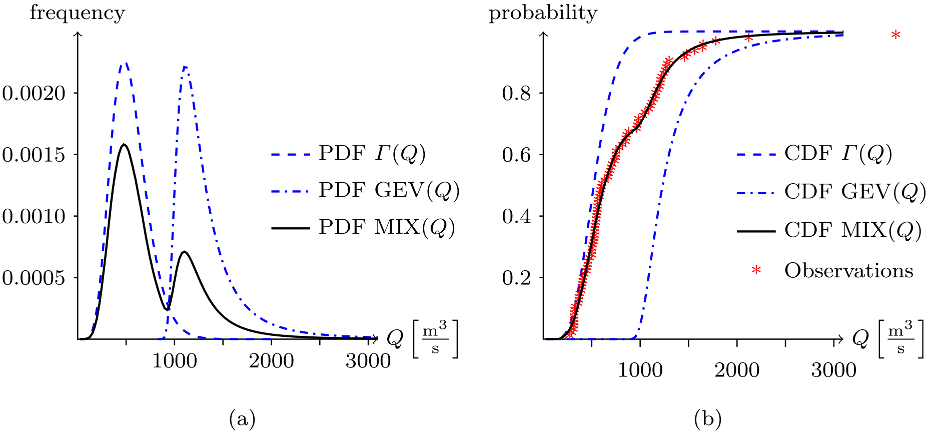

Fig. 4 Estimated distribution of the MIX (2) of maximum flows, river Odra, water gauge Trestno

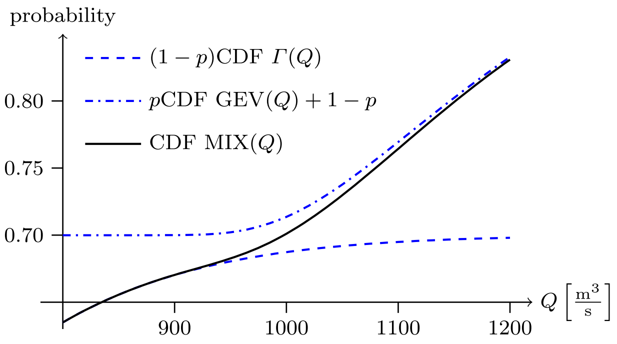

Fig. 5 Analysis of the influence of component cumulative distributions on the resulting cumulative distribution of the MIX (2), river Odra, water gauge Trestno

In this method the variance V in the formula (4) is calculated as follows:

$$V = \nabla x_q^T W_\theta \nabla x_q, \qquad \nabla x_q = \left[\frac{\partial x_q}{\partial p}, \ldots, \frac{\partial x_q}{\partial \xi}\right]^T, \tag{5}$$

where the matrix W_θ is approximated by a matrix reverse to the expected information matrix:

$$I_O(\theta) = \begin{bmatrix} -\frac{\partial^2}{\partial p^2} l(\theta) & \ldots & -\frac{\partial^2}{\partial p \partial \xi} l(\theta) \\ \vdots & & \vdots \\ -\frac{\partial^2}{\partial p \partial \xi} l(\theta) & \ldots & -\frac{\partial^2}{\partial \xi^2} l(\theta) \end{bmatrix} \tag{6}$$

calculated for the estimated values of the parameters.

In the applications presented below, the values of the information matrix (6) were determined numerically, while the derivatives in formula (5) were calculated on the basis for formulae presented in the attached Appendix. The calculations were performed using the program Mathematica 9, Wolfram Research Inc., and our own original software.

