An analysis of the hydrological regime as a factor influencing on the distributions of maximum annual flows

Wiesław Szulczewski¹,², Wojciech Jakubowski¹ and Tamara Tokarczyk²

¹Wrocław University of Environmental and Live Sciences, Department of Mathematics, ul. Norwida 25, 50-375 Wrocław, Poland
²Institute of Meteorology and Water Management National Research Institute, Branch Wrocław, ul. Parkowa 30, 51-616 Wrocław, Poland

Abstract. Statistical models of freshet flows are the basis for the design of hydrotechnical structures and for undertaking all and any activities related with flood threat. With regard to the method of data preparation for estimation and to the estimation procedure itself, the methods applied in such situations can be divided into two parts - FFA (Flood Frequency Analysis) and POT (Peak Over Threshold). In this study a comparison of those methods is made, using an original mixture of distributions (FFA) and an original procedure of distribution estimation (POT) for six selected water gauges on the river Odra.

1 Introduction

For years, statistical elaboration of freshet flows has been the basis for the estimation of high quantities of transgression. Those values are necessary at the stage of design of hydrotechnical structures, identification of areas with elevated risk of flood threat, etc. Therefore, starting with the nineties of the 20th century that subject is constantly in the area of interest of science. There is no way to analyse all the studies concerned with this problem. However, there are a few studies that partially fulfill the task, e.g. [1, 2]. For some time now more and more studies are devoted to the problem of non-stationarity of freshet flows. Those studies include also review elements covering the methods of estimation of maximum flows [3]. The methods used in the case of estimation of maximum flows can be divided into two parts [4], in relation to the method of preparation of data for the estimation, and to the estimation procedure itself.

In the first method, called FFA (Flood Frequency Analysis), only one observation – of the maximum flow – is taken from each hydrological year. On the basis of multi-year observations obtained in this manner, probability distributions of maximum flows are estimated. In this case the list of distributions used is extensive [4, 5] and contains over a dozen distributions of various types, starting with the three-parameter gamma distribution (Pearson type III), and ending with GEV type distributions. The choice of distributions applied in individual countries is highly diverse. For instance, in the USA the basic distribution is the log-Pearson distribution [6], in Poland the set of allowable distributions includes such distributions as Pearson type III, log-normal, Gumbel, and GEV [7]. Applying the above unimodal distributions it is assumed that the observed maximum flows originate from a simple sample. This assumption is highly debatable, as there are dry years, with no significant high flows, in which it is hard to classify the maximum flow as a freshet flow. To solve that problem of genetic heterogeneity, in the study [8] it was assumed that the sought distribution is a mixture of two probability distributions. The first is responsible for non-freshet maxima, and the second for freshet ones. This type of approach is now fairly often applied in similar problems related to e.g. economics [9] or ecology – e.g. [10, 11]. The MIX distribution proposed in the paper by [8] will be the basis for the comparative analyses presented herein. It is defined as follows:

\[ \text{MIX}(x) = p \Gamma(x; a, \nu) + (1 - p) \text{GEV}(x; \mu, \sigma, \xi) \]  

(1)

where:

\[ \Gamma(x; a, \nu) = \frac{1}{\Gamma(\nu)} x^{-a} \left( 1 - x^{-1} \right)^{-1/\nu}, \]

\[ \text{GEV}(x; \mu, \sigma, \xi) = \begin{cases} 
\exp \left[ - \left( 1 + \frac{(x - \mu)}{\sigma} \right)^{-1/\xi} \right], & \text{if } \xi > 0, \\
\exp \left[ - \left( \frac{(x - \mu)}{\sigma} \right) \right], & \text{if } \xi = 0
\end{cases} \]

is a three-parameter distribution of generalised extreme values;

\[ p, p \in [0,1] \] is a parameter of the mixture.

