

1 **Development of cluster analysis methodology for identification of model rainfall**  
2 **hyetographs and its application at an urban precipitation field scale**

3

4 Karol Mikołajewski<sup>1,2</sup>, Marek Ruman<sup>2,\*</sup>, Klaudia Kosek<sup>3,\*</sup>, Marcin Glixelli<sup>4</sup>, Paulina  
5 Dzimińska<sup>5</sup>, Piotr Ziętara<sup>4</sup>, Paweł Licznar<sup>5</sup>

6

7 <sup>1</sup>RETENCJAPL Sp. z o.o., Gdańsk 80-868, Poland; karol.mikolajewski@retencja.pl

8 <sup>2</sup>Faculty of Natural Sciences, University of Silesia in Katowice, Sosnowiec 41-200, Poland;

9 marek.ruman@us.edu.pl

10 <sup>3</sup>Faculty of Civil and Environmental Engineering, Gdansk University of Technology, Gdańsk 80-233, Poland;

11 klaudia.kosek@pg.edu.pl

12 <sup>4</sup>Krakow Water, Kraków 30-106, Poland; marcin.glixelli@wodociagi.krakow.pl (M.G.),

13 piotr.zietara@wodociagi.krakow.pl (P.Z.)

14 <sup>5</sup>Faculty of Environmental Engineering, Wrocław University of Science and Technology, Wrocław 50-377,

15 Poland; dziminska.paulina@gmail.com (P.D.), pawel.licznar@pwr.edu.pl (P.L)

16 \*Corresponding authors: marek.ruman@us.edu.pl (M.R.); klaudia.kosek@pg.edu.pl (K.K.)

17

18

19 **Abstract:** Despite growing access to precipitation time series records at a high temporal scale,  
20 in hydrology, and particularly urban hydrology, engineers still design and model drainage  
21 systems using scenarios of rainfall temporal distributions predefined by means of model  
22 hyetographs. This creates the need for the availability of credible statistical methods for the  
23 development and verification of already locally applied model hyetographs. The methodology  
24 development for identification of similar rainfall models is also important from the point of  
25 view of systems controlling stormwater runoff structure in real time, particularly those based  
26 on artificial intelligence. This paper presents a complete methodology of division of storm  
27 rainfalls sets into rainfalls clusters with similar temporal distributions, allowing for the final

28 identification of local model hyetographs clusters. The methodology is based on cluster  
29 analysis, including the hierarchical agglomeration method and k-means clustering. The  
30 innovativeness of the postulated methodology involves: the objectivization of clusters  
31 determination number based on the analysis of total within sum of squares (wss) and the  
32 Caliński and Harabasz Index (CHIndex), verification of the internal coherence and external  
33 isolation of clusters based on the bootmean parameter, and the designated clusters profiling.  
34 The methodology is demonstrated at a scale of a large urban precipitation field of Kraków city  
35 on a total set of 1806 storm rainfalls from 25 rain gauges. The obtained results confirm the  
36 usefulness and repeatability of the developed methodology regarding storm rainfall clusters  
37 division, and identification of model hyetographs in particular clusters, at a scale of an entire  
38 city. The applied methodology can be successfully transferred on a global scale and applied in  
39 large urban agglomerations around the world.

40

41 **Keywords:** precipitation modelling, storm rainfalls, cluster analysis, classification quality  
42 assessment indices

43

## 44 **1. Introduction**

45 Conducting calculations of non-stationary surface runoff of stormwaters and their  
46 further transformation in the stormwater system, including detention and retention in water  
47 bodies, requires the scenarios availability of rainfall temporal distribution. Precipitation  
48 scenarios employed by hydrological models usually determine the values of outflow volumes  
49 observed in the calculation results, the dynamics of their changes in time, and changes in the  
50 water volume retained in various elements of the drainage system. Due to historical conditions,  
51 and particularly calculation limitations, for decades, only the simplest models naturally

52 employing the simplest rainfall scenarios could be implemented in practice in hydrology at its  
53 various scales, and particularly in urban hydrology. An extreme example can be the rational  
54 model, still commonly used today in designing urban and road drainage systems. It employs  
55 block rainfall, i.e., a scenario of rainfall with a specified duration and constant intensity  
56 determined based on the local IDF (Intensity-Duration-Frequency) model for the adopted  
57 frequency of its occurrence (Kundzewicz and Licznar, 2021(2022)). The block model is  
58 obviously a very specific rainfall scenario. Unlike in that model, as observed for natural rainfall,  
59 hydrological analyses usually assume the necessity of taking into consideration the temporal  
60 variability of point rainfall intensity, usually described in the graphic form by means of  
61 hyetographs. Adopting hyetographs in design aimed at reflecting rainfall temporal distributions  
62 analogical to the temporal courses of locally recorded rainfalls, and implicitly also potential  
63 rainfalls that may occur in the future. This has encountered numerous obstacles since the very  
64 beginning, resulting not only from the stochastic (random) nature of rainfall, but also from the  
65 multifractal nature of rainfall still unknown at the time. It is currently already evidenced that  
66 even simple time series of rainfall records from single rain gauges have a multifractal structure  
67 escaping the principles of simple Euclidean geometry (de Lima, 1998; Deidda et al., 1999;  
68 Licznar, 2009). Due to this, and due to the limited availability of research material (sets of  
69 precise rainfall records at a high temporal resolution), the historical methodology of  
70 hyetographs representative development for modelling, customarily called model hyetographs,  
71 had to be based on generalisation and simplification, and assumptions that should be currently  
72 rejected or subject to suitable validation. All this creates the need for undertaking new research  
73 in the scope of identification of model hyetographs.

74 Interest in model hyetographs certainly increased with first attempts of transition from  
75 stationary methods of calculation of urban stormwater systems towards non-stationary methods  
76 of simulation of their operation during stormwater runoff. It became possible due to the



77 implementation of digital hydrodynamic models of sewage networks, such as e.g., programme  
78 SWMM (Storm Water Management Model) (Nix, 1994). Initially, however, due to the limited  
79 computing power of the available computers, focus was only placed on flows simulations in the  
80 underground canal system for single rainfalls. The model of stormwater runoff transformation  
81 in the sewage system was only combined with the hydrological surface runoff model, describing  
82 stormwaters inflow to network nodes (manholes and inlets) from the catchment area, and the  
83 stormwater system effect. The interaction of the stormwater system with the rainfall receiver  
84 was only possible to reflect by means of a suitable threshold condition on the network outlet.  
85 Rainfall introduced to the first generation of stormwater systems hydrodynamic models was a  
86 hyetograph of either a subjectively selected actual storm rainfall recorded by a pluviograph, or  
87 an artificial model hyetograph (often called model rainfall).

88         With rapid improvement of the computing power of computers, but without excessive  
89 complication of the algorithms of hydrodynamic models, it became possible to conduct  
90 simulations for entire series of local storm rainfalls. Their use enabled actual implementation  
91 of long-term simulations. After statistical processing, their final results could provide the basis  
92 for the probabilistic verification of the drainage system functioning in terms of stormwater  
93 system damming up frequency (Gires et al., 2012, 2013; Licznar et al., 2008), as well as e.g.,  
94 the necessary volume of retention reservoirs of stormwaters (Licznar, 2013). This new reality  
95 seemingly appeared to bring an end to the application of model hyetographs, at least in the area  
96 of urban hydrology. Moreover, with the installation of urban precipitation monitoring networks  
97 equipped with new generation electronic rain gauges, high temporal resolution rainfall series  
98 became much more available. In simulations of large stormwater systems, particularly in terms  
99 of development of their Real Time Control, records of rainfalls from entire networks of rain  
100 gauges began to be applied, or even spatial data from weather radars (Jakubiak et al., 2014).  
101 Afterwards at operational phase of RTC systems above mentioned observational rainfall data



102 sources are often coupled with numerical weather predictions. It allows proper short-term  
103 prediction of extreme rainfall events, indispensable for urban flood alert broadcasting and  
104 selection of best scenarios of next RTC strategies. Here, there is a vivid need for effective  
105 whether pattern recognition algorithms, based on machine learning technique, as for example  
106 support vector machine implemented by Nayak and Ghosh (2013) in case of Mumbai, India.  
107 Finally, in response to the stochastic nature of precipitation processes, it was determined that it  
108 is the most credible to supply hydrodynamic models with not so much historical precipitation  
109 data (i.e., implementations of local precipitation processes that already took place) as much  
110 richer synthetic data generated by means of local precipitation generators (Molnar et al., 2006).  
111 Owing to progress in the scope of multiplicative random cascades, it became possible to  
112 develop generators not only for generating rainfall series (1-D) (Güntner et al., 2001; Hingray  
113 and Ben Haha, 2005; Licznar et al., 2011a, 2011b), but also spatial data (2-D) (Over and Gupta,  
114 1994; Rupp et al., 2012), or even spatiotemporal data on rainfall (3-D) (Deidda, 2000).