3 Results and Analysis

In the nine water gauges described above, the parameters of standard distributions (log-normal, Pearson type III and GEV) were estimated with the method of the maximum likelihood, and compared with the distribution of the MIX, also estimated with the method of the maximum likelihood. The results of the estimation – adequate p-values of the χ^2 goodness-of-fit test and corresponding values of mean absolute relative error (MARE) given in %, are presented in Table 2. The MARE is determined between observed flows exceeding the value of the median and their equivalents calculated from the estimated MIX distribution. The measure of model fit error defined in this manner estimates quantitatively high flows that are of the highest importance in engineering practice. For maximum annual flows observed on a given water gauge the goodness-of-fit χ^2 tests were conducted for a predetermined and constant division into classes, irrespective of the tested distribution. The p-test values given in Table 2 are obviously dependent on the number of estimated parameters. In the case of the MIX distribution, it is much more difficult to work with the doubled number of parameters in trying to fit the mixture distribution.

Table 2 Goodness-of-fit of the χ^2 tests and mean absolute relative error (MARE) of the estimated distributions with maximum flows at selected water gauge stations

Water gauge	Catchment	n		Pearson III	log-normal	GEV	MIX
River	basin						
Kamienna Góra	mountain	40	p-value	0.96	0.96	0.91^{**}	0.33
Bóbr			MARE	7.37	6.76	8.32	**2.62**
Międzylesie Nysa	mountain	54	p-value	**0.005**	0.21	0.42	0.22^{***}
Kłodzka			MARE	24.41	11.85	8.77	**7.68**
Wilkanów Wilczka	mountain	50	p-value	**0.005**	0.14	0.19^{**}	0.15^{**}
			MARE	34.09	12.33	7.59	**6.22**
Łażany Strzegomka	highland	60	p-value	0.19	0.12	0.05^{**}	0.35^{*}
			MARE	13.62	19.88	28.86	**4.72**
Trestno Odra	lowland	94	p-value	0.06	**0.026**	**0.005**	0.13
			MARE	5.37	6.05	8.41	**3.16**
Ścinawa Odra	lowland	95	p-value	0.35	0.60	0.54	0.24
			MARE	4.99	3.29	4.53	**3.20**
Zbytowa Widawa	lowland	44	p-value	0.30	0.12	0.26^{*}	0.19
			MARE	6.18	10.04	7.02	**2.60**
Korzeńsko Orla	lowland	60	p-value	**0.016**	**0.005**	–	0.38^{*}
			MARE	6.62	12.81	–	**2.89**
Kłodnica Kłodnica	urban	55	p-value	0.48	0.30	0.18^{**}	0.24
			MARE	6.90	9.88	15.00	**6.03**

* – $\hat{\xi} = 0$ (Gumbel distribution), ** – $\hat{\xi} > 0.5$, *** – $\hat{\xi} > 1.0$;

n – number of years of observation;

red color – the goodness-of-fit hypothesis is rejected;

blue color – minimum value of MARE criterion

In addition, also indicated in Table 2 are those water gauges for which the estimated parameter $\hat{\xi}$ of the GEV and MIX distributions equalled 0, was higher than 0.5, or higher than 1.0. This is significant in view of the existence of GEV distribution moments. In the case when $0.5 \leq \hat{\xi} < 1.0$ is it not possible to determine its variance, while when $\hat{\xi} \geq 1.0$ it is also impossible to determine its mean value.

Analysis of the results obtained allows to draw the conclusion that the application of the MIX distribution improved the goodness of fit. Contrary to the standard distributions, at none of the analysed water gauges there were grounds to reject the hypothesis of good fitting of annual maximum flows with the mixture distribution. This is firmly supported by the MARE measure of fit, an index which is very close to the engineering intuition. Regarding the goodness of fit of the applied distributions, the water gauges presented in Table 2 can be classified in three groups.

I One of them groups those gauging stations at which the standard methods work correctly. In this group the level of p-values of χ^2 test, calculated for the standard distributions, did not provide justification for rejecting the goodness-of-fit hypothesis. In such a situation it was up to the decision maker to choose which distribution to use in engineering practice. In the examples presented in Table 2, this group includes the water gauges at Kamienna Góra, Ścinawa and Kłodnica, and those at Łażany and Zbytowa.