In the second method, called POT (Peak Over Threshold), proposed by Teodorović and Zelenhaisić [12] in 1970, all flood events are determined that exceed an arbitrarily chosen cut-off threshold in the analysed multi-year period. In consequence, the result of elaboration of such data can be either the volume of the high-flow event, its duration, or the highest flow in the event. In the context of comparison of the two methods, only one parameter of a high-flow event was adopted for analysis – its maximum flow. To determine the maximum annual flow it is also necessary to analyse the number of high-flow events in a year. Hence, the estimation of the
distribution of maximum annual flows is a two-stage procedure. In the first step we determine the distribution of maximum flows is all high-flow events of the analysed multi-year period (Fig.1). The high-flow event is adopted as a sequence of days with flows above the threshold \( Q_{95\%} \). In the study it was assumed that it corresponds to the flow \( Q_{70\%} \). The conducted goodness-of-fit tests showed that the best fit is displayed by the two-parameter generalised Pareto distribution (GPD):

\[
GPD(x) = 1 - (1 - \alpha x \lambda)^k, \quad x > 0
\]

In the second step the number of high flows in a year is determined. As a standard it is assumed that their distribution is a Poisson one. Then the distribution of maximum annual flows \( Q \) is determined as a compound Poisson distribution (CP) in accordance with the formula:

\[
Pr(Q \leq x + b) = \exp[-\lambda(1 - GPD(x))]
\]

where:

- \( x + b \geq 0 \) is the value of the maximum annual flow;
- \( b > Q_{95\%} \) is the threshold and \( x \geq 0 \) is the flow above the threshold;
- \( \lambda \) is the mean number of high-flow events in the year.

Formally it is a distribution with three parameters \( \alpha, \lambda \) of GPD distribution and \( \lambda \) of Poisson distribution. However, the GPD distribution is determined for flows exceeding a certain additional threshold \( b \) – it constitutes the fourth, hard to estimate, parameter of the distribution.

The paper presents the results of estimation of the distribution of maximum flows at selected water gauges on the river Odra, using the MIX distribution for the FFA method, and for the POT method – the compound Poisson distribution CP described above.

## 2 Material and method

### 2.1 River Odra and its catchment basin

The river Odra is one of the longest rivers in the catchment basin of the Baltic Sea. It is the second longest river in Poland, after the Vistula. Its length is 854 km, out of which 742 km within the territory of Poland. The total area of the catchment basin of the river Odra is 118861 km². Its major part, approx. 89%, is within the territory of Poland, 5% in Germany, and 6% in the territory of the Czech Republic.

Over the entire length of the Odra river one can distinguish two parts: the mountain part, covering the initial 50 km of its run, and the lowland part – on the remaining length of the river. The Odra river is canalised over 186 km, with 24 weirs damming the water.

The landscape of the Odra Valley within the territory of Poland is highly diversified (Tab.1). The Valley cuts across areas with diverse geological structure and relief, formed by various factors. Also the age of the particular sections is varied: the upper section, from the border of Poland to the Ścinawskie Depression, was formed ultimately after the recession of the ice-sheer of the Odra glaciation (300-280 thousand years ago), while the section from Siekierki to the mouth of the Odra into the Szczecin Lagoon is only 13-15 thousand years old.

The Valley of the river Odra in included in nine physiographic units with the rank of mesoregions, that form seven units with the rank of macroregions. The On the basis of Kondracki’s [13] physiographic division of Poland, the Odra Valley is divided into 9 sections:

| SECTION | MESOREGION | MACROREGION | SUBPROVINCE |
|---------|------------|-------------|-------------|
| National Border – Krapkowice | Raciborska Basin | Silesian Lowland | Central Poland Lowlands |
| Krapkowice – Łuża | Wroclaw Ustronreal | Trzebnicki Ridge |
| Łuża – Chobienia | Ścinawskie Depression | Millicko-Głogowskie Depression |
| Chobienia – Stany n. Nowa Sól | Głośów Ustronreal | South-Baltic Lakeland |
| Stany n. Nowa Sól – Cigacice | Kargowska Basin | Warta-Odra Ustronreal |
| Cigacice – Rybocice n. Ślubice | Central Odra Valley |
| Rybocice n. Ślubice – Górzycy | Lubuski Gap of the river Odra | Lubuskie Lakeland |
| Górzycy – Osinów Dolny n. Cedymina | Freienwalde Basin | Toruń–Eberswald Ustronreal |
| Osinów Dolny – Szczecin Lagoon | Lower Odra Valley | Szczecin Coastland |
| Szczecin Lagoon | Lower Odra Valley |
| Southern Baltic Coastland |

