

Attempts to establish a regional probabilistic model of intense rainfall for the Upper and Middle Oder River basin

Marcin Wdowikowski^{1,*}, *Bartosz Kaźmierczak*², *Ondrej Ledvinka*³, and *Matěj Jedlička*³

¹Institute of Meteorology and Water Management - National Research Institute, Podlesna 61, 01-673 Warsaw, Poland

²Wrocław University of Science and Technology, Faculty of Environmental Engineering, Plac Grunwaldzki 13, 50-377 Wrocław, Poland

³Czech Hydrometeorological Institute, Hydrology Database and Water Budget Department, Na Šabatce 2050/17, 143 06 Prague 412, Czechia

Abstract. The main scope of this study was to present the possibility of developing a heavy rainfall regional model, based on a probabilistic approach. Following the methods and conclusions of the scientific and technical work of the research team in the Upper and Middle Oder basin, carried out on rainfall data between 1961 and 2010, it was found that the generalized exponential distribution enables a satisfactory probabilistic description of heavy rainfall, and the results of the models constructed on its basis are obtained in an accessible and not time-consuming way. The regional model presented in the article sufficiently describes the empirical heavy rainfall at eight analyzed meteorological stations, especially in the range of rainfall amounts occurring most frequently from a statistical point of view, i.e. with the p -value from 0.3 to 1.0.

1 Introduction

The safe design of urban areas drainage requires reliable information on rainfall conditions in a specific region, derived statistically from long-term meteorological measurements and observations [1–4]. The determination of probable rainfall is also necessary to identify areas exposed to the occurrence of floods, regardless of their nature (flash floods or urban floods) [5–9]. In Poland, there is still a deficiency of scientific works in the field of probabilistic modelling of intense rainfall; therefore, this article is an attempt to reduce this deficit. The paper is a brief description of a consistent continuation of the research activities undertaken by the authors in the works [8, 9]. Thus, for the description of the study area and methodological details we rather refer to the previously published works and indicate only the necessary differences or modifications here. At the beginning, before the preparation of this work, there were primarily the conclusions of previous studies [1, 2, 8–11]. This paper attends to the attempt if the development of the regional rainfall model is possible for data series with different precipitation characteristics at locations with different

* Corresponding author: marcin.wdowikowski@imgw.pl

elevation. In order to achieve the intended goal, the quantiles of the probability distribution best fitted to random variables representing eight individual stations will be compared with the quantiles of an averaged time series variable that theoretically represents the largest precipitation heights for the whole Upper and Middle Oder basin. The purpose of the proposed simplification is to answer the question whether the generalization of rainfall data, the quality of which is usually the subject of many scientific discussions, as the input to the rainfall model, can give satisfactory results from the viewpoint of engineering applications.

2 Study area and methods

Following the paper [9], the basic study material was composed of one-minute rainfall series from meteorological stations located in the Upper and Middle (also called Central) Oder basin covering the years from 1961 to 2010, provided by the Polish Institute of Meteorology and Water Management – National Research Institute (IMWM-NRI). Figure 1 (adopted from [9]) shows the eight stations located in the Oder basin with elevation depicted in meters a.s.l. The Upper and Middle Oder was limited by the profiles of water gauges in Chalupki, directly located on the Polish-Czech border (downstream of Bohumin), and Slubice, located near the Polish-German border. Due to these research assumptions, only the Polish part of the Upper Oder basin was taken into account. With respect to [9], some potential meteorological stations such as Cieszyn, Leszno, Grabik and Słubice (not shown in Fig. 1) had to be excluded from further analyses because of the violation of the data series homogeneity due to the analyzed long-term period.