II The second one groups two water gauges (Międzylesie and Wilkanów), for which the hypothesis of good fitting of maximum flows with the Pearson type III distribution should be rejected. The tests of conformance of the remaining standard distributions, as well as the MIX distribution, did not provide grounds for the rejection of the conformance hypothesis. In this group the value of the estimator of parameter $\hat{\xi}$ of the MIX distribution was alarming. In both cases it significantly exceeded the value of 1.0. This means a lack of the expected value in the GEV distribution, and thus serious difficulties in the assessment of variability of the estimated quantiles.

III The third group, i.e. those water gauges where all the standard methods indicated the necessity of rejecting the tested hypotheses. In the selected set of water gauges, the third group includes two highly different lowland catchment basins: Trestno, Odra river, and Korzeńsko, Orla river. Only the MIX distribution provides a correct fit.

The most important application of the proposed probabilistic model is the inference from analysis of quantiles determined for the purposes of design of hydrotechnical structures, flood protection etc. Figure 6 presents graphs of probability distributions from four water gauges included in Table 2. They provide a comparison of the best fitting standard distribution and the MIX distribution with appended one-sided area of confidence determined by a quantile of the order of 0.84 (standard value in engineering applications). In the case of water gauge Trestno, due to the lack of fit none of the standard distributions was plotted. The horizontal axis in Fig. 6 was specially formatted in the scale of probability so as to facilitate the interpretation of the results for engineering applications. In Fig. 6a – water gauge Łażany, on the graph of the boundary of the area of confidence determined by the MIX distribution one can observe its local maximum. It is related to the area of determinacy of the GEV distribution when the estimated value of its lower limit falls within the range of variation of low maximum annual flows.

Frequently considered in engineering practice are qualtiles of flow corresponding to the probability of occurrence of 1% of maximum flow, together with its interval of confidence. Those values imply the costs of various kinds of hydrotechnical, flow protection and similar investment (see e.g. Mogollon et al. 2016). Therefore, it is important that they are estimated correctly, so that the estimate is in conformance with engineering intuition and practical knowledge. In this context, Table 3 compiles results obtained with the use of the standard models and the MIX for nine test water gauges. For each of those cases one column presents the 1% quantiles of maximum flow, while another column shows their 84% limits of intervals of confidence. In the case of the standard methods, according to the results presented in Table 3, if there were no grounds for the rejection of the goodness-of-fit hypothesis, a maximum of two best-fitting distributions were selected. In the case of a lack of possibility of estimating a quantile, a dash sign was used. In addition, the last column of the Table 3 presents the highest observed values in the analysed set of maximum annual flows. Those values are given as a kind of reference to the one-hundred-year flood obtained from estimation, and in consequence – to the possibility of conducting a qualitative analysis of fit that is obtained using the standard methods and the proposed MIX distribution. Comparison of