The catchment basin of the river Odra is highly developed and exceptionally asymmetrical. The basins of the left-bank tributaries situated in the Sudetes and the Sudetes Foreland and that of the river Olza flowing out of the Silesian Beskid are classified as basins of mountain-lowland rivers. The hypsometric system of the entire mesoregion is diversified and characterised by a story system of geo-ecological units. Three types of landscape are distinguished here: mountain and submontane, highland and lowland. The diversity of the environment of the area affects not only the amount of precipitations, but also the rate of runoff and the retention capacity of the catchment basins.

The main left-bank tributaries of the Odra are the rivers: Osobłoga, Nysa Kłodzka, Otawa, Słęza, Bystrzyca, Kaczawa, Bóbr, Nysa Łużycka. Right-bank tributaries, such as the rivers Klodnica, Mała Panew, Widawa and Barycz, are lowland rivers. The right-bank tributary Warta, whose length and catchment basin area are nearly equal to those of the Odra at the point where
the rivers join near Kostrzyń, has a significant effect only on the lower section of the Odra.

The Odra catchment basin is an interesting case among European rivers due to the impact of various civilisation cultures on the state of economic development in consecutive historical periods. The current hydrological regime of the river is an effect of natural geographic and climatic conditions and of ages of human activity, especially intensive in the 19th and 20th centuries. This is clearly observable due to the regulatory and reservoir structures on the river network in the catchment basins of the upper and central river Odra. The bed of the river Odra was once very meandering and over the last 200 years has been shortened about 160 km on many sections due to the digging of cross-cuts on many meanders [14].

In the area of the Odra catchment basin two climates clash – the marine and the continental, which causes a high variation of the weather. According to Polish Institute of Meteorology and Water Management (IMGW PIB) data, the annual precipitation totals in the period of 1961-1980 varied from 545 mm in the catchment basin of the Kaczawa to 1380 mm in the catchment basin of the Bóbr. Among the European rivers, the Odra is one of the least abundant in water. This is shown by analyses of the average flow volume quotient (SSQ) of the Elbe, the Vistula, the Danube and the Rhine to that of the Odra, the mean value of which is 1.06, 1.7, 2.9 and 5.0, respectively. Therefore, the Odra has less water than the Elbe, by about 6%, than the Vistula – by ca. 60%, nearly threefold less than the Danube, and fivefold less than the Rhine.

With regard to the longitudinal slope, the Odra is divided into the upper, central and lower:
- the Upper Odra: from the springs to Koźle, with length of 202 km. On the initial 54 km the Odra has the character of a mountain river, with slope of 7.2‰, while in the territory of Poland the slope is notably less, at about 0.33‰,
- the Central Odra: from Koźle to the mouth of the Warta, with slopes from 0.28 to 0.19‰, has the length of 522 km (including a canalised section of 187 km and a free-flowing section of 335 km);
- the Lower Odra: from the mouth of the Warta to the Szczecin Lagoon, with slopes from 0.05 to 0.00‰.

In the valley of the river Odra floods occur almost every year, either in its upper or lower section, or else over its entire length. Analyses of the frequency of occurrence of maximum diurnal precipitations indicate that the months with a high flood threat on the Odra are July and August.