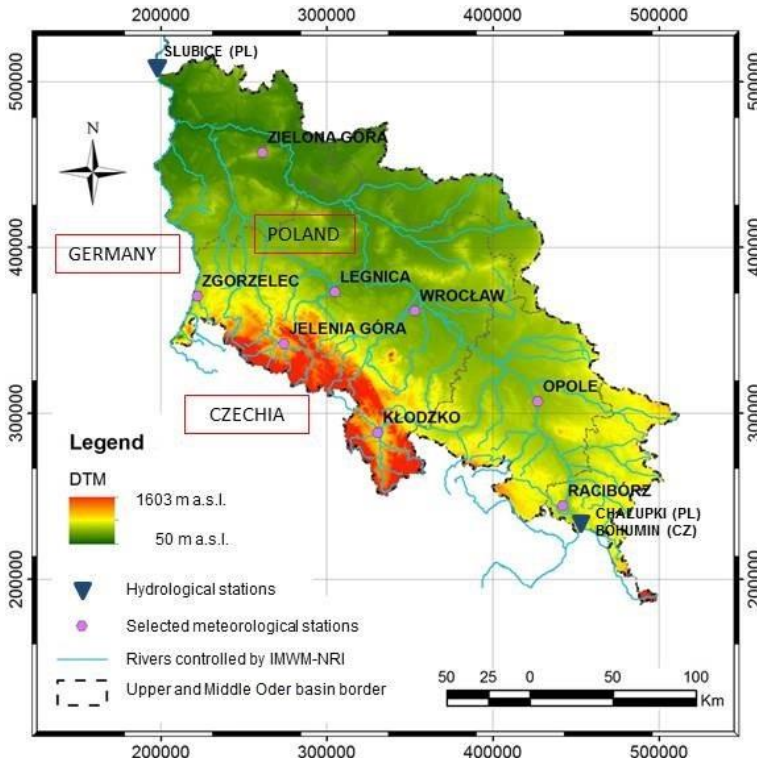


Fig. 1. Location of the eight analyzed meteorological stations of the IMWM-NRI within the Upper and Middle Oder basin (adopted from [9], based on the PUWG-92 coordinate system).

Probabilistic modelling of intense rainfall was made on the basis of the methodology developed and constantly improved by the research team of the Wrocław University of Science and Technology (Faculty of Environmental Engineering) and the IMWM-NRI, specifically for rainfall data from Wrocław (2009–2015) [1, 2, 10] and then from Legnica [11]. Simplified procedure for the development of a probabilistic model assumes the following tasks:

1. separation of maximum interval precipitation heights for rainfall duration from 5 minutes to 3 days based on the available precipitation events in the assumed 16 time periods (5, 10, 15, 30, 45, 60, 90, 120, 180, 360, 720, 1080, 1440, 2160, 2880 and 4320 minutes); this method is sometimes called the Partial Duration Series (PDS) approach,
2. selection of representative random samples of the largest precipitation values for the analyzed multi-year period, in this case 50 years of measurements, Important: For further analyses, the selection of the highest rainfall values can be made directly if one takes only the 50 largest values from the non-increasing sequence of all peaks found. In such a case, however, the independence of rainfall data above the accepted cut-off threshold (POT) must be maintained. Otherwise, the annual maximum precipitation (AMP) method should be used.
3. selection of the probability distribution that best fits the measurement data (through a numerical estimation of probability distribution parameters followed by a goodness-of-fit criterion assessment),
4. determination of the dependence of selected probability distributions parameters using the rainfall duration and the probability of exceedance.

Following the methods and results given in [1, 2, 8–11] and also the theoretical basis in [12], in this research, the four probability distributions were fitted to the data series from the eight stations, namely the Gumbel, Weibull, Log-Normal and Generalized Exponential Distribution (GED). The goodness of fit was provided by the BIC and rRMSE criteria. The results were presented in figures containing a comparison between the regional model and the eight station maximum empirical data series for chosen time periods (editorial frames of this paper did not allow to present all 16 time periods for eight stations; moreover, the comparison of all time periods for all analyzed stations would be unclear). All necessary numerical calculations were performed using the R-package statistical modules ('FAdistr', 'fExtremes', 'MASS' and 'bbmle') compiled in our own constructed scripts.