115         Despite all the aforementioned conditions, the end of application of model hyetographs  
116 in urban hydrology has not arrived yet. To a certain extent this probably results from the  
117 conservatism of the engineers themselves, preferring design and drainage systems modelling  
118 based on simple and long-familiar model hyetographs. This facilitates their work e.g., in terms  
119 of time and equipment requirements necessary for conducting hydrodynamic simulations, and  
120 in accordance with the conservative provisions of rarely modified technical standards (Schmitt,  
121 2000). Paradoxically, however, the application of model hyetographs is currently not limited to  
122 simple engineering works. They prove useful in the most advanced consulting works. Such  
123 works employ a completely qualitatively new approach to modelling stormwater networks,  
124 called integrated modelling. The software used for the purpose is much more advanced, because  
125 it combines three models: the hydrological streams model (receivers of stormwaters), the  
126 hydrodynamic stormwater network model, and the hydrological surface runoff model. The first



127 two models are 1-dimensional, whereas the surface runoff model is a 2-dimensional model with  
128 an additional fill parameter. Naturally, the latter must be coupled with the digital terrain model  
129 (DTM). Although the application of integrated models brings numerous benefits, e.g., in the  
130 form of the possibility of tracing floodings on the surface of the DTM, it requires the use of  
131 complicated software launched on equipment with high computing power. Also in this case,  
132 simulations of larger drainage systems prove time-consuming and prone to numerical  
133 instabilities. As a result, simulation of runoff from the drainage system ceases to be feasible for  
134 tens or even hundreds of scenarios of storm rainfalls. In practice, complicated simulations on  
135 an integrated model can be only conducted for certain characteristic precipitation. This explains  
136 the return to the concept of the model hyetograph application or a narrow group of model  
137 hyetographs in the case of time-consuming simulations. It therefore remains an important issue  
138 to improve the methodology that could efficiently and objectively determine what model  
139 hyetographs would reflect local rainfall distributions variable in time in a satisfactory way.

140 In the case of the methodology modernisation of the model hyetographs development,  
141 it should be remembered that it began forming in the situation of strongly limited access to high  
142 temporal resolution rainfall records and statistical tools supporting their processes. Rainfall was  
143 recorded by means of simple rain gauges, e.g., pluviographs, and the records in the analogue  
144 form were difficult to process. Their processing usually involved a review of the records with  
145 designation of maximum rainfall with different durations and structure, and determination of  
146 the resulting empirical DDF (Depth-Duration-Frequency) or IDF (Intensity-Duration-  
147 Frequency) dependencies. Due to this, it was proposed to use the information contained in DDF  
148 and IDF dependencies as the starting point for constructing model hyetographs. Next to the  
149 rectangular shape of block rainfall mentioned in the introduction, different authors  
150 recommended simple modifications of the shape of hyetographs. For example, Sifald (1973),  
151 analysing the hydraulic operation of the stormwater network, postulated a hyetograph with a

152 trapeze shape. At the scale of small catchments, Yen and Chow (1980) proposed transformation  
153 of rainfall depth read for the predefined duration and frequency from the DDF model into a  
154 triangular hyetograph. Probably due to the common use of the unit hydrograph method in the  
155 contemporary hydrology aimed at providing an estimate of direct runoff hydrographs resulting  
156 from given excess rainfall hyetograph, hyetographs with possibly simple shapes were eagerly  
157 used, composed of regular geometric figures. For example, in urban hydrology, Desbordes  
158 (1978) implemented the application of a hyetograph composed of “three” triangles. Peyron et  
159 al. (2002) attempted to simplify the shape through replacing two triangles representing the start  
160 and end impulses with rectangular courses. At the same time, in the conditions of Taiwan, Lee  
161 and Ho (2002) postulated the application of a model hyetograph with a shape built from two  
162 triangles. Even earlier, independently from the aforementioned papers, a concept appeared for  
163 the structure of the model hyetograph to be completely based on the reading of the entire IDF  
164 curve. The most classic example of the approach can be the continuous hyetograph for the city  
165 of Chicago developed by Keifer and Chu (1957). Another representative of this concept can be  
166 Euler type II model hyetograph (Schmitt, 2000), frequently encountered in the hydrodynamic  
167 modelling practice in Poland and Germany. The underlying idea of this type of hyetographs  
168 structures assumes a single artificial rainfall scenario including maximum point intensities for  
169 the entire hierarchy of partial times. It appears to be at variance with the observation of nature,  
170 where all maximums of point intensities are usually not recorded in a single rainfall at a  
171 specified level of frequency for durations lower and equal to total duration. Nonetheless, relying  
172 on the Euler type II hyetograph theoretically offers engineers the possibility to test the operation  
173 of the drainage system in a single simulation for all maximum point intensities simultaneously  
174 at a specified level of frequency of occurrence. Engineers are accustomed to using IDF curves,  
175 and their transformation into a Euler type II model hyetograph requires only simple algebra  
176 operations. In the case of analyses of larger drainage systems, it is even simpler for engineers



177 in Poland and Germany to reach for a hyetograph recommended in the guidelines of DVWK  
178 (1984). Its development is based on the distribution of rainfall depth read from the DDF model  
179 to three rectangles with different intensity levels.

180 With time, the methodology of model hyetographs determination could be improved  
181 owing to the availability of increasingly richer measurement data from rain gauge networks. It  
182 became feasible to determine model hyetographs based on the large rainfall datasets analysis.  
183 The precursor of such an approach was Huff (1967), who analysed data from 49 rain gauges  
184 and 12 years (1955-1966) from the state of Illinois in the USA. Because the rainfalls retrieved  
185 from the records had different total durations and depths, he proposed normalisation of  
186 hyetographs, and application of dimensionless hyetographs. This permitted comparison of a  
187 dataset of 261 rainfalls that Huff (1967) classified according to whether the greatest percentage  
188 of cumulative rainfall occurred in the first, second, third, or fourth quarter of the storm duration.  
189 The resulting model Huff mass curves found numerous applications in hydrology. The  
190 analogical methodology of temporal variability analysis of rainfall distributions was not only  
191 repeated in further research by Huff himself (1970, 1990), but also by many other scientists,  
192 e.g., Pani and Haragan (1981), Bonta and Rao (1987), Bonta (2004), Terranova and Iaquina  
193 (2011), Elfeki et al. (2014), Pan et al. (2017). Hydrologists in the USA also commonly apply  
194 model hyetographs defined by means of dimensionless mass curves recommended by the SCS  
195 (Soil Conservation Service) (McCuen, 1986). The increasingly rapidly growing digital  
196 databases of rainfalls records from electronic rain gauges, and the progress in the scope of data  
197 mining techniques nowadays allows for the continuation of the trend determined by Huff  
198 (1967), and model hyetographs determination based on the actual local rainfall analysis.

199 Access to a large data base from 25 electronic weight rain gauges from the municipal  
200 precipitation monitoring network of Warsaw (Poland) has become the impulse for the  
201 verification whether the Euler type II model hyetograph applied in practice, developed based



202 on the IDF curve, corresponds with temporal distributions of actual rainfalls. For this purpose,  
203 out of approximately 20-year precipitation series, Licznar and Szeląg (2014) desinated a total  
204 of 669 storm rainfalls. The set was then divided into subsets with increasing durations,  
205 expressed in minutes: [0-45], (45-60], (60-90], (90-120], (120-180], (180-240], (240-300],  
206 (300-360], (360-420]. Each of the subsets was moreover supplemented by an additional Euler  
207 precipitation model (type II) developed based on the local IDF curve. All subsets were then  
208 analysed with the application of the hierarchical agglomeration method. Based on the obtained  
209 dendrograms, Licznar and Szeląg (2014) observed that precipitation recorded in Warsaw, even  
210 those with approximate durations, have evidently differing hyetographs. Moreover, in each of  
211 the subsets, Euler model precipitation (type II) was an extreme outlier in the structure of the  
212 dendrogram with the highest bond distance towards all actual rainfalls. Results of the study  
213 have become an impulse for the search for a method of designation of more representative local  
214 model hyetographs. For this purpose, Licznar et al. (2017) proposed the application of cluster  
215 analysis tools in the form of not only the hierarchical agglomeration method, but also non-  
216 hierarchical k-means clustering. The methodology also found applications for example in  
217 research by Licznar (2018) and Wartalska et al. (2020). Nonetheless, it leaves evident gaps. Its  
218 greatest weakness is completely subjective a priori adoption of number  $k$  of clusters, i.e., the  
219 number of final rainfall models. Two further missing components of the methodology include  
220 the objective assessment of the classification results and cluster profiling. They are increasingly  
221 important, because in the context of new challenges in urban hydrology, striving for ‘smart city’  
222 solutions involves not so much searching for model hyetographs themselves as the possibility  
223 of fast and efficient search of similar temporal models of rainfalls. Runoff control systems  
224 employing artificial intelligence aim at the implementation of the most effective strategy of the  
225 forecasted scenario control of stormwater runoff, adopted based on the already implemented in  
226 nature and recorded precipitation phenomenon and runoff caused by rainfall with possibly



227 similar rainfall temporal distribution in reference to local rainfall forecast from the nowcasting  
228 system. The final question to be answered from the engineering practice point of view is: to  
229 what extent the application of the cluster analysis brings results repeatable at the natural spatial  
230 scale of extensive municipal drainage systems? Can for example model hyetographs developed  
231 based on rainfall series from nearby rain gauges be treated as credible for analyses of  
232 precipitation-runoff phenomena in the territory of the entire city?