Over the period of 988-1774 the chronicles recorded 36 great floods in the catchment basin of the Odra. In the 19th century catastrophic floods occurred in 1813, 1854, 1855 and 1888 [15]. In the 20th century there were a number of serious floods, e.g. in 1903, 1915, 1924, 1938, 1940, 1947, 1958, 1960, 1963, 1964, 1965, 1970, 1972, 1977, 1980, 1985, 1997. The greatest flood, with an unprecedented scale, exceeding the most catastrophic estimates, took place in July 1997, second to it being the flood of July 1903, that was considered to be the greatest flood of the 20th century until July 1997.

The formation of a flood wave depends on the parameters of the catchment basin, such as the local topographic, hydrographic and geological conditions, and on human economic activity (retention reservoirs, relief canals, polders, embankments, building-up of mountain brooks, changes in afforestation and land use). On the basis of analysis of historical runs of flood events in the catchment basin of the Odra we can distinguish three types of flood wave formation:
- the reach of critical rainfalls encompasses the mountain part of the catchment basin, causing a violent rise of water levels on the upper Odra, and also in all the tributaries of its upper and central section – e.g. the flood wave of 1997,
- the reach of critical rainfalls encompasses the spring part of the Odra catchment basin. The concentric layout of rivers in that part of the catchment basin (Odra, Ostrawica, Olza, Opava) and their similar lengths, cause wave overlapping on a short section of the Odra, from the mouth of the Opava to that of Ostrawica. The Opava most often contributes to building up of the falling part of the wave, while the Olza causes an extension of the duration of high water levels in the culmination phase, or else forms a separate peak wave, as e.g. in the flood wave of 1985.
- the reach of critical rainfalls encompasses the catchment basins of tributaries of the central section of the Odra (Kaczawa, Bóbr, Nysa Łużycka), e.g. flood wave of 1965.

The probability of occurrence of catastrophic flood waves that would encompass the upper, central and lower sections of the Oder is very low. In the lower section of the Odra the most frequent are spring floods, often related with ice progress down the river, and also during unfavourable storms on the Baltic, inhibiting flood wave waters discharge in to the sea.

2.2 Methods

From among all water gauge stations along the river Odra (Tab. 2) six were selected for comparative analyses of distributions of the MIX and the compound Poisson distribution with the FFA and POT methods. The selected water gauges are marked in bold type in Table 2.

In each case the analysis covered flows from the period of 1966-2010. The water gauge stations were selected so as to emphasise the diversity of the maximum annual flows. The main difference between the standard, unimodal, distributions of maximum flows and the analysed MIX distribution and the compound Poisson distribution is the possibility of taking into account two fundamentally different states of flow in the catchment basin. In the FFA method, due to the manner of gathering of observations, it is difficult to assume that the data set obtained that way will constitute a simple sample. The MIX distribution, due to its structure, allows the description of high-flow events that are of the
order of mean flows from a multi-year period and actual high-flow events. In the POT method, as flows lower than the estimated cut-off threshold are neglected, that problem is solved automatically. The genetic heterogeneity is illustrated in Fig. 1. It is plotted for the water gauge stations in Krzyżanowice (Silesian Lowland) and Słubice (Warta-Odra Urstromtal).

Table 2. Water gauge stations along the river Odra.

| Water gauge station | Km of river length [km] | Catchment basin area [km²] | Water gauge zero ordinate m a.s.l |
|---------------------|-------------------------|----------------------------|----------------------------------|
| Chałupki            | 725.12                  | 4663.69                    | 192.60                           |
| Krzyżanowice        | 712.43                  | 5874.90                    | 184.66                           |
| Racibórz-Miedonia   | 693.29                  | 6728.48                    | 176.28                           |
| Koźle               | 656.36                  | 9040.20                    | 162.51                           |
| Oława Most          | 543.87                  | 19791.24                   | 121.98                           |
| Brzeg Dolny         | 476.08                  | 26448.79                   | 97.73                            |
| Malczyce            | 456.36                  | 26820.38                   | 93.03                            |
| Ścinawa             | 429.53                  | 29611.48                   | 86.73                            |
| Głogów              | 367.49                  | 36411.50                   | 67.89                            |
| Nowa Sól            | 332.95                  | 36839.82                   | 58.82                            |
| Cigacle             | 291.00                  | 39901.33                   | 47.40                            |
| Połęcko             | 233.43                  | 47164.90                   | 32.62                            |
| Słubice             | 179.04                  | 53511.26                   | 17.45                            |
| Gozdowice           | 117.38                  | 109810.77                  | 3.02                             |