3 Results

With respect to the research methods described in Section 2, maximum rainfall data series were derived from the long-term period database representing the eight analyzed stations. As a result of this data separation, we got eight data frames combined with 50 maximum values of 16 rainfall durations (from 5 to 4320 minutes). In the next step, the averaging of particular rainfall durations from all eight stations was done in order to get regional data frame containing maximum values for the Oder River basin. The procedure of choosing the best-fitting probability distribution was applied to the maximum rainfall data prepared in described way. The results of the BIC and rRMSE criteria for each rainfall duration are summarized in Table 1. The smallest values (in bold) indicate the greatest correlation between the theoretical distribution and the regional average obtained using the eight measured data series. In general, for rainfall durations between 5 and 2160 minutes both criteria pointed out that GED is the best-fitting probability distribution. These results are in strict accordance with the conclusions of the work for Wrocław [1, 2, 10], Legnica [11] and several stations in the catchment area of the Lusatian Neisse River basin [8]. Despite major differences for the 2880- and 4320-minute rainfall duration, GED was adopted for further

analysis as the best-fitting. In fact, for the purposes of designing and operating drainage systems of urbanized areas, the rainfall duration limited to one day is sufficient [1, 2]. The GED goodness of fit is shown in Fig. 2.

Generalized Exponential Distribution

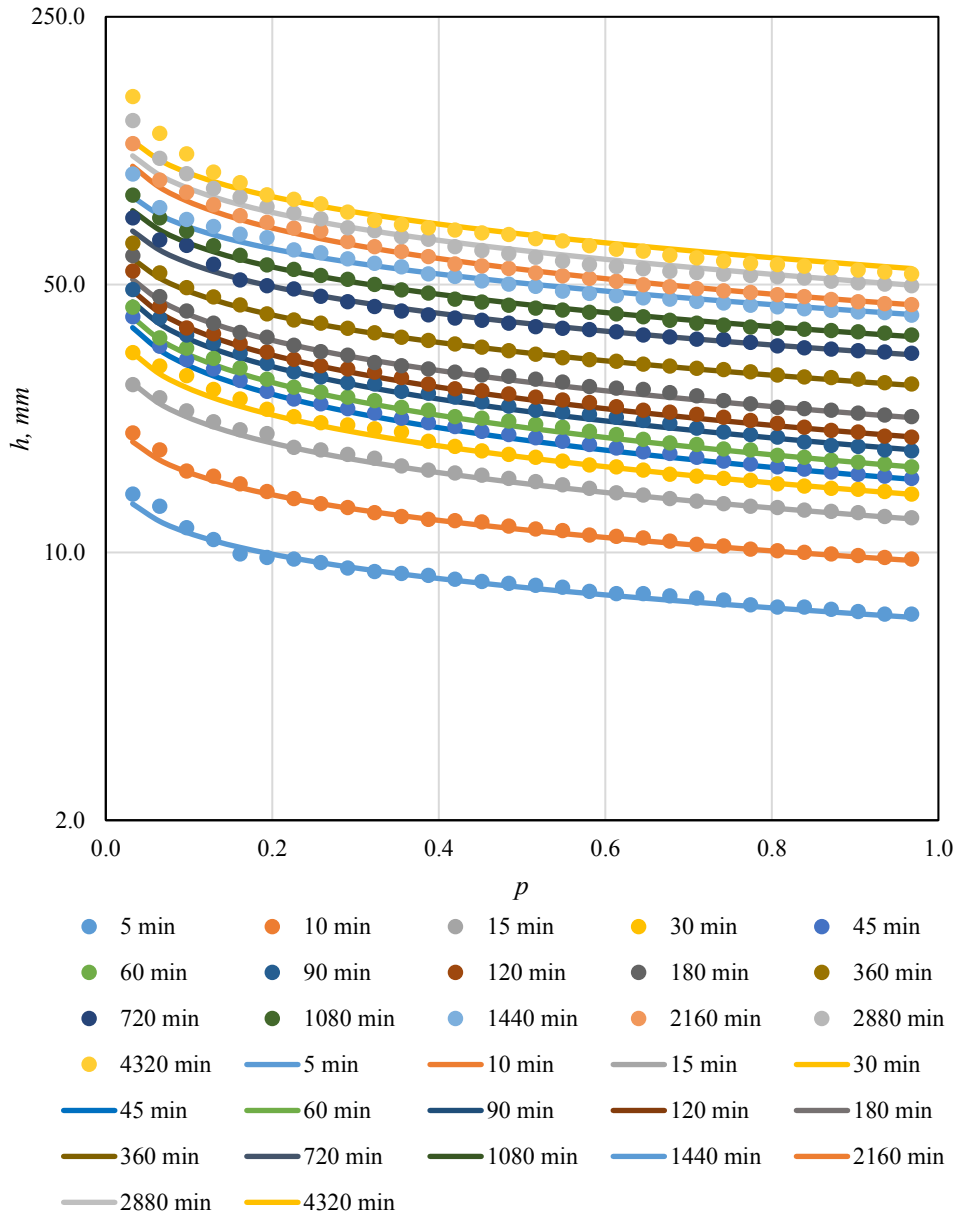


Fig. 2. Comparison of the GED theoretical maximum rainfall heights (lines) and their empirical regional-average counterparts (points) in the Upper and Middle Oder basin for all 16 rainfall durations. Vertical axis represents rainfall heights (h [mm]) while horizontal axis the probability of exceedance (p [-]).