233         Considering the overview of the methodology state of rainfall classification in terms of  
234 its temporal distribution and model hyetographs identification, the primary objective of this  
235 paper is to present the complete cluster analysis methodology supplemented by objective  
236 determination of the number of classes, credible assessment of the classification results, and  
237 cluster profiling. Another primary objective of the paper is to demonstrate the postulated  
238 complete cluster analysis methodology at the scale of a large urban precipitation field, and the  
239 resulting answer to the question whether model hyetographs retrieved from records of different  
240 rain gauges show mutual compatibility.

241         The pragmatic objective of the study is the complete methodology development of  
242 designation of local model hyetographs throughout Poland. In the years 2016-2020, the project  
243 of the Polish Atlas of Rains Intensities PANDa was implemented, resulting in a digital base of  
244 rainfalls at high temporal resolution (after 30 years of observation from 100 stations), and the  
245 national rainfall atlas composed of 12885 local IDF models ascribed to areas designated through  
246 the division of Poland with a grid with field dimensions of 5 km per 5 km (Burszta et al.; 2019;  
247 Licznar et al., 2020). The PANDa atlas provides the basis for the publicly accessible digital  
248 design platform [www.waterfolder.com](http://www.waterfolder.com), where you can design and select among others:  
249 retention reservoirs of stormwaters, infiltration reservoirs for stormwaters, linear drainage units,  
250 gravitational canals of stormwater systems, stormwater pumps and pumping stations, and green  
251 roof surfaces. In the scope of the new WaterFolder Connect project, works are currently

252 undertaken aimed at the integration of tools of selection and enabling hydrodynamic  
253 simulations of newly designed drainage systems. This will require the designation of local  
254 hyetographs for particular areas of Poland. The developed methodology presented in this article  
255 is planned to be applied at the scale of the entire country, and then globally.

## 256 **2. Materials and methods**

### 257 2.1. Study area

258 Research on objective classification of storm rainfall hyetographs by means of  
259 classification quality assessment indices was conducted at a large scale of an urban precipitation  
260 field in the territory of Kraków. Kraków is the second largest city in Poland in terms of  
261 population (~ 767 000) and surface area (327 km<sup>2</sup>), located in the south of the country on the  
262 Vistula River. The study employed part of the resources of the Polish precipitation data base of  
263 the Polish Atlas of Rains Intensities (PANDa) project, and records from the local rain gauge  
264 network of the Municipal Water Supply and Sewerage Company (MWSSC) in Kraków.

265 At the initial stage of the PANDa project in the years 2016-2017, a digital base of rainfall  
266 series was developed for a total number of 100 rain gauges in Poland. Then it was analysed in  
267 terms of occurrence of maximum rainfall intensity (Burszta-Adamiak et al., 2019). The base  
268 included among others records from a rain gauge installed in the area of the Kraków–Balice  
269 airport (50°04'40", 19°47'42") at a height of 237 m a.s.l., and rain gauge in station Kraków–  
270 Wola Justowska (50°03'50", 19°53'25") at a height of 204 m a.s.l. (Figure 1). For station  
271 Kraków–Balice, records from the multiannual period 1986-2006 were available, and for station  
272 Kraków–Wola Justowska records from the multiannual period 2007-2015. In that set, digital  
273 record series from the years 1986-1998 resulted from the digitalisation of pluviograph recording  
274 strips. The computer-aided method of conversion of pluviograph recording strips to digital  
275 format similar to that proposed by Licznar et al. (2011a) was adopted. Records from a standard

276 unheated pluviograph (200 cm<sup>2</sup> orifice) covered warm year periods between spring (April or  
277 May) and autumn (October or November) when most of storm rainfall events occur in Poland.  
278 For later years, i.e., for the multiannual period 1999-2015, all-year records from electronic  
279 tipping bucket gauges were already available. On station Kraków–Balice, rainfall was recorded  
280 by an electronic rain gauge Aster TPG, and on station Kraków–Wola Justowska, by an  
281 electronic rain gauge Met One Instruments 60030. The resolution of recording rainfall depth in  
282 the case of the aforementioned devices was 0.1 mm and 0.2 mm, respectively, and their inlet  
283 surface was 200 cm<sup>2</sup>. Records of local rainfall provided for the research by the MWSSC came  
284 from a network of a total of 23 rain gauges distributed throughout the city (Figure 1).

285

286 **Figure 1.** Location of rain gauges belonging to Waterworks Kraków and the Institute of  
287 Meteorology and Water Management

288

289 The entire network was composed of electronic tipping bucket gauges operating during  
290 both the summer and winter half-year. In all stations, rain gauges Hobo RG3M were installed,  
291 with rainfall record resolution of 0.2 mm and inlet surface of 200 cm<sup>2</sup>. Unfortunately, records  
292 of rainfall series from these rain gauges were considerably shorter than in the case of the data  
293 base of the PANDa project, covering a period of 30 years (1986-2015). Detailed information  
294 regarding periods of rainfall records by particular rain gauges of the measurement network in  
295 Kraków is provided in Table 1 in the chapter discussing results obtained in the scope of  
296 designation of storm rainfalls. It should be emphasised that all rain gauges used in the analysis  
297 were located within the administrative boundaries of the city. The only exception was rain  
298 gauge Kraków–Balice located in the direct vicinity of the city boundaries, in the area of the  
299 nearby airport. All rainfall series from the entire period of monitoring of the precipitation field,  
300 recorded directly in digital form by tipping bucket gauges, as well as those resulting from

301 digitalisation of records on pluviograph record strips, had a uniform temporal resolution of 1  
302 minute.

303

## 304 2.2. Applied methodology

305         Based on the available time series of rainfall records (from the rain gauges of the  
306 PANDa project and from the network of MWSSC), storm rainfalls were designated by standard  
307 criteria proposed by Schmitt (2000) for identifying storms for urban drainage systems  
308 modelling. They are standard criteria applied in Germany and Poland, and used in the already  
309 cited paper by Licznar et al. (2011a). The adopted threshold minimum value of the total amount  
310 of storm rainfall was 10 mm, and the minimum time interval between single rainfall events was  
311 at least 4 hours. The designation of rainfalls in reference to dry periods also employed the  
312 minimum value of rainfall depth of 0.1 mm during 5 minutes as the threshold value for the  
313 interval to be considered as part of a rainfall event in terms of duration and precipitation amount.  
314 The result of designation of storm rainfalls separately from rain gauges of the PANDa project  
315 and the MWSSC network were data sets called set No. 1 and No. 2, respectively. Set No. 3  
316 analysed in the final part of the research constituted a combination of sets No. 1 and No. 2.

317         Due to the differing durations and total depths of the designated storm rainfalls, further  
318 analysis of the sets, i.e., mutual comparison and identification of typical (quantifiable, model)  
319 distributions during rainfalls, involved double normalisation of cumulative hyetographs. This  
320 procedure was conducted in accordance with the methodology described in the publication  
321 (Licznar et al., 2017). It corresponded with the methodology of development of dimensionless  
322 hyetographs by Huff (1967). For this purpose, for each storm rainfall, total duration and total  
323 rainfall depth was determined. Next, each cumulative storm rainfall hyetograph with known  
324 total duration was divided into 100 even time intervals. For each of the subsequent intervals, a  
325 corresponding cumulative rainfall increase was determined. Subsequent cumulative

326 precipitation increases were divided by total rainfall depth, obtaining unitary cumulative  
327 precipitation increases. As a result, the shape of each of the storm rainfalls was reflected by a  
328 hyetograph normalised to a range from 0 to 100% for duration and to a range from 0 to 1 (100%)  
329 for rainfall depth.

330 The mutual comparison of shapes of normalised (dimensionless) hyetographs within the  
331 analysed sets, and further designation of typical storm hyetographs that could be considered  
332 model hyetographs, useful e.g., for modelling of local drainage systems, employed tools for  
333 mining large data sets in the form of cluster analysis algorithms. Unlike in the case of earlier  
334 attempts to apply cluster analysis for the classification of storm rainfall hyetographs and for  
335 identification of model hyetographs (Licznar et al. 2017; Licznar, 2018; Wartalska et al., 2020),  
336 this study applied a complete cluster analysis methodology. The complete cluster analysis  
337 methodology involves seven stages (Milligan, 1996; Zhou et al., 2014):

- 338 1) Selection of objects and variables;
- 339 2) Selection of the formula of normalisation of variable values;
- 340 3) Selection of distance measure;
- 341 4) Selection of classification method;
- 342 5) Determination of the number of classes;
- 343 6) Assessment of classification results;
- 344 7) Class description (interpretation) and profiling.

345 The implementation of the first two points of the methodology, i.e., selection of objects and  
346 their variables, combined with normalisation of their values, was already characterised earlier,  
347 and aimed at the development of sets of normalised (dimensionless) hyetographs. Two further  
348 stages 3 and 4, covering the selection of classification methods and distance measures, were  
349 analogical to those in the already published papers by Licznar et al., 2017, Licznar, 2018,  
350 Wartalska et al., 2020. In comparison to these publications, three last missing stages of the

351 cluster analysis were added in this paper, including: objective determination of the number of  
352 classes, and assessment of classification results, combined with simplified class profiling. This  
353 resulted in a coherent research methodology allowing for the classification of storm rainfall  
354 hyetographs, their division into the objectively determined number of clusters, determination  
355 of courses of model hyetographs, and their simplified profiling.