For each of the two stations there are two graphs in the Figure. The upper graph shows all the high-flow events in the period of 1975-2010 that exceeded the reference flow $Q_{ref}$ corresponding to the 70% flow taken from the curve of the duration sums and the flows obtained with the FFA method. The lower bar diagram represents the number of high-flow events in a year, subsequently analysed with the POT method. One can observe a distinct decrease in the number of high-flow events on the lower diagram, and even an absence of such events, three times at the turn of the nineteen nineties.

2.3 Estimation of parameters of MIX distribution

The estimation of parameters of the MIX distribution was determined with the method of the maximum likelihood [8]. The global maximum of the likelihood function was determined using the genetic algorithm of search for the global extreme of a multivariate function developed by Kenneth Price and Rainer Storn [16].

2.4 Estimation of parameters of compound Poisson distribution

The POT method [12] was applied to determine consecutive high-flow events from the daily values of flows in the water gauge profiles. In the climatic and hydrological conditions of Poland, for the 45-year series of daily flows the number of high-flow events varies from 50 to 200.

The first step of estimation of the distribution was the analysis of the fit of the maximum flows in a high-flow event to the unidimensional generalised Pareto distribution. Estimation of unknown parameters of the distribution was performed as follows:

- the observed maximum flows were arranged in an increasing order, obtaining sequence $y_i$, $i = 1, \ldots, n$
3 Results

The correctness of the choice of the analysed distributions was verified using the goodness-of-fit test \( \chi^2 \). At confidence level \( \alpha = 0.05 \), in each of the analysed distributions there were no grounds for the rejection of the hypothesis of conformance of the maximum annual flows with the MIX and GPD distributions. The relevant p-values of the test \( \chi^2 \) are presented in Table 3.

| Water gauge station | MIX p-value | GPD p-value |
|---------------------|-------------|-------------|
| Krzyżanowice        | 0.14        | 0.71        |
| Oława Most          | 0.10        | 0.12        |
| Głogów              | 0.16        | 0.09        |
| Cigacice            | 0.10        | 0.09        |
| Słubice             | 0.13        | 0.08        |
| Gozdowice           | 0.17        | 0.13        |

The fitting of both distributions to the empirical data is presented graphically in Fig. 2. The distributions were compared by marking the distribution functions in each graph:

- empirical – blue stars;
- estimated MIX distribution – red line;
- estimated compound Poisson distribution – black line.

The horizontal axis in the Figure is specially formatted in the scale of probability so as to facilitate the interpretation of the results obtained. The standard linear scale is in this case transformed by an inverse function to the gamma distribution [17].

Analysis of the distributions presented in the graphs allows the formulation of the following observations:

1. The MIX distribution is less sensitive to outstanding high values of flows. In all the graphs its distribution function limits the variation of maximum flows from below. This means that the n-year mean water estimated with the MIX will limit its actual variation from below.

2. The CP distribution is more sensitive to high values of flows. In almost all cases the graph of the distribution function limits the variation of flows from above. At all analysed water gauges the curve of the CP distribution lies within the area of confidence [8] determined by the MIX distribution.

3. At two water gauge stations situated on the lower Odra (Słubice, Gozdowice) the graph of the CP distribution diverges strongly from the observed maximum annual flows. In those cases the conducted tests of conformity (Kolmogorov’s \( \lambda \) test) suggested the rejection of the hypothesis of conformity.

