

## Criteria for identifying maximum rainfall determined by the peaks-over-threshold (POT) method under the Polish Atlas of Rainfall Intensities (PANDa) project

Ewa Burszta-Adamiak

*Wrocław University of Environmental and Life Sciences, Institute of Environmental Engineering, pl. Grunwaldzki 24, 50-363 Wrocław, Poland, e-mail: ewa.burszta-adamiak@upwr.edu.pl*

Paweł Licznar

*Wrocław University of Technology, Faculty of Environmental Engineering, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland, e-mail: pawel.licznar@pwr.edu.pl*

Janusz Zaleski

*Institute of Meteorology and Water Management – National Research Institute, Podleśna 61, 01-673 Warsaw, Poland, e-mail: janusz.zaleski@imgw.pl*

**Abstract.** Determination of rainfall maxima from long-term series is one of the more important tasks in urban hydrology. These maxima are useful both in designing land drainage systems and for flood protection in a catchment. The identification of rainfall maxima for the hierarchy of rainfall durations from 5 min to 4320 min is a fundamental stage of the creation of the first version of the Polish Atlas of Rainfall Intensities (PANDa), which will ultimately be a source of updated and reliable information on design rainfall intensities for designing and modeling rainwater drainage and retention systems in Poland. One of the methods for identifying extreme rainfall events is to use criteria for selecting rainfall based on their depth for a given rainfall frequency and duration. Existing national experience in this respect is based on the results of analyses usually conducted with regard to records from single weather stations. This article presents the results of a study designed to verify the usefulness of the literature-based criteria for identifying rainfall maxima using the peaks-over-threshold (POT) method at a much broader nationwide scale. The study analyzed data from a previously created digital database of rainfall series, which includes 3 000 station-years (consisting of a 30-year measurement series from 100 weather stations of the Institute of Meteorology and the Water Management – National Research Institute (IMGW-PIB)). The study results show that as far as the investigated measurement series are concerned, the criteria based on the literature sources have limited application and can only be used for identifying the largest short-duration rainfall events. To determine rainfall maxima for all of the time intervals analyzed (from 5 minutes to 3 days), it was necessary to develop our own criteria that would allow the methodology for identifying extreme rainfall events to be standardized for all 100 stations.

**Keywords:** rainfall model, Intensity-Duration-Frequency (IDF), Depth-Duration-Frequency (DDF), extreme value theory (EVT)

**Submitted** 23 January 2018, **revised** 4 June 2018, **accepted** 19 July 2018

### 1. Introduction

Over the last several years, the frequency of extreme rainfall events has clearly increased, and the coincidence of their various characteristics over a short time has produced multiplied detrimental effects. Because of this, many areas in Poland must cope with regular hydraulic overloads on their drainage systems, minor localized flooding, high-water events on rivers, and floods. This situation is aggravated by the increasing level of catchment imperviousness. During intense rainfall events, rapid destructive surface runoff occurs, which additionally results in contamina-

tion of receiving water bodies with substances leached from their drainage basins. The destructive impact of the extreme meteorological and hydrological characteristics affects the social and economic life of the impacted areas. Thus, the extreme value theory finds an ever widening application not only in natural science, but also in many other areas such as social sciences, urban planning and economics. Today, reliable data regarding extreme values are useful not only in analyses designed to estimate credit, insurance or market risk (Sobiecki 2013; Muela et al. 2017), but also in hydrology and meteorology (Towler et al. 2010). The knowledge in this area is used in designing

flood control facilities and in developing risk maps, and it is also used for sizing stormwater drainage systems whose task is to provide the required land drainage standard. For all of these tasks, it is necessary in the first place to determine the highest rainfall amounts from a long-term measurement series (a minimum 30-year series) (Licznar 2008; WMO 2011). It is also important to identify the depths of maximum rainfall from different locations and, based on that, to be able to rely on reliable data for a specific region where measures aimed at enhancing a rainwater management system are taken.

The above-mentioned factors inspired work associated with the development of the Polish Atlas of Rainfall Intensities (PANDa). This atlas is patterned after the idea of the KOSTRA rainfall atlas (in German *Koordinierte Starkniederschlags-Regionalisierungs Auswertungen*) that has already been used in Germany for 30 years (Bartels et al. 1997). The PANDa atlas will contain rainfall models in the form of Intensity-Duration-Frequency (IDF) and Depth-Duration-Frequency (DDF) relationships for the entire area of Poland. Similar to the German atlas, it is assumed that the rainfall models being currently developed as part of industrial research and development (this project is entitled *Development and implementation of the Polish Atlas of Rainfall Intensities* (PANDa) – the Smart Growth Operational Programme 2014-2020 POIR.01.01.01-00-1428/15) are to be periodically updated in the future. Owing to this, the PANDa atlas will be a source of updated and indisputable information on design rainfall intensities for designing and modeling rainwater drainage and retention systems across Poland. Currently under the PANDa project, design rainfall intensities are analyzed for a network of 100 rain gauges (Fig. 1). Thanks to a geostatistical analysis of this data set, actual values of the current design rainfall intensities are planned to be ultimately obtained for each city in Poland. One of the essential elements of the PANDa methodology is the issue of identifying rainfall maxima for the hierarchy of rainfall durations from 5 min to 4320 min.

First, in section 2, the standard procedures for determining rainfall maxima are discussed, and the study goals are formulated. Then, section 3 presents a description of the PANDa's digital rainfall series database elaboration and the methodology of criteria for determining the limit value (threshold) of the maximum rainfall. Section 4 reports and discusses the development of new criteria and their performance compared with criteria from the literature.

## 2. Methods for determining rainfall maxima

Selection of maximum rainfall events from rainfall series can be potentially accomplished using three methods. The first one is the so-called annual maximum series (AMS) method (Hailegeorgis, Alfredsen 2017; Nguyen et al. 2017). This selection consists of identifying exactly one maximum rainfall (with the maximum interval intensity) in each calendar year, regardless of the actual rainfall duration. This method was employed, for example, by Bogdanowicz and Stachý (1998) to construct regionalized Depth-Duration-Frequency (DDF) rainfall models for Poland (excluding mountainous areas) and by Avanzi et al. (2015) to analyze the spatial variability of DDF parameters over Italy. When ranking rainfall events using this method, there is a risk that a large number of intense rainfall events will not be included if they are not the highest in a specific period analyzed. This is of great importance, particularly in years in which high-depth or high-volume rainfall occurred relatively frequently. Because of this, the above-mentioned annual maximum series method was clearly criticized by Kotowski et al. (2011). Based on detailed analysis of rainfall records for Wrocław, they showed that the omission of large numbers of intense rainfall events leads to a dangerous underestimation of design rainfall intensities for high probabilities and low rainfall frequencies.