356 Cluster analysis tools find broad practical application in collating large data sets. Their  
357 implementation in the case of such sets permits the separation of their objects into a certain  
358 number of subsets, called clusters, covering mutually similar objects. The requirement of  
359 decouplability and sufficiency of the designated clusters is met, i.e., each of the elements  
360 belongs to a specific single cluster (Larose, 2005; Stanisiz, 2007). Therefore, the sum of all  
361 clusters corresponds with the initial large set of objects, and particular clusters are separate and  
362 have no elements in common. In each cluster, objects are approximate, mutually similar, and  
363 simultaneously different from objects in other clusters. Depending on the adopted method, the  
364 division into clusters can be conducted to an *a priori* determined or undetermined number of  
365 clusters. In research in the Kraków polygon, cluster analysis was applied to the division of sets  
366 No. 1, No. 2, and No. 3 of dimensionless cumulative hyetographs of storm rainfalls into an  
367 undetermined, and then determined number of clusters. For this purpose, the hierarchical  
368 agglomeration method and non-hierarchical k-means clustering method were implemented.

369 Agglomeration methods have already found application in research on precipitation, not  
370 only in the context of search for similarities in temporal distribution of storm rainfalls for  
371 particular locations (Licznar et al., 2017; Wartalska et al., 2020), but also in the case of research  
372 on the variability of precipitation conditions within a large municipal precipitation monitoring  
373 network (Licznar et al., 2015). In their research conducted on the municipal monitoring network  
374 in Warsaw (including 25 rain gauges), comparable to that in Kraków in terms of size, Licznar  
375 et al. (2015) successfully implemented agglomeration methods to evidence similarities of



376 empirical distributions of breakdown coefficients (BDCs) of rainfall from rain gauges located  
377 in different points in the city, for hierarchies of timescales, corresponding with time from 5 min  
378 to 1280 min. Agglomeration methods applied in this type of research aim at combining  
379 mutually similar objects through the application of appropriate measures of their mutual  
380 distance and the agglomeration method. The starting point is treating each object as a separate  
381 cluster. At the subsequent stages of agglomeration, objects most mutually approximate  
382 according to the defined measure (i.e., most similar by default) are combined into new clusters,  
383 covering objects and clusters resulting from earlier stages, until a single cluster is obtained.  
384 Based on the experience of other authors (Licznar et al., 2017; Wartalska et al., 2020), in  
385 research in the Kraków polygon, the already verified distance metrics were applied: Euclidean  
386 and Euclidean squared distance. The Euclidean distance metric  $d(x,y)$  for two objects  $x = (x_1,$   
387  $\dots, x_r)$  and  $y = (y_1, \dots, y_r)$ , characterised by  $r$  measurement values, is expressed in the following  
388 formula (Larose 2005, Stanisiz 2007):

$$389 \quad d(x, y) = \sqrt{\sum_{i=1}^r (x_i - y_i)^2}. \quad (1)$$

390 The Euclidean distance metric has simple and natural interpretation in the case of objects  
391 defined by only two or three measurement values ( $r = 2$  or  $r = 3$ ), because in that case its  
392 equivalent is distance on a plane and in space of two points  $x$  and  $y$ . Through analogy, in the  
393 case of the analysed sets of normalised hyetographs, the Euclidean distance between pairs of  
394 hyetographs was determined in space with a considerably higher number of dimensions (for  $r$   
395 = 100).

396 Next to adopting a particular mutual distance metric for objects in the measurement  
397 space, the objective of combining mutually similar objects and clusters also requires the  
398 application of a particular method of their agglomeration. In this case, also based on the already  
399 cited papers (Licznar et al., 2017; Wartalska et al., 2020), the popular unweighted pair-group  
400 method was applied. In this method, differences in distances between all pairs of elements



401 included to particular clusters are calculated. Values of the averaged differences between all  
402 pairs are adopted as the measure of distance between particular clusters. Due to this, it is known  
403 which elements of the sets are mutually similar and can be included to shared clusters, and  
404 moreover to what extent particular clusters are mutually similar and can be agglomerated into  
405 structures of larger clusters. A natural consequence of this is forming on the dendrograms (the  
406 resulting diagrams of the set structure in relation to the increasing bond distance, and therefore  
407 decreasing similarity between its elements) characteristic ‘chains’ made of similar objects  
408 developing increasingly extensive clusters.

409         Considering the primary study objective, i.e., the determination of model hietographs,  
410 a more significant research tool was k-mean clustering. The application of this tool permitted  
411 separation of the analysed sets into  $k$  independent clusters differing to the greatest possible  
412 degree. In the case of earlier research on model hietographs (Licznar et al., 2017, Wartalska et  
413 al., 2020), the application of the k-means method always involved questionable adopting of  $k$   
414 number of clusters subjectively estimated based on the analysis of previously prepared  
415 dendrograms. It was assumed that at a certain level of bond distance, the chains of clusters can  
416 be cut to obtain several separate subsets of mutually similar dimensionless hietographs. The  
417 obvious weakness of such an approach was lack of justification of the adopted level of bond  
418 distance at which the dendrogram was divided (cut). Moreover, as not observed in earlier papers  
419 (Licznar et al., 2017; Wartalska et al., 2020), simple cut-off of dendrograms at a given level of  
420 and distance and obtaining  $k$  clusters was not equivalent to the determination of the same  
421 number  $k$  of independent clusters by means of the k-means clustering method. Particular  
422 clusters could include differing subsets and objects as a result of differences in the classification  
423 methods between the hierarchical agglomeration method and non-hierarchical method, namely  
424 k-means clustering.



425 Unlike in the case of the hierarchical agglomeration method, aimed at combining objects  
426 and subclusters, agglomeration by means of the k-means clustering method aims at fragmenting  
427 the entire set into the *a priori* defined number  $k$  of clusters, whereas none of the  $k$  clusters is the  
428 subcluster of another cluster. The algorithm of agglomeration by means of the k-means  
429 clustering method therefore involves development of  $k$  subclusters, and then moving objects  
430 across them for distances between them within the subclusters to be as small as possible, and  
431 for distances between subclusters to be as large as possible. The moving procedure is repeated  
432 iteratively aiming at the most efficient separation of clusters (Larose, 2005). The final objective  
433 is arriving at a solution in which the designated clusters meet two criteria: that of internal  
434 coherence and external isolation (Gordon, 1999). This task was implemented in the computing  
435 environment of language R due to the resulting calculation difficulties. Their primary source  
436 was the number of classified objects. The number of all divisions of a set of  $n$  elements into  $k$   
437 non-empty clusters is escribed by the following formula (Everitt et al., 2001; Gordon, 1999):

$$438 \quad L(n, k) = \frac{1}{k!} \sum_{s=1}^k (-1)^{k-s} \binom{k}{s} s^n, \quad (2)$$

439 where  $s$  is the number of class ( $s = 1, \dots, k$ ). In accordance with formula (2), even in the case of  
440 a very small set of 10 objects with their division into 4 non-empty clusters, the number of all  
441 possible divisions is 34 105 ( $L(10, 4) = 34105$ ). For comparison, research has involved division  
442 of large sets of hyetographs including hundreds or even thousands of storm rainfalls.

443 These calculation challenges were further considerably multiplied during analyses of  
444 measurement sets No. 1, 2, and 3 due to multiple launching of the clustering algorithm with the  
445 application of the bootstrap method. In statistics, bootstrap methods are used to estimate the  
446 distribution of estimation errors by means of multiple sampling with replacement. Their  
447 implementation in the case of research on storm hyetographs from Kraków meant that the  
448 clustering algorithm was performed 150 times, each time for random samples from the analysed  
449 sets. Results obtained in subsequent iterations were compared, allowing for the designation of

450 values of the bootmean parameter. The bootmean parameter was calculated as a mean value of  
451 the Jaccard index (Jaccard similarity coefficient) for each of the designated clusters. The  
452 Jaccard coefficient itself measures similarity between two sets, and is determined as the ratio  
453 of power set of the intersection of sets and power set of sum of these sets. High values of the  
454 Jaccard coefficient approximate to 1 strongly suggest perfect repeatability of the separation of  
455 objects into clusters. It is assumed that exceeding the threshold of 0.6 for Jaccard coefficients  
456 for each of the clusters suggests no occurrence among the designated clusters of clusters with  
457 random character, i.e., those including rainfall models deviating from the remaining clusters,  
458 but simultaneously not mutually similar. The verification of the above criterion allows drawing  
459 conclusions regarding the resulting clusters meeting the criteria of internal coherence and  
460 external isolation (Gordon, 1999). The bootmean parameter was used in the study for  
461 conducting assessment of results of the classification of particular sets No. 1, 2, and 3.