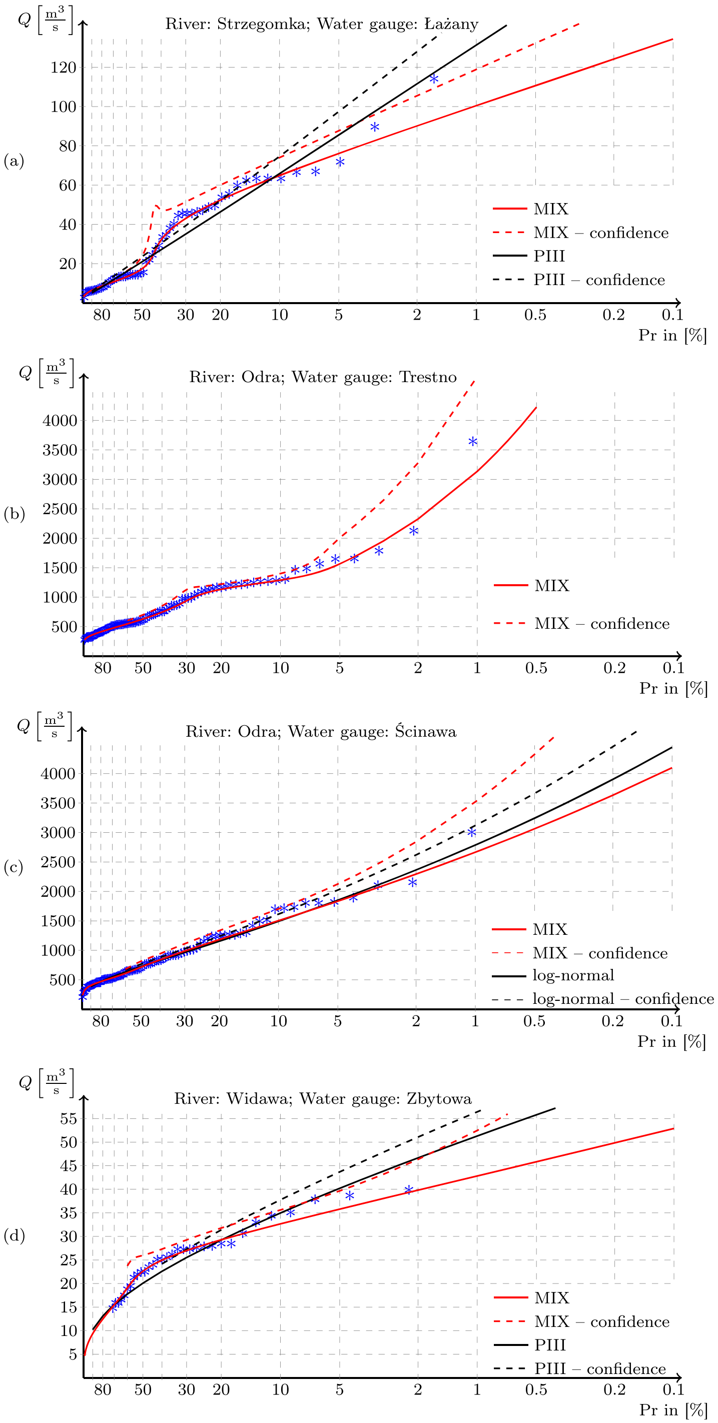

Fig. 6 Distribution curves and confidence areas of the non-exceedance quantile of the best-fitting standard distributions and the MIX distribution for chosen water gauges: **a**-Łażany, **b**-Trestno, **c**-Ścinawa, **d**-Zbytowa

Table 3 One-percent quantiles and their 84% confidence areas of maximum flow at selected water gauges

Water gauge	Standard method		MIX		Q^*[m^3/s]
River	$Q_{1\%}$[m^3/s]	$Q_{1\%,c}$[m^3/s]	$Q_{1\%}$[m^3/s]	$Q_{1\%,c}$[m^3/s]	
Kamienna Góra Bóbr	118/155 PIII/LN	155/198 PIII/LN	136	284	109
Międzylesie Nysa Kłodzka	32/36 LN/GEV	38/44 LN/GEV	58	–	63.1
Wilkanów Wilczka	88/98 LN/GEV	116/– LN/GEV	160	–	105
Łażany Strzegomka	138/292 PIII/LN	169/415 PIII/LN	100	119	114
Trestno Odra	2455 PIII	2701 PIII	3136	4744	3640
Ścinawa Odra	2395/2787 PIII/LN	2598/3121 PIII/LN	2666	3521	3000
Zbytowa Widawa	51.3/55.3 PIII/GEV	56.5/60.3 PIII/GEV	42.8	52.5	39.8
Korzeńsko Orla	–	–	71	81	69.1
Kłodnica Kłodnica	66/194 PIII/GEV	77/247 PIII/GEV	82	147	68.3

* the highest observed annual flow value in the data set

those values in the particular cases allows to note that the standard methods often cause a large underestimation or a far-reaching extrapolation. Each of those cases is undesirable in practice. The first may cause e.g. the building of flood levees that will not perform their role, while the second – an absurd increase of the costs of the investment. Apart from a few exceptions, the basic conclusion that follows from Table 3 is that the behaviour of the MIX is characterised by a high stability relative to the results obtained by means of the 3 classic three-parameter distributions. This is supported by the MARE index which is closer to engineering intuition (Table 1).