It can be clearly seen in Fig. 2 that, for averaged measurements, GED was best fitted to rainfall durations t ranging from 5 to 1440 minutes. For $t \in [2160;4320]$, the differences ranged generally from 11% to 23%, and in the extreme case of the highest precipitation for the duration $t = 4320$ minutes almost 31%.

Table 1. BIC information criterion and rRMSE values for analyzed probability distributions and rainfall duration of averaged data set.

t [min]	BIC				rRMSE			
	GED	Gumbel	Log-Nor	Weibull	GED	Gumbel	Log-Nor	Weibull
5	182.90	190.60	185.60	182.58	2.07	4.62	2.19	2.24
10	215.64	226.50	218.84	215.48	1.42	5.13	1.77	1.66
15	250.92	266.50	255.92	251.20	1.54	6.17	3.08	1.60
30	268.72	288.24	273.52	269.06	1.89	7.15	4.02	1.97
45	278.02	303.52	286.72	278.90	1.68	7.52	4.90	1.66
60	290.62	310.86	296.96	291.18	1.67	7.20	3.68	1.67
90	298.20	319.44	304.42	298.78	1.80	7.32	3.59	1.78
120	302.34	325.28	308.52	302.90	2.07	7.43	3.72	2.00
180	312.46	329.94	316.90	312.76	2.19	6.69	2.54	2.14
360	325.64	337.20	327.96	325.60	1.66	5.78	1.48	2.04
720	338.34	352.12	340.72	338.48	2.27	6.29	1.79	2.32
1080	351.10	366.14	353.14	351.24	2.19	6.61	1.68	2.21
1440	369.42	379.48	370.20	369.50	2.60	6.04	1.89	2.65
2160	372.14	397.04	377.24	372.58	2.36	7.77	2.64	2.04
2880	394.48	408.50	393.86	394.44	3.80	7.73	2.33	3.26
4320	406.20	419.94	393.86	406.40	4.91	8.41	10.09	4.12

The final step of this research was to compare the GED quantile determined based on the regional-average data set with the quantiles corresponding to the empirical data for the eight analyzed stations. Figure 3 shows the correlation level between the GED theoretical regional-average rainfall and the empirical maximum rainfall for the eight meteorological stations.