462 The compute-intensive process of k-means clustering was preceded by the stage of  
463 determination of the number of classes. For this purpose, a methodology was adopted analogical  
464 to that applied in research on temporal distributions of hourly water uptakes by Dzimińska et  
465 al. (2021). The analysis of previously prepared dendrograms provided the basis for the  
466 determination of a potential range of the number of clusters that should be considered in the  
467 case of division of storm rainfalls in Kraków. The optimal number of clusters was designated  
468 from that range based on the analysis of the total within sum of squares (wss) and values of the  
469 Caliński and Harabasz Index (CHIndex) for a variable number of clusters. Values of total within  
470 sum of squares (wss) and the Caliński and Harabasz Index (CHIndex) were calculated in  
471 accordance with the following formulas (Walesiak and Gatnar, 2009):

$$472 \quad wss = \sum_i^k \sum_{x \in C_i} \|x - m_i\|^2, \quad (3)$$

$$473 \quad CHIndex = \frac{SS_B}{SS_W} \cdot \frac{N-k}{k-1}, \quad (4)$$

474 where:  $k$  – number of clusters,  $x$  – element of a set,  $C_i$  –  $i$ -th data cluster,  $m_i$  – cluster centroid  $i$ ,  
475  $\|x - m_i\|^2$  - Euclidean distance between two vectors,  $N$  – total number of observations  
476 (elements of a set),  $SS_B$  – total variance between clusters (trace of interclass covariance matrix),  
477  $SS_W$  – total internal cluster variance (trace of intraclass covariance matrix). The value of  $w_{SS}$   
478 naturally decreases with an increase in the number of clusters  $k$ . Nonetheless, the gradient of  
479 the decrease evidently decreases after reaching a certain number of clusters considered optimal.  
480 In the case of the Caliński and Harabasz Index, the number of clusters is searched, considered  
481 optimal, for which its value is the highest. In accordance with formula (4), this means the  
482 occurrence of maximisation of the ratio of  $SS_B$  and  $SS_W$ , i.e., particular clusters differ from one  
483 another very significantly, and the elements of the set agglomerated in particular clusters are  
484 strongly mutually similar (relatively weakly variable).

485 The last seventh stage of the research covered the description (interpretation) and  
486 profiling of the obtained clusters. For all the clusters, their centroids were determined  
487 (arithmetic averages calculated from the original values of each variable based on objects  
488 developing a given cluster). This provided the basis for obtaining sets of normalised  
489 hyetographs for each of the analysed sets No. 1, 2, and 3. They were subject to mutual  
490 comparison. Finally, simplified profiling of clusters was conducted. The objective of cluster  
491 profiling is the identification of characteristic features of particular clusters allowing for the  
492 determination of differences between them. Cluster profiling is conducted based on variables  
493 that did not take part in the process of classification of the set of objects. In the case of research  
494 on hyetographs of storm rainfalls from Kraków, the available variables that did not participate  
495 in the process of classification of the set of objects were total rainfall depths and total rainfall  
496 durations. These general rainfall characteristics permitted the determination of distributions of  
497 mean precipitation intensities in particular clusters, i.e., simplified profiling of clusters in terms



498 of explanation of differences in the obtained types of model hyetographs from the point of view  
499 of intensities of the represented storms.

500

### 501 **3. Results and discussion**

502           Based on the adopted criteria (Schmitt, 2000), from set No. 1, covering time series of  
503 precipitation records from 30 years for rain gauges of the PANDa project, a total of 313 storm  
504 rainfalls were designated (Table 1), corresponding to the frequency of their occurrence at a level  
505 of approximately 10 storms per year. Analogically, from the larger set No. 2 of precipitation  
506 series from 126 years, recorded on 23 rain gauges of the MWSSC network, a total of 1494  
507 storm rainfalls were designated (Table 1), corresponding to the frequency of their occurrence  
508 at a level of approximately 12 storms annually. The obtained quantities of data sets were in  
509 accordance with the expectations and results of previously published research from the territory  
510 of Poland. For example, from a set of digitalised pluviography storm records from 38 years  
511 from Wrocław (south-western Poland), Licznar et al. (2011a) designated 250 storm rainfalls  
512 based on the same criteria, determining their frequency of occurrence at a level of 6.6 times per  
513 year. The lower precipitation frequency results from the fact that at each stage, precipitation  
514 records covered only periods of 5-6 months with positive air temperatures that allowed for the  
515 exposure of non-heated pluviographs. For comparison, in the polygon of the rain gauge network  
516 of the city of Warsaw composed of 25 electronic weighing rain gauges recording precipitation  
517 throughout the year, the criteria proposed by Schmitt (2000) permitted the designation of a total  
518 of 669 storm rainfalls (Licznar and Szelağ, 2014). The latter figure, in the context of records  
519 somewhat longer than two years (114 weeks) for each of the rain gauges, translates into the  
520 frequency of occurrence of storms equal to 12.3 events per year per single observation point.

521

522 **Table 1.** List of rain gauges belonging to sets No. 1 and 2 with characteristics of designated  
523 storms

524

525 Particular rainfalls differed in their courses in time. It is evident based on the example  
526 of cumulative normalised (dimensionless) hyetographs of set No. 1 in figure 2. A large majority  
527 of rainfalls, both in the case of set No. 1 and No. 2, had hyetographs considerably differing  
528 from theoretical hyetographs recommended for application as rainfall scenarios for  
529 hydrodynamic modelling of municipal drainage systems, such as: block rainfall, model rainfall  
530 according to DVWK, or model rainfall according to Euler type II. This observation also remains  
531 in accordance with research of other authors who already previously questioned the justification  
532 of drainage systems modelling in Poland commonly being based on model rainfall according  
533 to Euler type II (Licznar and Szeląg, 2014; Licznar et al., 2017).

534

535 **Figure 2.** Cumulative dimensionless hyetographs of 313 storm rainfalls from Kraków (set  
536 No. 1)

537

538 The designated storm rainfalls in particular sets No. 1 and 2 also differed in terms of  
539 total durations and total rainfall depths. In set No. 1, total depths of storm rainfalls were in a  
540 range from 10.0 to 105.0 mm, and their durations varied from 24 to 2329 minutes. These  
541 parameters in set No. 2 varied within similar ranges from 10.0 to 145.4 mm and from 14 to  
542 4576 minutes in the case of total depths and total durations, respectively. These values also  
543 showed similar distributions in both analysed sets, as confirmed by histograms included in  
544 figure 3. A vast majority of storm rainfalls was characterised by total rainfall depth in a range  
545 from 10 mm to 20 mm, and their total durations varied from 2 h to 12 h. The similarity of total  
546 distributions of rainfall depths and durations in sets No. 1 and No. 2 presented in figure 3 at

547 least partially justifies the possibility of use of sets of storm rainfalls recorded in stations in city  
548 outskirts and belonging to the national precipitation monitoring network in the context of issues  
549 regarding urban hydrology at a scale of a city as large as Kraków.

550

551 **Figure 3.** Histograms of rainfall depths and durations for the designated sets of rainfalls from  
552 Kraków, diagrams in the top row for set No. 1 (313 rainfalls), diagrams in the bottom row for  
553 set No. 2 (1493 rainfalls)

554

555 The analysed sets No. 1 and No. 2 of normalised cumulative precipitation were used for  
556 preparing dendrograms presented in figures 4a and 4b, respectively. As demonstrated based on  
557 the example of the least abundant set No. 1 in figure 4a, even in the case of number of objects  
558 only slightly exceeding 300, their abundance makes the prepared detailed dendrogram largely  
559 illegible. Due to this, the same dendrogram and all remaining dendrograms for more abundant  
560 sets were prepared for a number of leaf nodes reduced to 30 (Figures 4a, 4b, 5a, and 5b),  
561 considerably improving their legibility. In practice, this measure corresponded with cut-off of  
562 detailed dendrograms at a level of bond distance of approximately 1.0. On the prepared  
563 dendrograms in figures 4a and 4b, it is easy to recognise the characteristic structures in the form  
564 of chains connecting similar objects or subclusters of objects occurring for lower bond  
565 distances. The obtained image of dendrograms therefore corresponded in terms of quality to  
566 dendrograms published in earlier papers regarding the application of agglomeration methods in  
567 the analysis of precipitation hyetographs (Licznar and Szeląg, 2014; Licznar et al., 2017,  
568 Licznar, 2018; Wartalska et al., 2020). In the case of figure 4b, notice that its structure included  
569 a cluster with number 29 with extremely high bond distance (approximately 5.5). Detailed  
570 analysis of cluster No. 29 in figure 4b showed that it is composed of only one object.