The second method of determining rainfall maxima is based on selection of an annual exceedance (AE) series. Rainfall maxima (with the maximum interval intensity) are selected irrespective of the date of their occurrence, and the number of selected maximum rainfall events is equal to the number of observation years. The application of this method can be found, among others, in the publications of Kupeczyk and Suligowski (1997) as well as Suligowski (2004), which address the Intensity-Duration-Frequency (IDF) relationships for Poland.

Some authors propose that the number of measurement series be increased by including not only the annual maxima but also secondary extremes occurring in a given year. The selection of upper extremes, i.e. the maxima from a set of observations, consists of selecting rainfall events that exceed a certain arbitrarily high level determined as the threshold value. This method is called the peaks-over-threshold (POT) method. It is well known in financial market analysis, but it also applies to climatic or hydrological risk analysis such as flood and low flow frequency analysis (Norbiato et al. 2008; Joshi, St-Hilaire 2013), flash flood warning (Miao et al. 2016; Gharib et al.

2017) and regional climate model simulations (Roth et al. 2014). Such an approach allows us to obtain a larger statistical sample, for which it is necessary to define a threshold value that determines whether a particular rainfall event can be classified as an extreme event (Dupuis 1998). This approach is recommended by many authors (Lang et al. 1999; Re, Barros 2009; Trambly et al. 2013; Bezak et al. 2014; Wdowikowski et al. 2016; Gharib et al. 2017) and by the Intergovernmental Panel on Climate Change (IPCC 2014), among others. A detailed description of the POT method can be found in the papers by Coles (2001), and Bommier (2014).

In the POT method, there are several possibilities for determining the threshold value, including both analytical and graphical approaches. Solutions can also be found where the threshold excess is predefined and assumes values depending on the specific needs of the respective study and not on the threshold value arising from the characteristics of the investigated phenomenon. Such a case may occur when determining the probability of occurrence of flood events causing flood damage. The allowable flow is the boundary flow that is frequently set as the threshold (Byczkowski et al. 2008). The limits for threshold airborne pollutant concentrations are another example of a predefined threshold excess. Some authors (Coles 2001; Scarrott, MacDonald 2012) have made an attempt to determine the threshold excess using graphical methods (e.g., the mean excess over threshold plot and the Gerstengarbe-Werner plot). For the needs of urban hydrology, rainfall maxima above the threshold are most often selected based on analytical methods.

The aim of this study was to verify the usefulness of the literature-based cutoff criteria for determining rainfall maxima for series obtained from a 30-year recording period at 100 IMGW-PIB weather stations. At the final stage of this analysis, an attempt was made to develop our own (new) criterion that would allow the methodology for determining maximum rainfall events to be standardized for all of the stations analyzed.

### 3. Data and methods

#### 3.1. Study area and data

The research material used in this study was the digital rainfall series database developed during the first stage of the PANDa project. This database included 30-year rainfall records (1986-2015) from 100 stations belonging to the Institute of Meteorology and Water Management – National Research Institute (IMGW-PIB). The locations of rain gauges included in the digital database are shown in Figure 1.

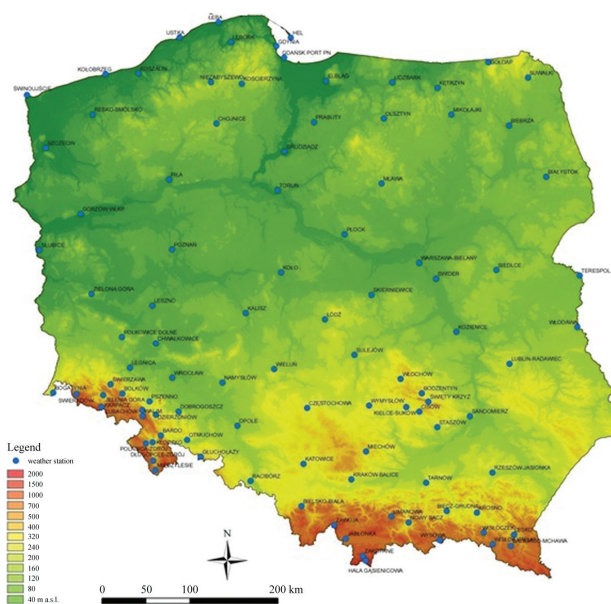


Fig. 1. Locations of rain gauges included in the digital database developed under the PANDa project

The original rainfall series collected in the database were recorded using rain gauges that met the national standards valid at the time when relevant measurements were made. Over the last 30 years, these standards have significantly changed. Because of this, only some of the rainfall measurements came from new-generation electronic rain gauges. Their remaining measurements (from most of the stations for the period between 1986 and 2000) were recorded by pluviographs on strip charts. Because these instruments did not have a heating system and could freeze in winter, they were only used during the period of positive temperatures: the period from approximately May 1 to October 30. The analog pluviographic records were converted to digital format using the methodology developed by Licznar (2005). Based on the converted pluviographic data and the available records from electronic rain gauges, pseudo time-series were generated with uniform 1-minute time resolution.

To search for maximum rainfall depth values, a specialized software application was developed. Using iteration, this application calculated the highest interval-specific rainfall depths, expressed in mm, for the entire hierarchy of time interval durations of 5, 10, 15, 30, 45, 60, 90, 120, 180, 360, 720, 1080, 1440, 2160, 2880, and 4320 minutes, respectively. In seeking the highest interval-specific rainfall depths, the average number of occurrences of a phenomenon was adopted for each weather station. With 30-year rainfall records (1986-2015), this meant that it was necessary to find 60 maximum values of interval-specific rainfall depth. During the next step, the frequency distributions were ranked according to the non-increasing interval values of rainfall depth (in the a.m. period) and

the exceedance probabilities ( $p$ , %) calculated according to equation:

$$p(m, N) = \frac{m}{N+1} \quad (1)$$

where  $m$  represents the place of a specific term in the rainfall distribution series, and  $N$  is series size; these values were assigned to the synthetic rainfall time series created in this way.