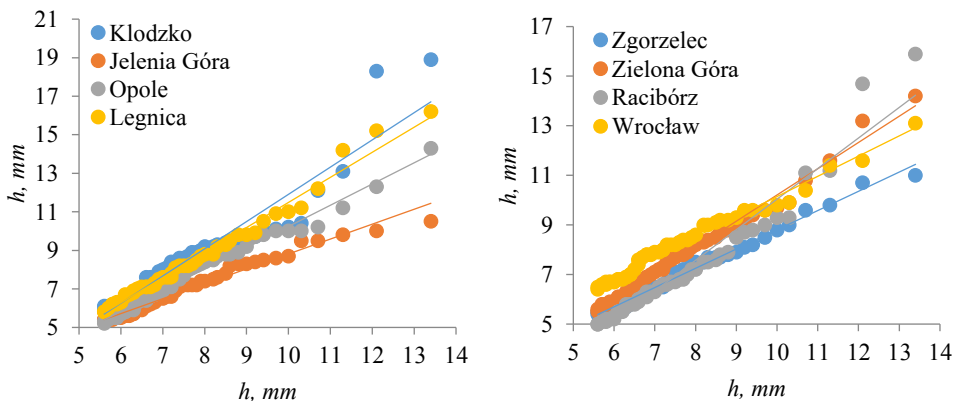


Fig. 3. Correlation between the GED regional-average rainfall height (h [mm]) and the empirical maximum rainfall height (h [mm]) (coloured dots and linear trend lines with R^2 described in text) for the eight meteorological stations of the IMWM-NRI for the rainfall duration $t = 5$ minutes.

Simple regression correlation factor indicated a high level of model conformity to measured data. The values of the R^2 coefficient, for the rainfall duration $t = 5$ minutes, were respectively (decreasing order): Zgorzelec 98.95%, Zielona Góra 98.91%, Legnica 98.42%,

Opole 98.35%, Wrocław 98.18%, Jelenia Góra 97.08%, Racibórz 94.70%, Kłodzko 90.64%. A strong correlation in the range of heavy rainfall for durations such as $t = 5$ minutes is very important for the drainage design purposes. Figure 4 shows that the differences between theoretical and measured data were primarily due to the least-frequent rainfall with the probability of exceedance up to $p = 0.1$. The largest differences in precipitation values were 5.5 mm and 6.2 mm for Kłodzko, 2.9 mm for Jelenia Góra and 2.4 mm for Zgorzelec. The remaining differences did not exceed 2.0 mm, and in the majority of cases 1.0 mm, which can be considered a very good result.