571

572 **Figure 4 a.** Dendrograms obtained for set No. 1 composed of 313 dimensionless cumulative  
573 rainfall hyetographs from Kraków (top panel). Below find the same dendrogram prepared for  
574 the reduced number of 30 leaf nodes (bottom panel). In the diagrams, vertical axes show bond  
575 distances for particular rainfalls and rainfall clusters. The horizontal axis of the dendrogram on  
576 the bottom panel shows numbers of rainfalls in particular clusters; **b.** Dendrogram obtained for  
577 set No. 2 composed of 1494 dimensionless cumulative rainfall hyetographs from Kraków. The  
578 dendrogram was prepared for the reduced number of 30 leaf nodes. The vertical axis of the  
579 diagram shows bond distances for particular rainfall clusters, and the horizontal axis shows  
580 their numbers

581

582 **Figure 5 a.** Dendrogram obtained for adjusted set No. 2 composed of 1493 dimensionless  
583 cumulative rainfall hyetographs from Kraków. The dendrogram was prepared for the reduced  
584 number of 30 leaf nodes. The vertical axis of the diagram shows bond distances for particular  
585 rainfall clusters, and the horizontal axis shows their numbers; **b.** Dendrogram obtained for  
586 adjusted set No. 3 composed of 1806 dimensionless cumulative rainfall hyetographs from  
587 Kraków. The dendrogram was prepared for the reduced number of 30 leaf nodes. The vertical  
588 axis of the diagram shows bond distances for particular rainfall clusters, and the horizontal axis  
589 shows their numbers

590

591 The object was a hyetograph of a storm rainfall designated from records of rain gauge  
592 No. 12 in Płaszów in 2014 (Figure 6). The rainfall had a total depth of 13.2 mm and total  
593 duration of 507 min. Although the rainfall event formally meets the adopted criteria of storm  
594 rainfall (Schmitt, 2000), it showed very specific course in time (Figure 6). For the first 460 min.  
595 of the rainfall, only 0.8 mm of rain was recorded. In the second, considerably shorter, final part  
596 of the rainfall lasting 47 min, 12.4 mm of rain was recorded. Considering this very specific



597 course of the rainfall in time, it was excluded from set No. 2, and the analysis was continued  
598 for a set of 1493 normalised hyetographs of the remaining rainfall events.

599

600 **Figure 6.** Hyetograph of a rainfall event recorded on station Kraków-Płaszów in 2014

601

602 For the adjusted set No. 2 including 1493 objects, the dendrogram presented in figure 5a  
603 was obtained. As a result of elimination of outliers in the form of a single rainfall event from  
604 rain gauge No. 12, a dendrogram was obtained with a maximum bond distance at a level  
605 somewhat higher than 4.0, comparable like in the case of set No. 1 (Figure 3). Importantly, after  
606 the adjustment, the most deviating cluster with number 24 was not composed of a single object,  
607 but covered normalised hyetographs for 9 different storm rainfalls (combined into a single  
608 cluster for bond distances lower than 1.0). At the final stage of application of hierarchical  
609 methods for analysing structures of sets of normalised hyetographs, a dendrogram was also  
610 prepared for set No. 3, constituting a combination of sets No. 1 and No. 2, presented in  
611 figure 5b. The dendrogram obtained in the case of set No. 3, prepared for 1806 dimensionless  
612 storm rainfall hyetographs, had a structure analogical to that of the already discussed  
613 dendrograms in figures 4a and 5a. Its maximum bond distance did not exceed 4.0, and it showed  
614 characteristic chain connections of similar subclusters of objects, mutually connecting for bond  
615 distances of less than 1.0. Irrespective of the similarities indicated herein, the comparison of  
616 three dendrograms obtained for sets No. 1, 2 and 3 (Figures 4a, 5a, and 5b) provided no basis  
617 for answering the question regarding the appropriate and unquestionable number of classes in  
618 the division of normalised hyetographs from Kraków into clusters. To illustrate the problem in  
619 a simple way, even if a certain subjective level of bond distance is adopted a priori, e.g., 1.75,  
620 then the structure of the dendrogram of sets No. 1, 2 and 3 allows for designating the following  
621 mutually divergent numbers of clusters: 11, 9 and 6. Considering the substantial discrepancy in

622 terms of number of clusters in further research, a potential range of the number of clusters  $k$   
623 was subject to analysis, covering values from 2 to 20.

624 The results of calculations of the Caliński and Harabasz Index values ( $CHIndex$ ) as well  
625 as total within sum of squares ( $wss$ ) for the number of clusters  $k$  within a range from 2 to 20 for  
626 set No. 1 are presented in figure 7.

627

628 **Figure 7.** Value of the  $CHIndex$  and total within sum of squares ( $wss$ ), and for a set 1 of 313  
629 rainfalls from Kraków, depending on the adopted number of clusters  $k$

630

631 Diagrams of both parameters directly suggest that the optimum number of clusters for  
632 set No. 1 should be adopted as equal to 4 ( $k = 4$ ). By theory,  $wss$  values naturally decrease with  
633 an increase in the number of clusters  $k$ , although the decrease gradient evidently decreases after  
634 reaching number of clusters  $k = 4$ . For the same number of four clusters, maximum  $CHIndex$   
635 value is also observed ( $CHIndex=170$ ). In accordance with formula (4), a high  $CHIndex$  value  
636 is correlated with maximisation of the ratio of  $SS_B$  and  $SS_W$ . This means that particular clusters  
637 very significantly differ from one another, and elements of the set grouped in particular clusters  
638 are strongly similar to one another (relatively weakly variable).

639 The determination of the optimum number of clusters  $k = 4$  was followed by the  $k$ -  
640 means clustering process for set No. 1 of normalised cumulative hyetographs from Kraków.  
641 Through the application of the bootstrap method, the following values of the bootmean  
642 parameter were obtained for subsequent clusters from 1 to 4: 0.72; 0.65; 0.84, and 0.74.  
643 Bootmean parameter values were higher than 0.6 for all four clusters. This evidences that the  
644 designated four clusters included no cluster with random character, i.e., one that includes  
645 rainfall models deviating from the remaining three clusters, but at the same time not mutually  
646 similar. As a result of the  $k$ -mean clustering process for set No. 1 including 313 cumulative

647 normalised hyetographs from Kraków, clusters No. 1, 2, 3, and 4 were ascribed 102, 93, 68,  
648 and 50 rainfalls, respectively. This corresponded with the share of 35%, 32%, 23%, and 17%,  
649 respectively, throughout the set of analysed rainfalls.

650 For the designated clusters, averaged dimensionless cumulative storm rainfall  
651 hyetographs were determined, presented in figure 8. The cumulative dimensionless  
652 hyetographs were also transformed to the form of storm rainfall hyetographs presented in  
653 figure 9.

654

655 **Figure 8.** Diagrams of averaged dimensionless cumulative rainfall hyetographs for four clusters  
656 designated by means of k-means clustering for Kraków based on set No. 1 for 313 storm  
657 rainfalls

658

659 **Figure 9.** Model dimensionless rainfall hyetographs developed by means of the k-means  
660 clustering method for set No. 1 for 313 rainfalls from Kraków. The horizontal axis shows  
661 percent increase in rainfall duration, and the vertical axis shows percent shares in total  
662 precipitation amount

663

664 The analysis of the obtained model hyetographs shows that the most frequently  
665 occurring hyetographs of type 1 and 2 (35% and 32%, respectively) have relatively even values  
666 of point rainfall intensity (point rainfall depths for unitary duration intervals of 1/100 of total  
667 duration do not exceed 2% total rainfall depth). The rainfalls, however, do not correspond with  
668 the simplified block rainfall model, commonly used in designing drainage systems, and even  
669 sporadically applied in their hydrodynamic modelling. In the case of numerous unitary duration  
670 intervals, point rainfall depths differ from 1%, they are considerably lower, and frequently  
671 approximate to 0.5%, or considerably higher, reaching approximately 2% of total rainfall depth.

672 The substantially more seldom occurring model hyetograph for cluster 3 has a shape very  
673 generally approximate to model rainfall according to Euler type II recommended by Schmitt  
674 (2000) for modelling stormwater systems. Unlike in the case of rainfall model according to  
675 Euler type II, the greatest rainfall accumulation occurs not in 1/3 of rainfall duration, but already  
676 in its initial part, during the first 10% of the duration of the entire rainfall. Moreover, this  
677 accumulation has no character of a very sharp peak (point rainfall depths for unitary duration  
678 intervals of 1/100 of total duration do not exceed 4% of total rainfall depth). The most seldom  
679 occurring hyetograph shape is that determined in the case of cluster 4. With a high degree of  
680 generalisation, the hyetograph can be treated as a mirror reflection of the model hyetograph of  
681 cluster 3. This is not strictly accurate, because the rainfall accumulation occurring in the final  
682 part of the hyetograph is very obscure. It is observed in the final interval covering approximately  
683 30% of the entire rainfall, and point rainfall depths for unitary duration intervals of 1/100 of  
684 total duration do not exceed 2.5% of total rainfall depth in that case. Moreover, at the very  
685 beginning of the model hyetograph for cluster 4, a small rainfall peak occurs, with no equivalent  
686 in the final part of the model hyetograph for cluster 3. Referring to the discussed results obtained  
687 in the analysis of set No. 1 from Kraków, it is worth emphasising that the obtained courses of  
688 cumulative model hyetographs in figure 8 were very approximate to the cumulative model  
689 hyetographs obtained by Licznar (2018) for another rain gauge from Poland, as a result of  
690 application of clustering of a set of 213 storm rainfalls by means of the k-means clustering  
691 method for a subjectively adopted number of 4 clusters. The obtained results also remain in  
692 accordance with earlier studies from the territory of Poland that question the justification of  
693 common application of the synthetic rainfall model according to Euler type II due to its  
694 deviation from the vast majority of scenarios of temporal course of actual storm rainfalls  
695 (Licznar and Szelağ, 2014; Licznar et al., 2017).