The exceedance probability statistically corresponds to rainfall frequency ( $C$ , years), as determined from the following formula:

$$C = \frac{100}{p} \quad (2)$$

where  $p$  is the exceedance probability expressed in %.

The threshold values used to identify extreme rainfall were determined during the measurement data analysis as values with an empirical rainfall frequency of  $C = 1$  year ( $p = 100\%$ ).

### 3.2. Methods

#### 3.2.1. Literature criteria

In the literature, various criteria can be found, which are used to set a limit value (threshold) to determine the maximum rainfall depth (or sometimes intensity) for a given period. To select maximum rainfall events of the strong, heavy and torrential rainfall types, Chomicz (Kotowski et al. 2010) used a qualitative criterion of the following form:

$$U_k = \sqrt{2^k} \sqrt{t} \quad (3)$$

where  $U_k$  is the depth of rainfall of category  $k$  (with  $k \in \{0; 9\}$ ) [mm], and  $t$  is the rainfall duration [min]. Afterwards Błaszczuk developed a physical model commonly used for drainage systems design in Poland based on discontinuous rainfall events recorded in Warsaw during the periods 1837-1891 and 1914-1925 (Kotowski et al. 2010; Węglarczyk 2013). Within the whole observational set, Błaszczuk distinguished the groups of strong, heavy and torrential rainfall based on the criterion that determined rainfall depth according to the following inequality (Kotowski et al. 2010):

$$h \geq t^{0.5} \quad (4)$$

where  $h$  is rainfall depth [mm], and  $t$  is rainfall duration [min].

Based on the research material used by Błaszczuk, Chomicz eliminated 8 strong rainfall events, limited the statistical analysis to 71 heavy and torrential rainfall

events, and identified rainfall events that met the criteria according to the following inequality for the entire rainfall duration (that is, above the curve  $U_1$  according to the Chomicz scale) (Kotowski et al. 2010; Węglarczyk 2013):

$$h_{U_1} \geq (2t)^{0.5} \quad (5)$$

where  $h_{U_1}$  is rainfall depth above the curve  $U_1$  in the Chomicz scale [mm], and  $t$  is the rainfall duration [min].

To identify intense rainfall events, Kotowski proposed his own selection criterion for interval-specific rainfall depths, based on the qualitative criterion proposed by Chomicz (equation 3):

$$h_{min} \geq 0.75t^{0.5} \quad (6)$$

where  $h_{min}$  is the lowest value of the rainfall depth [mm] from the maximum rainfall series with an assigned rainfall frequency of  $C = 1$  year, and  $t$  is the rainfall duration [min].

The selected criterion allowed him to distinguish in each year from several to over a dozen synthetic rainfall events as design rainfall events for statistical analysis of the probability and frequency of their occurrence (Kotowski et al. 2010; Kotowski 2011a, b). The implementation of the adopted criterion for interval-specific rainfall depths for a 50-year observation period (from the months of May-October), recorded at IMGW-PIB's Wrocław-Strachowice weather station, led to the selection of a total of 514 synthetic rainfall events.

#### 3.2.2. New criteria

To determine the threshold excess for maximum depth rainfall for all of the stations analyzed, two new criteria were developed, which are referred to as criteria I and II in this article. The first one (criterion I) included maximum rainfall depths for each station, based on the distribution frequency determined for all time intervals using one power functional dependence. In this way, rainfall with depth greater than  $h_{min}$  was selected as follows:

$$h_{min} \geq a \cdot t^b \quad (7)$$

As an alternative, the other criterion (criterion II) was based on distinguishing two rainfall duration groups, which differ in the slope of the point pattern in the plot. For each group, a different functional dependence was determined, described by formula (8) for the durations from 5 min to some preselected breakeven time  $t_b$  and by formula (9) for the durations from  $t_b$  to 4320 min (3 days).



$$h_{min} \geq a_1 \cdot t^{b_1} \quad (8)$$

$$h_{min} \geq a_2 \cdot t^{b_2} \quad (9)$$

Our own threshold selection criterion was sought in order to identify rainfall maxima for the 30-year observation period and for a rainfall frequency of  $C = 1$  year ( $p = 100\%$ ).

For the above-mentioned formulas, parameters fitting all inequalities were linearized as follows:

$$\ln(h_{min}) \geq \ln(a) + b \cdot \ln(t) \quad (10)$$

Assignment of the parameters  $a$  and  $b$  was performed by means of the least squares method. In the case of criterion II, retrieval of the  $a_1$ ,  $a_2$ ,  $b_1$  and  $b_2$  parameters was performed for a whole hierarchy of breaktimes  $t_b$  equal to analyzed intermediate time durations between 5 and 4320 minutes. Numerical calculations for thirteen separate breaktimes  $t_b$  were performed in Matlab using the `lsq_lut` piecewise function, allowing for a piecewise linear least squares fit with a continuity constraint.

The mean absolute error ( $MAE$ ) was used as a prime measure of divergence between the rainfall depth values determined according to the criteria based on our own interpretation and those recorded in the nationwide network of weather stations over the period 1986-2015. Regarding the absolute value of rainfall depth, the  $MAE$  shows the mean deviation of the value determined based on the criterion from the actual value over the period analyzed. Thus, the evaluation of  $MAE$  values allowed for comparison of the goodness of fit of different models relating  $h_{min}$  to time scales, as well as the final selection of the best criterion for threshold excess for maximum rainfall depths.  $MAE$  values were determined from the following formula:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_i^p| \quad (11)$$

where  $n$  is the number of determined differences (equal to the number of time intervals analyzed),  $y_i$  is the rainfall depth determined according to the criterion, and  $y_i^p$  is the rainfall depth recorded at the respective IMGW-PIB weather station.

As an extra measure of divergence between the rainfall depth values determined according to the criteria based on our own interpretation and those empirically retrieved as part of the PANDa project, the mean squared error ( $MSE$ ) was used:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y_i^p)^2 \quad (12)$$

Estimated values of the  $MSE$  were afterwards used for calculation of the  $AIC$  (Akaike Information Criterion):

$$AIC = N \cdot \ln(MSE) + 2 \cdot L_p \quad (13)$$

where  $L_p$  is the number of estimated model parameters. In the case of criterion I, the simple model  $L_p$  was equal to 2. In the case of criterion II, the model was defined by formulas (8) and (9); with the additional value of  $t_b$ , the total number of parameters has increased to 5 ( $L_p = 5$ ).