4. The poor fit of the CP distribution to the maximum annual flows at Słubice and Gozdowice results from the difficulty of estimation of the distribution of the number of high-flow events in a year. In the study it was assumed that it was a Poisson distribution [18], but due to the low variation of the number of high-flow incidents there is no possibility of correct conformance testing (in this case the estimated means values of 50- and 100-year waters in Tab. 4 are marked in red).

Table 4 presents the mean 50- and 100-year waters estimated with the MIX and CP distributions. In addition, in the last column, the estimated values of 100-year water are given, obtained with the standard methods (IMGW PIB). Except for the gauge station at Oława, one can note an overestimation of the flows obtained with the standard methods in relation to the MIX distribution. The values confirm also a considerable overestimation of all estimated means of n-year waters determined with the POT method.
On the basis of the conducted statistical inference one can formulate the following conclusions:

1. Both methods, the MIX distribution and the compound Poisson distribution (CP) can be used for the determination of maximum flow distributions. However, at the lower section of the Odra an additional analysis of the distribution of the number of high-flow events in a year with the POT method is necessary.

2. The advantages of using both distributions are the following:
   - The MIX and GPD distributions provide good fit to strongly diversified data;
   - The estimations of mean n-year waters obtained with their use limit them from above – CP distribution, and from below – MIX distribution;

3. The disadvantages of using both distributions are the following:
   - Long, at least 40-year stationary series of observed flows are required,
   - Tests of conformance of both distributions with observation data require high precision and, due to a small amount of data, are difficult to automate.

4. At the parameters $\xi > 0.5$ (GEV) and $\kappa < -0.5$ (GPD) both distributions have unlimited variance – it is then impossible to determine the areas of confidence for high quantiles;

5. The form of the estimated distributions is strongly affected by high noted flows, and their potential inaccurate measurement will have a significant impact on the determination of the value of estimation.

**References**

1. R. M. Vogel, I. Wilson., J. Hydrol. Eng. 69, 69–76 (1996)
2. R. W. Katz, M. B. Parlange, P. Naveau, Adv. Water Resour. 25, 1287–1304 (2002)
3. L. Xiong, T. Du, C. Xu, S. Guo, C. Jiang, C.J. Gippel, Water Resour. Manag. 29, 3615-3633 (2015)
4. R. Kidson, K.S. Richards, Prog. Phys. Geogr. 29(3), 392–410 (2005)
5. D.R. Maidment, Handbook of hydrology (Mc Graw Hill, 1993)
6. B. Bobee, Water Resour. Res. 11(5), 681-689 (1975)
7. S. Węglarczyk, Gospodarka Wodna 11, 323-328 (2015)
8. W. Szulczewski, W. Jakubowski, Water Resour Manage 32, 3519–3534 (2018)
9. R. Engle, A. Lunde, J Financ Economet 1(2), 159–188 (2003)
10. A. Żyromski, W. Szulczewski, M. Biniak-Pieróg, W. Jakubowski, Renew Sust Energ Rev 65, 1118–1126 (2016)
11. W. Szulczewski, A. Żyromski, W. Jakubowski, M. Biniak-Pieróg, Renew Sust Energ Rev 82(2), 1787-1795 (2018)
12. P. Teodorović, V. Zelenhasić, Water Resour. Res. 6(6), 1641-1648 (1970)
13. J. Kondracki, Geografia regionalna Polski (PWN, Warszawa, 2002)
14. A. Arkuszewski, Monografie Budowli Hydrotechnicznych w Polsce (Wydawnictwa Komunikacji i Łączności, Warszawa, 1985)
15. A. Dubicki, J. Malinowska-Małek, K. Strońska, Limnologica 35(3), 123-131 (2005)
16. K. V. Price, R. M. Storn, J. A. Lampinen, Differential Evolution - A Practical Approach to Global Optimization (Springer-Verlag, 2005)
17. Z. Kaczmarek, Metody statystyczne w hydrologii i meteorologii (Wydawnictwa Komunikacji i Łączności, Warszawa, 1970)
18. Q. Wang, J Hydrol 129, 263-280 (1991)