However, in the case of the MIX a big problem is the occasional load with the heavy tail of the GEV distribution. In this case there is theoretically no possibility of determination of the areas of confidence for maximum flows with a high probability of non-exceedance.

Unfortunately, an important aspect in the use of the standard methods is the necessity of subjective selection of a statistically substantiated distribution for the estimation of values that determine investment decisions. This is not an easy situation for the decision-makers. Table 3 presents two examples of such water gauges: Łażany and Kłodnica, where depending on the choice of distribution the estimations of quantiles of one-hundred-year flood and their areas of confidence differ at least two-fold. The model of the MIX distribution does not permit such extremely different estimations and, what is more, within the range in which it has been tested, it is much more effective in operation.

4 Conclusions

The MIX distribution (2) proposed in the paper has been analysed against the background of distributions used so far. The comparison was made on the base of observations of maximum annual flows in specially selected water gauges in the upper and central river Odra. They are characterised by a large diversity of natural and anthropogenic conditions. Based on the results obtained we can formulate the following conclusions:

1. The MIX distribution, even in hydrologically difficult cases, where standard methods fail, provides statistically correct estimates, although there are situations in which it is not possible to determine the area of confidence of the tested quantile.
2. In the case of water gauges where the standard distributions give correct estimates, the MIX distribution provides similar values.
3. The MIX distribution takes into account the genetic heterogeneity of observed maximum annual flows – it can be the basic tool in the work of the hydrologist.
4. The MIX model allows to unify the statistical procedure and inference.
5. As opposed to the standard methods, in which statistically justified estimations of high quantiles of non-exceedance are characterised by unacceptably high variation, the MIX model displays a more rational behaviour.

In spite of having conducted an analysis of the effectiveness of the proposed method on several dozen water gauges on Polish and European rivers, the authors realize that the method still needs further testing and verification. The applied genetic algorithm of estimation of the parameters of the MIX, employing the search for the global maximum of a multivariate function, does not ensure the obtainment of the maximum value of the target function. As in this algorithm the initial population is selected at random, the chances of obtaining erroneous values of the estimated parameters were reduced by multiple repetition of the calculation process.

Compliance with Ethical Standards

Conflict of interests The author declare that they have no conflict of interest.

Appendix

The partial derivatives occurring in model delta (Coles 2001) – formula (5), have been calculated by the Leibniz's rule for differentiation under the integral sign. Their values are as follows:

1. Parameter p:

$$\frac{\partial x_q}{\partial p} = \frac{\mathrm{GEV}(x_q, \mu, \sigma, \xi) - \Gamma(x_q, \alpha, \beta)}{p\gamma(x_q, \alpha, \beta) + (1-p)\mathrm{gev}(x_q, \mu, \sigma, \xi)} = \frac{\frac{1}{p}\left[\mathrm{GEV}(x_q, \mu, \sigma, \xi) - \pi\right]}{p\gamma(x_q, \alpha, \beta) + (1-p)\mathrm{gev}(x_q, \mu, \sigma, \xi)}$$

2. Parameter α:

$$\frac{\partial x_q}{\partial \alpha} = \frac{p\frac{\beta}{\alpha}\left[\Gamma(x_p, \alpha, \beta+1) - \Gamma(x_p, \alpha, \beta)\right]}{p\gamma(x_q, \alpha, \beta) + (1-p)\mathrm{gev}(x_q, \mu, \sigma, \xi)}$$

3. Parameter β:

$$\frac{\partial x_q}{\partial \beta} = \frac{p\,(\psi(\beta) - \ln(\alpha))\,\Gamma(x_q, \alpha, \beta) - p \int\limits_{0}^{x_q} \gamma(x, \alpha, \beta) \ln x dx}{p\gamma(x_q, \alpha, \beta) + (1 - p)\mathrm{gev}(x_q, \mu, \sigma, \xi)}$$

where $\psi(\beta) = \frac{\Gamma'(\beta)}{\Gamma(\beta)}$ is a digamma function – a logarithmic derivative of function Γ.