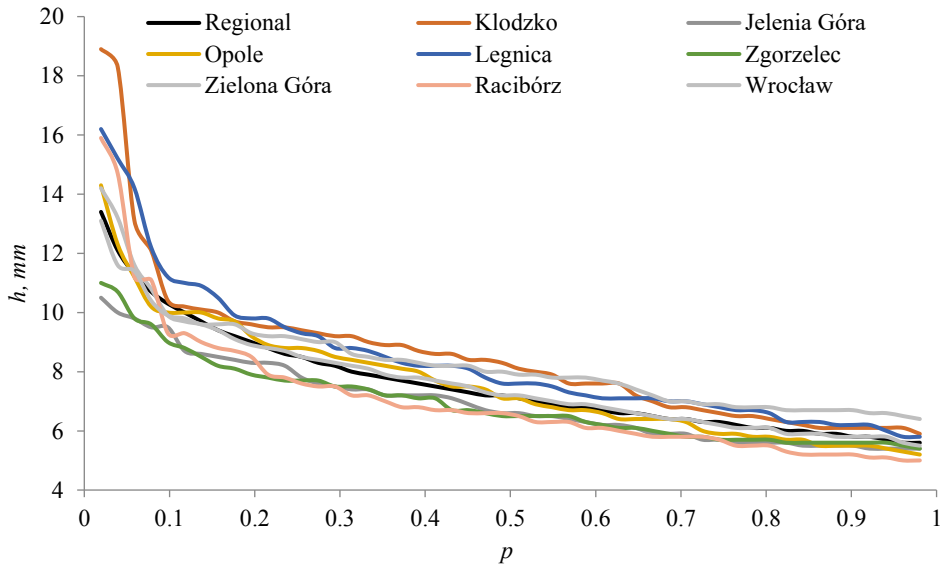


Fig. 4. The GED regional-average rainfall height (h [mm]) (black line) and the empirical maximum rainfall height (h [mm]) for the eight meteorological stations of the IMWM-NRI in the Upper and Middle Oder basin for the rainfall duration $t = 5$ minutes and their dependence on the probability of exceedance (p [-]).

The editing scope of the work does not allow to present, in a similar manner, all the results obtained for all analyzed rainfall durations. Therefore, Fig. 5 presents one of the negative examples for $t = 1440$ minutes. Significantly inferior-quality fit of the GED theoretical data to the measurement data for the duration of precipitation from a few to several hours was confirmed in the literature of the subject and has been confirmed in previous studies conducted for Wrocław [1, 2, 10], Legnica [11], several Lusatian Neisse basin stations [8] and 12 meteorological stations in the Upper and Middle Oder basin [9]. The example presented in Fig. 5 shows that the data from the Racibórz, Opole and Jelenia Góra stations are significantly different not only from the proposed regional model but also from other measurement stations, which, on the other hand, show satisfactory compatibility between themselves and the regional model. Distinct differences also apply to the least expected precipitation corresponding to the probability of exceedance p between 0.01 and 0.3. The differences here range from 22% to 39%, which gives a rainfall difference of up to 45.6 mm. The topographic location of these three stations is not without significance in the observed situations because it is a submontane area prone to the occurrence of intense rainfall that is hardly predictable. Also, in this area, floods harmful to the technical infrastructure, human property and the natural environment are very common [7].

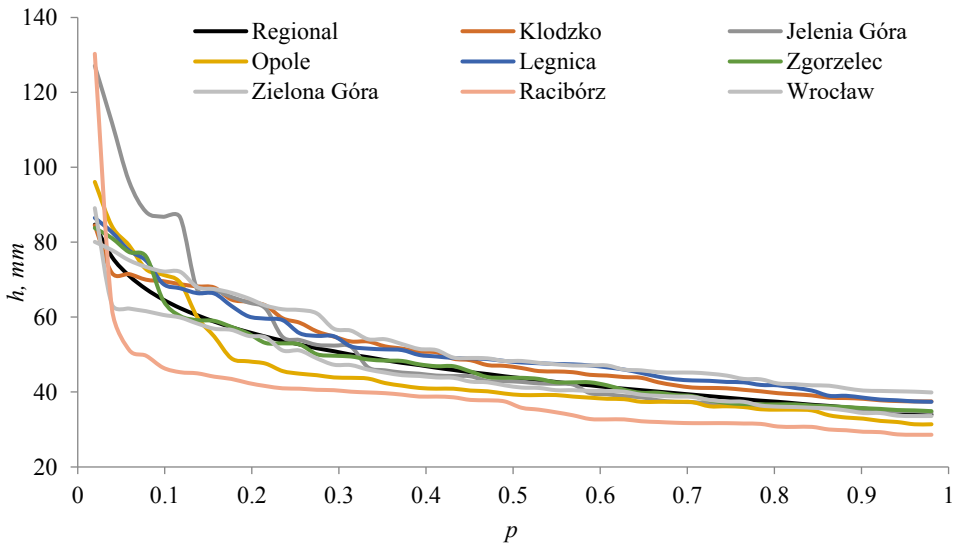


Fig. 5. The GED regional-average rainfall height (h [mm]) (black line) and the empirical maximum rainfall height (h [mm]) for the eight meteorological stations of the IMWM-NRI in the Upper and Middle Oder basin for the rainfall duration $t = 1440$ minutes and their dependence on the probability of exceedance (p [-]).