Models characterized by the lowest  $AIC$  value were assumed as optimal with respect to their performance, i.e. in agreement with predicted and empirical rainfall depths (low values of  $MAE$  and  $MSE$  errors), as well as from the perspective of model parsimony (reduction in the total number of model parameters  $L_p$ ).

#### 4. Results and discussion

In the peaks-over-threshold (POT) method, the values of examined rainfall events that are below the threshold excess are not included in the analysis. To be able to assign an exceedance probability of  $p = 100\%$  ( $C = 1$ ) over the 30-year observation series, at least 30 such rainfall events were necessary for each of the 16-time intervals and for each measuring station. Thus, the POT method gives a minimum of 480 rainfall events selected as maximum rainfall depths that occurred over the period 1986-2015. Using the digital rainfall database and the specially designed computer application, the rainfall series were reviewed, and the above-mentioned 480 values of maximum rainfall depths for the 16 durations analyzed were determined for each of the 100 stations considered. These values were ranked as decreasing distribution frequencies whose final values represent the maximum rainfall depths with an assigned frequency of  $C = 1$ . They are presented graphically in Figure 2 for all of the stations analyzed.

Spatial variations in maximum rainfall depths with an assigned frequency of  $C = 1$  can be clearly observed based on a comparison of the rainfall time series for the 100 stations analyzed. The lowest values were recorded at the Biebrza, Łeba and Gdynia stations, located in northern Poland (seaside). The highest maximum rainfall depths were recorded for a 5-min rainfall at the Polanica station, a 10-min rainfall at the Tarnów station, a 15-min rainfall in Wisłok Wielki, a 30-min rainfall in Wysowa, and at the Hala Gąsienicowa station for the other rainfall durations (45-4320 min). These stations are located in the south of Poland, with the Hala Gąsienicowa station being situated at the highest elevation (1523 m a.s.l.) among the stations

considered. For this reason, it is a station that clearly deviates from a rather dense cluster of plots obtained for the other 99 stations shown in Figure 2.

The literature-based criteria for setting threshold values of rainfall depths performed well in determining threshold excess for rainfall events that were recorded at one or several weather stations. However, the conducted analysis, whose results are shown in Figure 2, demonstrates that these criteria do not give satisfactory results for most of the 100 stations analyzed due to the temporal and spatial variations. The Błaszczyk criterion performed well only in setting the threshold excess for rainfall events with a duration of up to 360 min, which means that the rainfall depths recorded at the stations analyzed for the subsequent time intervals (from 720 min to 3 days) had lower values than those determined from inequality (4). The Chomicz cri-

terion was even less applicable to the database analyzed. The threshold excess could be determined by means of inequality (5) only for rainfall events lasting no more than 180 min. The recorded series of rainfall maxima with durations longer than 3 h had lower depths than those included in the analysis performed by Chomicz and classified as heavy and torrential rainfall according to his scale. Such a trend was observed at as many as 13, 69 and 78 stations out of 100 for the time intervals corresponding to rainfall lasting 180 min, 360 min and 720 min, respectively. Regarding long-duration rainfall events, the Chomicz criterion performed well only in the case of 2 stations out of 100 with regard to rainfall lasting from 1080 to 1440 min and only for 1 station with regard to rainfall events with durations from 2160 to 4320 min. Kaźmierczak and Kotowski (2012) drew similar conclusions, claiming that

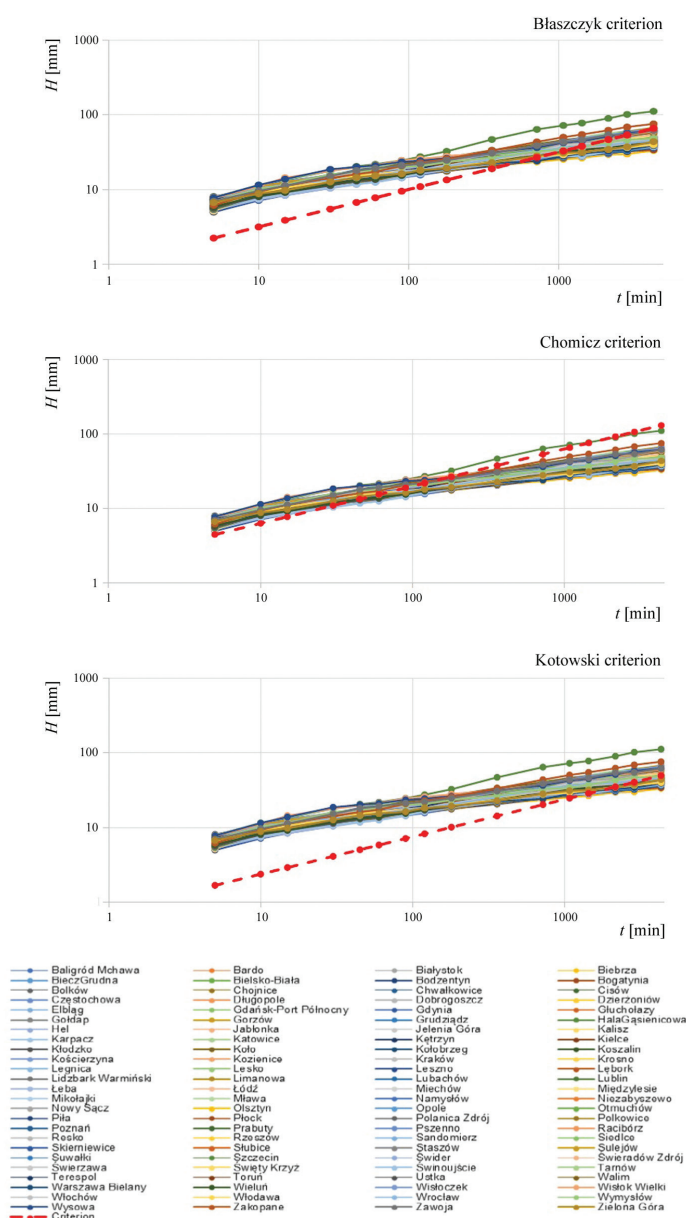


Fig. 2. Rainfall depths ( $h$ ) over time ( $t$ ) for the time series with an assigned frequency of  $C = 1$  ( $p = 100\%$ ) for each of the stations relative to the values calculated according to the Błaszczyk, Chomicz and Kotowski criteria (red dotted line) (in double logarithmic form)

the Chomicz criterion does not allow one to isolate a sufficiently large number of data for long-duration rainfall events. At this stage of the analysis, it is worth mentioning that neither Błaszczuk nor Chomicz took into account rainfall with durations of up to 3 days. Similarly to the authors of the present paper, Kotowski et al. (2010) included such a time interval in their study. The justification for analyzing design rainfall intensities for longer durations exceeding several hours, up to even 3 days, arises from the German experiences where the above-mentioned KOSTRA rainfall atlas provides engineers with information on design rainfall depths for rainfall durations from 5 min to 4320 min.