696           Diagrams of the dependencies of the Caliński and Harabasz (*CHIndex*) index values and  
697 total within sum of squares (*wss*) on the number of clusters *k* developed for set No. 2 are  
698 presented in figure 10.

699

700 **Figure 10.** Value of the CHIndex and total within sum of squares (*wss*), and for set No. 2 of  
701 1493 rainfalls from Kraków, depending on the adopted number of clusters *k*

702

703           Analogically to set No. 1, they provide the basis for the determination that the optimum  
704 number of clusters for set No. 2 should be adopted as equal to 4 ( $k = 4$ ). For four clusters ( $k =$   
705 4), an evident peak of the CHIndex is observed ( $CHIndex=801$ ), and the curve of total within  
706 sum of squares (*wss*) flattens out to below 2000 after a rapid decrease from a level of  
707 approximately 5000. The cited CHIndex and *wss* values cannot be referred to values obtained  
708 in the case of set No. 1 (Figure 7). Orders of magnitude in both cases are different, as results  
709 from different abundance of sets No. 1 and No. 2, and as accounted for by the structure of  
710 formulas (3) and (4).

711           The process of k-means clustering of set No. 2 of normalised cumulative hyetographs  
712 from Kraków into four clusters ended with the designation of clusters meeting the criteria of  
713 internal coherence and external isolation (Gordon, 1999). The confirmation of the above was  
714 obtaining values of the bootmean parameter substantially exceeding the threshold of 0.6 for  
715 each of the clusters. The parameter reached: 0.93; 0.91; 0.96, and 0.91, respectively for clusters  
716 from 1 to 4. As the final result of clustering, subsequent clusters were ascribed, respectively:  
717 510, 506, 232, and 245 storm rainfalls. This corresponded with the respective share of 34%,  
718 34%, 16%, and 16% in the entire population of analysed rainfalls in set No. 2. The cited percent  
719 share of particular clusters in set No. 2 proved approximate, like in the case of the previously  
720 discussed set No. 1. For particular clusters, averaged cumulative storm rainfall hyetographs

721 were also determined, presented in figure 11. The diagrams of averaged normalised cumulative  
722 hyetographs in the figure pointed to very high compatibility in terms of temporal distribution  
723 with the previously discussed results for set No. 1 (Figure 8).

724

725 **Figure 11.** Diagrams of averaged dimensionless cumulative rainfall hyetographs for four  
726 clusters designated by means of the k-means clustering method for Kraków based on set No. 2  
727 for 1493 storm rainfalls

728

729 The evident divergence of results of clustering of storm rainfall hyetographs for sets  
730 No. 1 and No. 2 became an impulse for undertaking analysis of set No. 3, constituting a  
731 combination of the aforementioned sets. As expected, the obtained results proved to be virtually  
732 identical to results obtained previously for sets No. 1 and No. 2. Curves of variability of values  
733 of the Caliński and Harabasz index (CHIndex) and total within sum of squares (wss) developed  
734 for this large set of 1806 storm rainfalls are presented in figure 12, and raise no doubts as for  
735 their interpretation. The evident maximum of the CHIndex value ( $CHIndex=1105$ ) is obtained  
736 for four clusters ( $k = 4$ ). For four clusters, a very steep gradient of decrease in wss values from  
737 more than 5500 to approximately 2000 is also rapidly levelled almost to zero. Clustering by  
738 means of the k-means method of hyetographs included in set No. 3 ended with creating four  
739 clusters meeting the criteria of internal coherence and external isolation. The obtained  
740 bootmean parameter values were practically approximate to 1.0, and for subsequent clusters  
741 they reached: 0.94; 0.93; 0.96, and 0.94, respectively. Clusters No. 1, 2, 3, and 4 were ascribed:  
742 613, 605, 288, and 300 rainfalls, respectively, corresponding to a respective share of 34%, 33%,  
743 16%, and 17% in the entire set No. 3.

744

745 **Figure 12.** Value of the CHIndex and total within sum of squares (wss), and for a set 3 of 1806  
746 rainfalls from Kraków, depending on the adopted number k of clusters

747

748 For particular clusters of set No. 3, averaged hyetographs of dimensionless cumulative  
749 storm rainfalls were also determined, collectively presented in figure 13.

750

751 **Figure 13.** Diagrams of averaged dimensionless cumulative rainfall hyetographs for four  
752 clusters designated by means of the k-means clustering method for Kraków based on set No. 3  
753 for 1806 storm rainfalls. For comparison, the diagram also shows Median Time Distributions  
754 of Heavy Storm Rainfall at a Point developed by Huff (1990)

755

756 The figure also presents dimensionless hyetographs developed by Huff (1990),  
757 constituting the subject of a separate discussion. Pursuant to the expectations, the diagram in  
758 figure 13 proved strongly similar to the diagrams of model hyetographs for clusters from 1 to 4  
759 determined for sets No. 1 and No. 2 (Figures 8 and 11). For better assessment of the scale of  
760 similarity of the model hyetographs obtained for particular clusters in the case of clustering sets  
761 No. 1, No. 2, and No. 3, figure 14 was additionally prepared.

762

763 **Figure 14.** Comparison of averaged dimensionless cumulative rainfall hyetographs for four  
764 clusters designated by means of the k-means clustering method based on sets No. 1, 2, and 3:  
765 313, 1493, and 1806 storm rainfalls recorded in Kraków. The horizontal axes present percent  
766 increase in rainfall duration, and the vertical axes show percent shares in total rainfall depth

767

768 The similarity of averaged normalised cumulative rainfall hyetographs obtained for all  
769 three sets of storm rainfalls from Kraków within particular clusters is unquestionable. In the

770 case of sets No. 2 and No. 3, the diagrams of normalised cumulative hyetographs even overlap.  
771 In the case of cluster No. 4, hyetographs resulting from the analysis of all three sets overlap.  
772 This provides the basis for the presumption that the developed methodology of application of  
773 cluster analysis for the designation of model hyetographs shows repeatability in the scope of  
774 obtained results within the urban precipitation field. In engineering practice, it can be therefore  
775 used for the determination of local sets of model hyetographs based on the analysis of  
776 approximately 30-year-long rainfall series even from single rain gauges located near city  
777 boundaries. They will be able to supply computer models for simulation of drainage systems in  
778 the centre of a city as large as Kraków with satisfactory precision.

779 For a better understanding of the obtained results, particularly including the  
780 determination of characteristic features of particular clusters that differentiate them from one  
781 another, their profiling was performed. Such profiling is called simplified, because it is only  
782 based on general variables describing storm rainfalls in the form of their total depths, total  
783 durations, and mean intensities. For this purpose, table 2 presents mean values of these variables  
784 within particular clusters for sets No. 1 and No. 2. Figures 15a and 15b present histograms of  
785 the frequency of occurrence of storm rainfalls with different mean intensities within different  
786 clusters, respectively for sets No. 1 and No. 2. Values of rainfall intensity in diagrams in  
787 figures 15a and 15b and in table 2 are expressed in a standard unit of  $\text{dm}^3/(\text{s}\cdot\text{ha})$  applied in urban  
788 hydrology.

789

790 **Table 2.** Mean values of total depths, durations, and intensities of storm rainfalls included in  
791 particular clusters for sets No. 1 and No. 2 from Kraków

792

793



794 **Figure 15 a.** Histograms of mean rainfall intensities for four designated clusters in rainfall set  
795 No. 1 from Kraków (313 rainfalls); **b.** Histograms of mean rainfall intensities for four  
796 designated clusters in rainfall set No. 2 from Kraków (1493 rainfalls)

797

798 Data included in table 2 and in diagrams in figures 15a and 15b suggest that rainfalls  
799 classified to cluster 3 usually showed mean intensities higher than those of rainfalls included in  
800 other clusters. For the majority of storm rainfalls included in cluster 3 in sets No. 1 and No. 2,  
801 mean intensities in sets No. 1 and No. 2 were usually within the range from 5 to 20  $\text{dm}^3/(\text{s}\cdot\text{ha})$ ,  
802 and their mean value within sets No. 1 and No. 2 exceeded 14  $\text{dm}^3/(\text{s}\cdot\text{ha})$ . High mean rainfall  
803 intensities in cluster 3 did not result from high total rainfall depths, but from evidently shorter  
804 durations. In set No. 2, mean rainfall depths for cluster 3 were even evidently lower than for  
805 the three remaining clusters. Mean rainfall durations for cluster 3, however, were approximately  
806 6 h, whereas for the remaining clusters they were considerably longer, within a range from  
807 approximately 8 h to 12 h. This suggests that cluster  $n = 3$  included short but very intensive  
808 rainfalls, probably with convection genesis. This hypothesis would also explain the greatest  
809 variability of point rainfall depths for unitary duration intervals shown on the model hyetograph  
810 of cluster 3 in figure 9. Completely different profiling results were obtained for cluster 2 which  
811 had a considerably more equalised course of the model hyetograph (Figure 9). Rainfalls  
812 classified to this cluster had not only longer durations within the clusters in particular sets No. 1  
813 and No. 2, but usually also very low intensities within a range from 0 to 5  $\text{dm}^3/(\text{s}\cdot\text{ha})$ . This  
814 encourages a hypothesis that this cluster primarily included frontal rainfalls with long durations,  
815 but low and considerably more even in time intensities. The research hypotheses stated here  
816 regarding the division of storm rainfalls into clusters by rainfall genesis should be verified in  
817 further research. It will however require access to synoptic records permitting more precise  
818 profiling of the designated clusters.