4. Parameter μ:

$$\frac{\partial x_q}{\partial \mu} = \frac{(1 - p)\mathrm{gev}(x_q, \mu, \sigma, \xi)}{p\gamma(x_q, \alpha, \beta) + (1 - p)\mathrm{gev}(x_q, \mu, \sigma, \xi)}$$

5. Parameter σ:

$$\frac{\partial x_q}{\partial \sigma} = \frac{(1 - p)\mathrm{gev}(x_q, \mu, \sigma, \xi)\frac{x_q - \mu}{\sigma}}{p\gamma(x_q, \alpha, \beta) + (1 - p)\mathrm{gev}(x_q, \mu, \sigma, \xi)}$$

6. Parameter ξ:

$$\frac{\partial x_q}{\partial \xi} = \frac{(1 - p)\mathrm{gev}(x_q, \mu, \sigma, \xi)\left[\frac{\sigma}{\xi^2} \ln\left(1 + \frac{\xi}{\sigma}(x_q - \mu)\right)(1 + \frac{\xi}{\sigma}(x_q - \mu)) - \frac{x_q - \mu}{\xi}\right]}{p\gamma(x_q, \alpha, \beta) + (1 - p)\mathrm{gev}(x_q, \mu, \sigma, \xi)}$$

References

Akaike H (1974) A new look at statistical model identification. IEEE Trans Automat Contr 16:716–722

Alila Y, Mtiraoui A (2002) Implications of heterogeneous flood-frequency distributions on traditional stream-discharge prediction techniques. Hydrol Process 16:1065–1084

Bobee B (1975) The log Pearson type 3 distribution and its application in hydrology. Water Resour Res 11(5):681–689

Calenda G, Mancini C, Volpi E (2009) Selection of the probabilistic model of extreme floods: the case of the river tiber in rome. J Hydrol 371:1–11

Cannon AJ (2010) A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology. Hydrol Process 24:673–685

Cassalho F, Beskow S, de Mello C, de Moura M (2018) At-site flood frequency analysis coupled with multiparameter probability distributions. Water Resour Manag 32(1):285–300

Coles S (2001) An introduction to statistical modeling of extreme values. Springer, Berlin

Engle R, Lunde A (2003) Trades and quotes: a bivariate point process. J Financ Economet 1(2):159–188

Escalante-Sandoval C (2007a) Application of bivariate extreme value distribution to flood frequency analysis: a case study of Northwestern Mexico. Nat Hazards 42:37–46

Escalante-Sandoval C (2007b) A mixed distribution with EV1 and GEV components for analyzing heterogeneous samples. Ingeniería Investigación y Tecnología 8(3):123–133

Frances F (1998) Using the TCEV distribution function with systematic and non-systematic data in a regional flood frequency analysis. Stoch Hydrol Hydraul 12:267–283

Gharib A, Davies E, Goss G, Faramarzi M (2017) Assessment of the combined effects of threshold selection and parameter estimation of generalized pareto distribution with applications to flood frequency analysis. Water 9:1–17

Hawkins R (1974) A note on mixed distributions in hydrology. In: Proceedings of a Symposium on Statistical Hydrology. U.S. Department of Agriculture, Agricultural Research Service; Washington, D.C., vol 1275, pp 336-335

Hess S, Bierlaire M, Polak J (2005) Capturing correlation and taste heterogeneity with mixed GEV models. Applications of simulation methods in environmental and resource economics. The Economics of Non-Market Goods and Resources 6:55–75

Hosking J (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. J R Stat Soc. Series B Stat Methodol 52:105–124

Hosking J, Wallis J (1987) Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29:339–349