The comparison of the time series of the highest rainfall with an assigned frequency of  $C = 1$  year, compiled by Kotowski et al. (2010) for Wrocław-Strachowice and for the 100 stations included in the analysis in this paper, demonstrates that it is possible to use the Kotowski criterion for the time series analyzed, but only for rainfall no longer than 1440 min (1 day). For longer durations, the recorded rainfall maxima have higher depths than those measured in Wrocław-Strachowice and those determined from inequality (6). In the analyzed case, this means that the Kotowski criterion, as described by formula (6), is not applicable to two stations when determining rainfall maxima lasting 2160 min and to 37, 48 and 74 stations out of 100 in the case of rainfall events with durations of 2160 min, 2880 min and 4320 min, respectively.

The impossibility of using the literature criteria to determine the threshold excess for rainfall maxima with an assigned frequency of  $C = 1$  ( $p = 100\%$ ), with respect to rainfall events with durations from 5 min to 3 days recorded at the 100 weather stations located across Poland, forced us to develop our own criteria. The goal set by the authors of the present paper was to find a functional dependence that could be considered representative across Poland. The sought-after formula was to have the quality of universality. That is, it should have a simple functional form, but at the same time it should allow one to determine the threshold at the level of rainfall maxima with an assigned rainfall frequency of  $C = 1$  year, which would be acceptable for all of the 100 weather stations and all of the time intervals, i.e. both short- and long-duration rainfall events (from 5 min to 3 days). The developed formula should maximally limit the number of frequency distributions of maximum rainfall depths, obtained based on the review of rainfall series for the individual durations. Simultaneously, as a safe formula, the criterion should not lead to elimination of any maximum rainfall depth values corresponding to frequencies above  $C = 1$  year.

To develop the above-characterized universal criterion, among the maximum rainfall depths with an assigned fre-

quency of  $C = 1$  ( $p = 100\%$ ) for the 100 analyzed stations as shown in Figure 2, their minimum values were selected for each of the time durations analyzed. A classical dependence with a structure as described by inequality (7), similar to those used by Chomicz, Błaszczuk and Kotowski, was fitted to these values with parameters  $a$  and  $b$  equal to 4.0845 and 0.2613, respectively. According to this formula (criterion I), only rainfall with the following interval depths should be included in the POT analysis:  $h \geq 6.2$  mm for time  $t = 5$  min,  $h \geq 7.5$  mm for  $t = 10$  min,  $h \geq 8.3$  mm for  $t = 15$  min,  $h \geq 9.9$  mm for  $t = 30$  min,  $h \geq 11.0$  mm for  $t = 45$  min,  $h \geq 11.9$  mm for  $t = 60$ ,  $h \geq 13.2$  mm for  $t = 90$  min,  $h \geq 14.3$  mm for  $t = 120$  min,  $h \geq 15.9$  mm for  $t = 180$  min,  $h \geq 19.0$  mm for  $t = 360$  min,  $h \geq 22.8$  mm for  $t = 720$  min,  $h \geq 25.3$  mm for  $t = 1080$  min,  $h \geq 27.3$  mm for  $t = 1440$  min,  $h \geq 30.4$  mm for  $t = 2160$  min,  $h \geq 32.7$  mm for  $t = 2880$  min, and  $h \geq 36.4$  mm for  $t \geq 4320$  min (3 days).

The determined interval-specific rainfall depths as described by functional dependence (7) for the entire range of rainfall durations are shown in Figure 3. This function was plotted on a double logarithmic scale of the x and y axes. In such a scale, the relationship between rainfall depths and rainfall durations has a characteristic pattern (similar to a linear dependence). With the data so compiled, one can clearly see the divergence between the maximum rainfall depths with an assigned frequency of  $C = 1$  recorded at the weather stations (marked as points in the plot) and the values determined based on criterion I. A detailed analysis of these data allowed us to observe that the rainfall depth values were not plotted along a single straight line for the entire range of the rainfall durations analyzed. For the intermediate rainfall durations of approximately 2 h, a change was observed in the slope of the point pattern in the plot. This phenomenon is most likely due to the discontinuity in the scale properties of the rainfall events arising at some specific timescale of so-called breaktime. A detailed discussion of this issue can be found in a monograph by Licznar (2009).

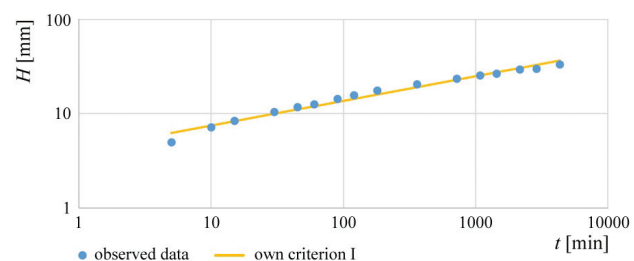


Fig. 3. The lowest maximum rainfall depths with an assigned frequency of  $C = 1$  ( $p = 100\%$ ) from the 100 weather stations relative to the values determined based on criterion I (in double logarithmic form)

Table 1. Estimated  $a_1$ ,  $a_2$ ,  $b_1$  and  $b_2$  parameters of criterion II models for a hierarchy of breaktimes  $t_b$ , with associated  $MAE$ ,  $MSE$  and  $AIC$  values