It is worth mentioning that the occurrence of extremely high rainfall values such as those from Racibórz (i.e. $h = 130.3$ mm for $t = 1440$ minutes or $h = 227.4$ mm for $t = 4320$ minutes) poses considerable difficulties when trying to describe the precipitation distribution. This is because most of the quality tests indicate such values as outliers. The statistical description of the precipitation data in mountain areas is difficult on the one hand, but necessary on the other in order to carry out the further research for various tasks, especially for the purpose of water supply and sewerage.

3 Conclusions

Eight meteorological stations and four probability distributions were used to describe the long-term rainfall data in the Oder River basin (1961–2010). Averaging of the maximum rainfall durations values allowed to create a regional model based on the GED distribution. An evaluation analysis led to the formulation of the following conclusions:

- general conclusion: we can provide well-fitting (to measured data) local probabilistic rainfall models, but it is much harder to develop a regional model for the terrain with varying elevation (in the analyzed Oder basin area from ~100 m to nearly 400 m a.s.l. – without higher parts of the Sudetes),
- the best-fitting probability distribution was GED,
- the differences between model and measured data were mainly higher in the extreme maximum values range (probability of exceedance p between 0.02 and 0.3).
- for Legnica, Kłodzko, Zielona Góra, Jelenia Góra and for most rainfall durations at Opole and Wrocław, regional model described measured data to a satisfactory degree,
- the largest differences between the data from the regional model and those from the eight stations were usually between 20–30%, and in extreme cases they did not exceed 40%.

- more and more detailed analyses (including the usage of other probability distributions and a spatial analysis) should be performed in the future for the Oder River basin for flash flood and urban flood protection purposes,
- if possible, similar analyses should also be done in the Czech parts of the Oder catchment and, especially, in its mountainous parts in both Czechia and Poland; maybe, in whole Poland (in order to find the regional schemes and also the probability distribution parameters similarities).

References

1. A. Kotowski, *The Principles of Safe Dimensioning of Sewer Systems* (Seidel-Przywecki, Warsaw, 2015)
2. A. Kotowski, B. Kaźmierczak, A. Dancewicz, *The Modeling of Precipitation for the Dimensioning of Sewer Systems* (Polish Academy of Sciences, Warsaw, 2010)
3. P. Licznar, K. Siekanowicz-Grochowina, M. Oktawiec, A. Kotowski, E. Burszta-Adamiak, *Ochrona Środowiska* **40**, 17 (2018)
4. P. Licznar, E. Burszta-Adamiak, A. Kotowski, K. Siekanowicz-Grochowina, M. Oktawiec, *Ochrona Środowiska* **40**, 21 (2018)
5. W. Szalińska, I. Otop, T. Tokarczyk, *Meteorology Hydrology and Water Management* **2**, 13 (2014)
6. L. Kuchar, W. Szalińska, S. Iwański, L. Jelonek, *Meteorology Hydrology and Water Management* **2**, 49 (2014)
7. G. Dumieński, M. Krzyżanowski, A. Tiukało *E3S Web of Conferences* **44**, 00037 (2018)
8. M. Wdowikowski, B. Kaźmierczak, O. Ledvinka, *Meteorology Hydrology and Water Management* **4**, 53 (2016)
9. M. Wdowikowski, A. Kotowski, P.B. Dąbek, B. Kaźmierczak, *E3S Web of Conferences* **17**, 00096 (2017)
10. B. Kaźmierczak, A. Kotowski, *J Hydrol* **525**, 345 (2015)
11. B. Kaźmierczak, M. Wdowikowski, *Periodica Polytechnica Civil Engineering* **60**, 305 (2016)
12. S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer Series in Statistics, Springer-Verlag, London, 2001)