Katz R, Parlange M, Naveau P (2002) Statistics of extremes in hydrology. Adv Water Resour 25:1287–1304

Khaliq M, Ouarda T, Ondo JC, Gachon P, Bobe'e B (2006) Frequency analysis of a sequence of dependent and or non-stationary hydrometeorological observations: a review. J Hydrol 329:534–552

Kidson R, Richards K (2005) Flood frequency analysis: assumptions and alternatives. Prog Phys Geogr 29(3):392–410

Kollu R, Rayapudi S, Narasimham S, Pakkurthi K (2012) Mixture probability distribution functions to model wind speed distributions. IJEEE 3:1–10

Kundzewicz Z, Graczyk D, Maurer T, Pińskwar I, Radziejewski M, Svensson C, Szwed M (2005) Trend detection in river flow series: 1. A,nnual maximum flow. Hydrolog Sci J 50(5):797–810

Maidment D (1993) Handbook of hydrology. McGraw-Hill, New York

Migoń P(ed) (2010) Wyjątkowe zdarzenia przyrodnicze na Dolnym Śląsku i ich skutki, vol 14. Instytut Geografii i Rozwoju Regionalnego Uniwersytetu Wrocawskiego

Mogollon B, Frimpong E, Hoegh A, Angermeier P (2016) An empirical assessment of which inland floods can be managed. J Environ Managment 167:38–48

Mudelsee M (2010) Climate time series analysis. Classical statistical and bootstrap methods. Springer, Berlin

Price KV, Storn RM, Lampinen JA (2005) Differential evolution - a practical approach to global optimization. Springer, Berlin

Roth M, Jongbloed G, Buishand T (2015) Threshold selection for regional peaks-over-threshold data. J Appl Stat 43(7):1291–1309

Stojković M, Prohaska S, Zlatanović N (2017) Estimation of flood frequencies from data sets with outliers using mixed distribution functions. J Appl Stat 44(11):2017–2035

Szulczewski W, Żyromski A, Jakubowski W, Biniak-Pieróg M (2018) A new method for the estimation of biomass yield of giant miscanthus (*Miscanthus giganteus*) in the course of vegetation. Renew Sustain Energy Rev 82(2):1787–1795

Teodorović P, Zelenhasić V (1970) A stochastic model for flood analysis. Water Resour Res 6(6):1641–1648

Vaidyanathan V, Vani Lakshmi R (2017) Estimation of parameters in a finite mixture of multivariate gamma distributions using gaussian approximation. Sri Lankan Journal of Applied Statistics 17-3:187–200

Vogel RM, Wilson I (1996) Probability distribution od annual maximum, mean, and minimum streamflow in the United States. J Hydrol Eng 69:69–76

Wang Q (1991) The POT model described by the generalized Pareto distribution with Poisson arrival rate. J Hydrol 129:263–280

Węglarczyk S (2015) Osiem powodów konieczności rewizji stosowanych w Polsce wzorów na maksymalne roczne przepywy o zadanym prawdopodobieństwie przewyższenia. Gospodarka Wodna 11:323–328

Xiong L, Du T, Xu C, Guo S, Jiang C, Gippel C (2015) Non-stationary annual maximum flood frequency analysis using the norming constants method to consider non-stationarity in the annual daily flow. Water Resour Manag 29:3615–3633

Yang C, Hill D (2012) Modeling stream flow extremes under non-time-stationary conditions. In: XIX international conference on water resources, University of Illinois at Urbana-Champaign June 17-22

Yen B (2002) System and component uncertainties in water resources. In: Bogardi J, Kundzewicz Z (eds) Risk, reliability, uncertainty and robustness of water resources systems. Cambridge University Press, Cambridge, pp 133–142

Żyromski A, Szulczewski W, Biniak-Pieróg M, Jakubowski W (2016) The estimation of basket willow (*Salix viminalis*) yield – New approach. Part I: Background and statistical description. Renew Sust Energ Rev 65:1118–1126