Breakttime $t_b$ [min]	Parameter				$MAE$ [mm]	$MSE$ [mm <sup>2</sup> ]	$AIC$
	$a_1$	$a_2$	$b_1$	$b_2$			
10	1.6820	4.4434	0.6714	0.2495	0.8079	1.2858	14.0221
15	2.0480	4.6831	0.5467	0.2412	0.6535	0.8447	7.2996
30	2.5656	5.0272	0.4281	0.2203	0.5452	0.4869	-1.5151
45	2.7759	5.2319	0.3909	0.2244	0.4798	0.3675	-6.0165
60	2.8883	5.4344	0.3733	0.2190	0.4382	0.2854	-10.0618
90	3.0127	5.8316	0.3557	0.2089	0.3696	0.1773	-17.6786
<b>120</b>	<b>3.0886</b>	<b>6.2068</b>	<b>0.3460</b>	<b>0.2002</b>	<b>0.3185</b>	<b>0.1344</b>	<b>-22.1110</b>
180	3.1811	6.9658	0.3353	0.1844	0.3236	0.1445	-20.9516
360	3.3559	9.0097	0.3178	0.1500	0.5097	0.4414	-3.0849
720	3.5708	12.8493	0.2988	0.1042	0.7934	1.0310	10.4885
1080	3.7085	17.4559	0.2887	0.0659	0.9121	1.4130	15.5314
1440	3.7947	12.8655	0.2811	0.1133	0.9497	1.4083	15.4781
2160	3.8883	20.9704	0.2744	0.0549	1.0141	1.6785	18.2864

For the above-mentioned reasons, criterion II was postulated, as described by equations (8) and (9), towards providing a better fit of the threshold values of rainfall depths with the values observed for the rainfall frequency of  $C = 1$  year. Results of numerical calculations of the  $a_1$ ,  $a_2$ ,  $b_1$  and  $b_2$  parameters for a hierarchy of breaktimes  $t_b$  are summarized in Table 1. In view of these results, the breakttime  $t_b$  should be adopted to be 120 minutes, and the  $a_1$ ,  $a_2$ ,  $b_1$  and  $b_2$  parameters of criterion II should be equal to 3.0886, 6.2068, 0.3460 and 0.2002 respectively. With such parameters, criterion II is characterized by the lowest  $MAE$  and  $MSE$  values and is ranked as the best model based on the lowest  $AIC$  value. Figure 4 provides a graphical representation of this model relative to the empirical values of the lowest maximum rainfall depths with an assigned frequency of  $C = 1$  ( $p = 100\%$ ) from the 100 weather stations in Poland.

The introduction of criterion II clearly reduces the  $MAE$  and  $MSE$  errors in comparison with criterion I.  $MAE$  and  $MSE$  errors for simple criterion I were high and equal to 1.0932 mm and 1.9512 mm<sup>2</sup> respectively. Clearly, the last values were close to the respective values reported in Table 1 for the longest considered breakttime ( $t_b = 2160$  min). In contrast, the same errors were one order of magnitude lower for the optimal breakttime  $t_b$  of 120 min. The  $MAE$  error (0.3185 mm), expressed as the absolute value of rainfall depth, was even comparable to the rainfall depth resolution of the original recordings from electronic tipping-bucket gauges (0.1 mm). Similarly, the low value of  $MSE$  error (0.1344 mm<sup>2</sup>) resulted in a very low  $AIC$  value (-22.1110) for criterion II. This value was still much lower for criterion II with respect to criterion I

( $AIC = 14.6954$ ), despite the obvious penalty for an extra three model parameters.

Figure 5 shows the differences (in %) in the threshold rainfall depth values determined based on the literature criteria and based on criteria I and II relative to the rainfall data (the lowest values of the maximum rainfall depths with an assigned rainfall frequency of  $C = 1$  year) recorded at the 100 stations over the 30-year observation period. A positive value in Figure 5 means that the threshold rainfall depth, calculated using the literature criterion or criteria I or II for determining the threshold excess, was higher than that recorded at the analyzed stations. This is an unfavorable situation because higher rainfall depth values (treated as the threshold) in the frequency distributions arranged in decreasing order would “cut off” days with rainfall maxima lower than those determined as the threshold values by the criterion. Hence, such “cut off” rainfall events could not be included in further analysis. Such a situation would occur in the case of the Chomicz criterion (already from  $t \geq 180$  min), the Błaszcyk criterion (from  $t \geq 720$  min), and the Kotowski criterion (from  $t \geq 1$  day). For rainfall depths determined as the threshold according to the criteria adopted based on the literature sources and the depths of rainfall maxima with an assigned frequency of  $C = 1$  year from the 100 weather stations, the greatest differences were observed in the case of the Chomicz criterion. These differences reach even 180% in the case of rainfall maxima with durations of 4320 min (3 days). Likewise, a negative value in Figure 5 means that the threshold rainfall depth, calculated based on the literature criteria or criteria I or II, was lower than that recorded at the analyzed stations. This is an acceptable situation



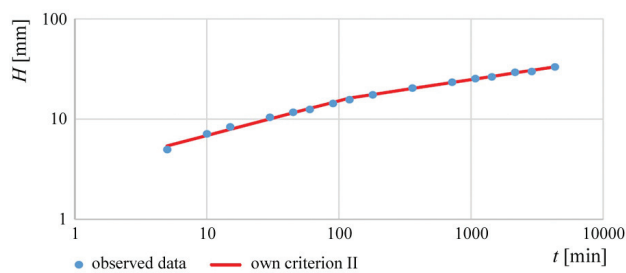


Fig. 4. The lowest maximum rainfall depths with an assigned frequency of  $C = 1$  ( $p = 100\%$ ) from the 100 weather stations relative to the values determined based on criterion II (in double logarithmic form)

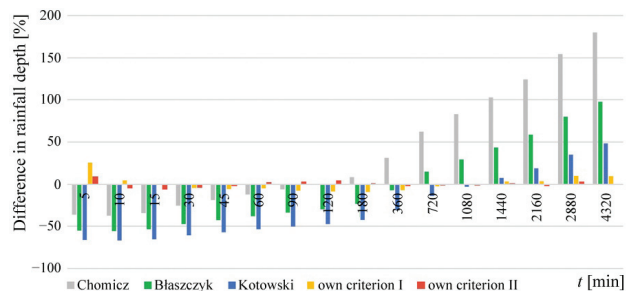


Fig. 5. Differences in % between rainfall depths calculated according to the criteria and rainfall depths for  $C = 1$  obtained from the 100 IMGW-PIB stations

from the reliability point of view. The data thus presented show that it is possible to use the Chomicz criterion to determine the threshold for rainfall with durations of up to 120 min, the Błaszczyk criterion for durations of up to 360 min, and the Kotowski criterion for durations of up to 1 080 min. In comparison, the performance of criteria I and II is much better, and the more advanced criterion II could be used in practice for all durations from 5 min to 3 days.

## 5. Conclusions

Rainfall characteristics are the result of many factors, among which location and measuring period are considered to be the most important. The conducted analysis shows that the literature criteria for determining the threshold excess for rainfall maxima, which performed well for single rain gauges analyzed by the other authors, could not be used to describe rainfall events from a 30-year measurement period for the 100 IMGW-PIB weather stations. The results revealed that as longer rainfall durations were taken into account, the number of stations decreased where the literature criteria for determining the threshold excess for rainfall maxima could be used. Thus, for rainfall events with the longest interpreted durations (3 days), the Chomicz criterion could be used only for 1 station, the Błaszczyk criterion could be used for 4 stations, and the Kotowski criterion could be used for 26 stations.

The Hala Gąsienicowa station, with the highest recorded rainfall depths, is among these stations.

The usefulness of the Chomicz criterion (for rainfall depths above the curve  $U_1$ , according to formula (5)) for determining rainfall maxima using the POT method is the lowest for the stations analyzed. Out of the 100 stations, it would not be possible to determine the threshold values at 13 stations for rainfall maxima with an assigned exceedance probability of  $p = 100\%$  ( $C = 1$  year) and with a duration longer than 180 min. This criterion performs well only for 1 station, i.e. the Hala Gąsienicowa station, with respect to rainfall events with durations from 2 160 to 4320 min. Due to the a.m. limitations in using the criteria established by the other authors, it was necessary to formulate a criterion to be universal for determining the threshold excess, with such a criterion being uniform for rainfall recorded at the 100 weather stations. The present study demonstrates that such a criterion cannot be satisfactorily described by a single functional dependence for the entire range of rainfall durations analyzed. It seems much more useful to develop criterion II, where the threshold excess is determined by two separate equations (8) and (9) relating to short and long rainfall durations, respectively. This is evidenced by the lower values of the mean absolute and mean squared errors obtained for this criterion and the lowest (not exceeding 9%) differences between the recorded rainfall depths from the frequency range of  $C = 1$  year and the depths calculated based on criterion II.

The developed criterion II is recommended by the authors of the present paper to be applied in POT analysis of rainfall maxima in Poland, with the assumption that one is interested in maximum rainfall depths with an exceedance probability of  $p = 100\%$  ( $C = 1$  year). This criterion is useful in practice, particularly when we do not yet have full records of rainfall series in digital form and hence it is necessary to review the records on paper pluviographic strips. The criterion allows us to quickly reject many rainfall records whose intensities are too low and that do not affect the sought-after maximum (design) intensities. From this perspective, the presented criterion II can also be used in the POT method in practice as a rough estimation of the minimum values of rainfall depths (intensities) with a probability of  $p = 100\%$  ( $C = 1$  year) that should be used, e.g., in designing rainwater drainage systems. In the authors' opinion, the developed criterion II has a definite cognitive value. Its structure clearly suggests that it is not possible to determine a power functional dependence that combines rainfall depth with rainfall duration by using a single set of parameters for a very wide range of rainfall durations from 5 min to 4320 min. It is necessary to separately consider rainfall maxima with short dura-

tions and rainfall maxima with longer durations, which may be due to the different origin of their occurrence. The maxima of short-duration rainfall can be an effect of short and intensive convective rainfall events, whereas long-duration rainfall maxima can be produced by frontal rainfall events. This issue will be the subject of in-depth analysis during a further stage of this study as part of the conducted quality assessment (*QA*) of the determined rainfall depth maxima by using, among others, historical synoptic records. Finally, the functional dependence of criterion II, with its break for the duration of 120 minutes, demonstrates the need to use similar, in terms of their structure, dual physical dependencies describing design rainfall intensities for the 100 stations analyzed. This suggestion is in agreement with the results of the study in this area currently conducted under the PANDa project, which will be the subject of a separate scientific note. In the future, the PANDa atlas should also be an objective tool for assessing trends in changes in short-duration rainfall intensities in Poland due to climate change. Finally, an important scientific and utilitarian purpose of this project is to formalize the tools used to process rainfall data and construct models for design rainfall intensities in Poland. For this reason, the outcomes of the PANDa project currently being implemented should include not only rainfall models themselves but also a comprehensive rainfall data processing methodology used for this purpose.

### Acknowledgments

The project entitled *Development and implementation of the Polish Atlas of Rainfall Intensities (PANDa)* – POIR.01.01.01-00-1428/15, is financed by the National Centre for Research and Development under the Smart Growth Operational Programme 2014-2020, Priority axis: Support for R&D in enterprises; Measure: R&D projects in enterprises; Sub-measure: Industrial research and development conducted by enterprises.

### Bibliography

Avanzi F., De Michele C., 2015, Orographic signature on extreme precipitation of short durations, *Journal of Hydrometeorology*, 16, 278-294, DOI: 10.1175/JHM-D-14-0063.1

Bartels H., Malitz G., Asmus S., Albrecht F. M., Dietzer B., Günther T., Ertel H., 1997, *Starkniederschlagshöhen für Deutschland (1951-2000)*, KOSTRA-DWD-2000, Selbstverlag des Deutschen Wetterdienstes. Offenbach am Main, available at [https://www.dwd.de/DE/fachnutzer/wasserwirtschaft/kooperationen/kostra/fortschreibung\\_pdf.pdf%3F\\_\\_blob%3DpublicationFile%26v%3D3](https://www.dwd.de/DE/fachnutzer/wasserwirtschaft/kooperationen/kostra/fortschreibung_pdf.pdf%3F__blob%3DpublicationFile%26v%3D3) (20.07.2018)

Bezak N., Brilly M., Šraj M., 2014, Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis, *Hydrological Sciences Journal*, 59 (5), 959-977, DOI: 10.1080/02626667.2013.831174

Bogdanowicz E., Stachy J., 1998, Maximum rainfall in Poland. Project characteristics, (in Polish), *Materiały Badawcze IMGW. Seria: Hydrologia i Oceanologia*, 23, 85 pp.

Bommier E., 2014, Peaks-Over-Threshold modelling of environmental data, U.U.D.M. Project Report 2014:33, Uppsala Universitet, available at <https://uu.diva-portal.org/smash/get/diva2:760802/FULLTEXT01.pdf> (20.07.2018)

Byczkowski A., Banasik K., Hejduk L., 2008, The calculation of probable annual flood flows, (in Polish), *Infrastruktura i Ekologia Terenów Wiejskich*, 5, 199-208

Coles S., 2001, *An introduction to statistical modeling of extreme values*, Springer Series in Statistics, Springer, 209 pp.

Dupuis D.J., 1998, Exceedances over high thresholds: A guide to threshold selection, *Extremes*, 1 (3), 251-261, DOI: 10.1023/A:1009914915709

Gharib A., Davies E.G.R., Goss G.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 (9), 692, DOI: 10.3390/w9090692

Hailegeorgis T.T., Alfredsen K., 2017, Analyses of extreme precipitation and runoff events including uncertainties and reliability in design and management of urban water infrastructure, *Journal of Hydrology*, 544, 290-305, DOI: 10.1016/j.jhydrol.2016.11.037

Joshi D., St-Hilaire A., 2013, Low Flow Frequency analysis of three rivers in Eastern Canada, Institut National De la Recherche Scientifique, Centre Eau, Terre et Environnement (INRS-ETE), Québec, available at <http://espace.inrs.ca/2663/1/I322.pdf> (20.07.2018)

Kaźmierczak B., Kotowski A., 2012, Depth-duration-frequency rainfall model for dimensioning and modelling of Wrocław drainage systems, *Environment Protection Engineering*, 38 (4), 127-138, DOI: 10.5277/EPE120411

Kotowski A., 2011a, Methodological basis for formulating models of dependable precipitations for measuring sewerage systems, (in Polish), *Przegląd Geofizyczny*, LVI (1-2), 45-97

Kotowski A., 2011b, The need to standardize the modeling for dimensioning storms sewers in Poland, (in Polish), *Gaz Woda i Technika Sanitarna*, 7/8, 260-266

Kotowski A., Kaźmierczak B., Dancewicz A., 2010, The modeling of precipitations for the dimensioning of sewage systems, (in Polish), *Studia z Zakresu Inżynierii*, 68, 170 pp.

Kotowski A., Kaźmierczak B., Dancewicz A., 2011, Safe of sewage systems based on the local rainfall models, (in Polish), *Czasopismo Techniczne. Środowisko*, 1 (108), 85-99

- Kupczyk E., Suligowski R., 1997, Statistical description of the precipitation time structure as an input element for hydrological models, (in Polish), [in:] *Przydziały opadów i wezbrań o zadanym czasie powtarzalności*, U. Soczyńska (ed.), University of Warsaw, Warsaw, 17-82
- Lang M., Ouarda T.B., Bobee B., 1999, Towards operational guidelines for over-threshold modeling, *Journal of Hydrology*, 225 (3-4), 103-117, DOI: 10.1016/S0022-1694(99)00167-5
- Licznar P., 2005, A proposal of precipitation records processing for the needs of urban drainage systems' design and exploitation, (in Polish), *Woda-Środowisko-Obszary Wiejskie*, 5, 197-207
- Licznar P., 2008, Calculating frequency of precipitation sewage system damming up, (in Polish), *Gaz, Woda i Technika Sanitarna*, 7-8, 16-21
- Licznar P., 2009, Synthetic rainfall time-series generators for needs of stormwater and combined sewage systems modelling, (in Polish), *Uniwersytet Przyrodniczy, Wrocław*, 180 pp.
- Miao O., Yang D., Yang H., Li Z., 2016, Establishing a rainfall threshold for flash flood warnings in China's mountainous areas based on a distributed hydrological model, *Journal of Hydrology*, 541 (Part A), 371-386, DOI: 10.1016/j.jhydrol.2016.04.054
- Muela S.B., Martín C.L., Sanz R.A., 2017, An application of extreme value theory in estimating liquidity risk, *European Research on Management and Business Economics*, 23 (3), 157-164, DOI: 10.1016/j.iedeen.2017.05.001
- Nguyen T-H., Outayek S.E., Lim S.H., Nguyen V.-T.-V., 2017, A systematic approach to selecting the best probability models for annual maximum rainfalls – A case study using data in Ontario (Canada), *Journal of Hydrology*, 553, 49-58, DOI: 10.1016/j.jhydrol.2017.07.052
- Norbiato D., Borga M., Esposti S.D., Gaume E., Anquetin S., 2008, Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins, *Journal of Hydrology*, 362 (3-4), 274-290, DOI: 10.1016/j.jhydrol.2008.08.023
- Re M., Barros V.R., 2009, Extreme rainfalls in SE South America, 2009, *Climatic Change*, 96 (1-2), 119-136, DOI: 10.1007/s10584-009-9619-x
- Roth M., Buishand T.A., Jongbloed G., Klein Tank A.M.G., van Zanten J.H., 2014, Projections of precipitation extremes based on a regional, non-stationary peaks-over-threshold approach: A case study for the Netherlands and north-western Germany, *Weather and Climate Extremes*, 4, 1-10, DOI: 10.1016/j.wace.2014.01.001
- Scarrott C., MacDonald A., 2012, A review of extreme value threshold estimation and uncertainty quantification, *REVSTAT – Statistical Journal*, 10 (1), 33-60
- Suligowski R., 2004, Temporal and spatial structure of precipitation in Poland. Regionalization attempt, (in Polish), *Prace Instytutu Geografii Akademii Świętokrzyskiej*, Kielce, 116 pp.
- Towler E., Rajagopalan B., Gilleland E.R., Summers S., Yates D., Katz R.W., 2010, Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory, *Water Resources Research*, 46 (11), DOI: 10.1029/2009WR008876
- Tramblay Y., Neppel L., Carreau J., Najib K., 2013, Non-stationary frequency analysis of heavy rainfall events in southern France, *Hydrological Sciences Journal*, 58 (2), 280-294, DOI: 10.1080/02626667.2012.754988
- Wdowikowski M., Kaźmierczak B., Ledvinka O., 2016, Maximum daily rainfall analysis at selected meteorological stations in the upper Lusatian Neisse River basin, *Meteorology Hydrology and Water Management – Research and Operational Applications*, 4 (1), 53-63, DOI: 10.26491/mhwm/63361
- Węglarczyk S., 2013, On the correctness of the Chomicz equations for design rainfall calculations, (in Polish), *Infrastruktura i Ekologia Terenów Wiejskich*, 3 (IV), 305-323
- WMO, 2011, *Climate observations, stations and networks*, [in:] *Guide to climatological practices*, WMO-No. 100, World Meteorological Organization, Geneva