819 In the summary of the entire discussion of results, it is also necessary to refer to the  
820 classic papers by Huff (1967, 1990). Although the research on Dimensionless, Cumulative  
821 Rainfall Hyetographs was implemented in different climate conditions (Illinois, USA), with a  
822 different approach to the identification of single rainfalls, and with the application of a  
823 considerably simpler method of their classification to the first, second, third, or fourth quartile  
824 (depending on whether the highest percent of cumulative rainfall occurred in the first, second,  
825 third, or fourth quarter of its duration, respectively), relatively high correspondence of the  
826 courses of medians is observed (50th-percentile) between dimensionless hyetograph curves  
827 derived from point rainfall values derived by Huff (1967, 1990) and model hyetographs in  
828 figures 8, 11 and 13. In the case of hyetographs from Kraków, cluster 3 corresponds with first-  
829 quartile storms, cluster 1 with second-quartile storms, cluster 2 with third-quartile storms, and  
830 cluster 4 with fourth-quartile storms. Further analogies can be sought in general characteristics  
831 of particular quartiles. For designing and modelling drainage systems, Huff (1990)  
832 recommended the application of first-quartile storm hyetographs for time scales of about 6  
833 hours or less, and second-quartile storm hyetographs for time scales of about 6 to 12 hours.  
834 These recommendations overlap with mean durations determined for clusters No. 3 and No. 1  
835 in sets No. 1 and 2, respectively (Table 2). Nonetheless, at a closer investigation of the study  
836 by Huff (1967), differences are observed in terms of frequencies of occurrence of rainfalls with  
837 the adopted model hyetographs. In the case of research from Illinois, the relative frequencies of  
838 the storms were 30, 36, 19, and 15 percent for the first, second, third, and fourth quartiles,  
839 respectively. They are not in accordance with frequencies of 16, 34, 33, and 15 percent obtained  
840 for the analogical model hyetographs designated from set No. 3, corresponding to subsequent  
841 clusters No. 3, 1, 2, and 4, respectively. The determined divergence, however, does not  
842 undermine results from Kraków, because the analogical divergence has already been signalled  
843 and discussed by Pani and Haragan (1981). Analysing a set of 117 rainfalls from Texas (USA),

844 recorded in months with the highest probability of occurrence of convection rainfalls, the  
845 authors obtained a median (50th-percentile) dimensionless hyetograph curve with shapes fully  
846 corresponding with results by Huff (1967), but the determined relative frequencies of the storms  
847 were 13, 41, 32, and 14 percent for the first, second, third, and fourth quartiles, respectively.  
848 The latter frequencies are considerably more approximate to results from Kraków, despite  
849 obvious differences in the location of both research polygons and approach to processing  
850 rainfall records and determination of model hyetographs. The qualitative compatibility of study  
851 results from Kraków with classic papers by Huff (1967, 1990) and Pani and Haragan (1981) is  
852 an additional premise confirming the accuracy of the methodology of identification of model  
853 hyetographs of storm rainfalls based on complete cluster analysis and quality indices.  
854

#### 855 **4. Summary and final conclusions**

856 Progress in the scope of atmospheric precipitation measurements techniques, and the  
857 occurring municipal rain gauge networks expansion requires simultaneous modernisation of the  
858 methodology of recorded precipitation series processing. A necessary element of such a  
859 methodology are certainly modern methods of objective and automatic search of rainfalls  
860 groups with similar courses in time that can be described in a general way by means of model  
861 hyetographs. The practical application area of such methods can currently exceed processing  
862 sets of local model hyetographs, and find implementation in the practical rainwater runoff  
863 control systems operation. Due to the growing number of rain gauges and rapidly increasing  
864 sets of rainfall records, also in the case of model hyetographs identification, it becomes justified  
865 and necessary to reach for data mining tools, primarily including the cluster analysis.

866 This paper is not pioneer in terms of the very idea of the cluster analysis application in storm  
867 rainfall hyetographs classification. The application of the cluster analysis in research on

868 temporal distributions of storm rainfalls has already been postulated by Licznar et al. (2017),  
869 and then tested for several locations in Poland (Licznar, 2018; Wartalska et al., 2020). The  
870 primary objective of the paper was the improvement of the research methodology to meet the  
871 requirements of the complete cluster analysis methodology covering seven stages (Milligan,  
872 1996), including: selection of objects and variables; formula selection of variable values  
873 normalisation; selection of distance measure; selection of the classification method; the number  
874 of classes determination; assessment of classification results; class description and profiling.  
875 The aforementioned objective covered the three primary detailed objectives involving:  
876 objectivization of the number of clusters determination, the internal coherence and external  
877 isolation of clusters verification, and profiling of the retrieved clusters.

878 Another substantial objective of the paper was to demonstrate the developed methodology at  
879 the scale of a large precipitation field in Poland. This objective covered two detailed objectives,  
880 namely testing the methodology on a large measurement set, and equally importantly, analysing  
881 its repeatability at the scale of a large urban precipitation field. The basic question was also to  
882 what extent model hyetographs developed based on records from nearby rain gauges (located  
883 e.g., at an airport or in suburbs) correspond with the shapes of model hyetographs for rain  
884 gauges of the urban rain monitoring network. Owing to the collaboration with the Municipal  
885 Water Supply and Sewerage Company (MWSSC) in Kraków, Poland, it was possible to apply  
886 the developed complete cluster analysis methodology to a large measurement set of 1806 storm  
887 rainfalls (Set No. 3), composed of set No. 1 – 313 storm rainfalls designated from two nearby  
888 rain gauges belonging to the countrywide network of IMGW, and set No. 2 – 1493 storm  
889 rainfalls designated from 23 rain gauges belonging to the municipal rain monitoring network  
890 of MWSSC. Three applications of the complete cluster analysis methodology for sets No. 1,  
891 No. 2, and No. 3 permitted its thorough testing, designation of a set of model hyetographs for



892 practical application in modelling of drainage systems in Kraków, and drawing the following  
893 final conclusions:

- 894 1) The complete methodology of the storm rainfall hyetographs cluster analysis should  
895 cover tools for both hierarchical and non-hierarchical analysis of the sets structure.  
896 Before the application of the cluster of normalised (dimensionless) hyetographs, key in  
897 terms of the final products, i.e., model hyetographs, by means of the k-means method,  
898 the hierarchical agglomeration method should be applied to prepare dendrograms of  
899 similarity of temporal courses of rainfalls in the analysed sets. The dendrograms should  
900 be subject to expert analysis not only in terms of determination of a potential number of  
901 clusters in the analysed sets, but more importantly the identification of particularly  
902 peculiar rainfall patterns. Like in the case of set No. 2 and rainfall recorded by rain  
903 gauge No. 12 in Płaszów in 2014, such records should be removed before the division  
904 of the set using a predefined number  $k$  of clusters by means of non-hierarchical methods;
- 905 2) The diagrams analysis of correlation of values of the Caliński and Harabasz index  
906 ( $CHIndex$ ) and total within sum of squares ( $wss$ ) with number  $k$  of clusters permits  
907 completely objective determination of the correct number of clusters for which the  
908 division of storm rainfall sets should be performed from the similarity point of view of  
909 their normalised (dimensionless) hyetographs. For the accurate, optimal number of  
910 clusters, maximisation of  $CHIndex$  values is observed combined with an evident  
911 decrease in the gradient of the decrease in  $wss$  values. In the case of all three analysed  
912 sets of storm rainfalls from Kraków, based on analyses of  $CHIndex$  and  $wss$  values, the  
913 adopted optimum number of clusters was four ( $k = 4$ ), and the choice was positively  
914 verified in all further research through obtaining clusters meeting the requirements of  
915 internal coherence and external isolation;



916 3) The fundamental element for the credibility of the obtained results of storm rainfalls  
917 divisions and model hyetographs identification is the assessment of classification  
918 results. The study conducted on three sets of rainfalls from Kraków justified the repeated  
919 launching of the clustering algorithm with the application of the bootstrap method.  
920 Although this undoubtedly complicates the computing algorithms and prolongs the time  
921 of calculations, it allows for calculating the bootmean parameter corresponding to the  
922 mean value of the Jaccard Index (Jaccard similarity coefficient) for each of the  
923 designated clusters. The bootmean parameter permits drawing objective conclusions on  
924 whether particular clusters meet the criteria of internal coherence and external isolation  
925 (Gordon, 1999). For all clusters designated from three sets of rainfalls from Kraków,  
926 the bootmean parameter usually exceeded the adopted threshold of 0.6, confirming that  
927 the designated subclusters included no clusters with random character, i.e., those  
928 including rainfall patterns deviating from the remaining clusters, but also evidently  
929 mutually different. Relatively lowest values of the bootmean parameter were obtained  
930 for the least abundant set No. 1 (in a range from 0.65 to 0.84), whereas for  
931 approximately five or six times more abundant sets No. 2 and 3, they were higher than  
932 0.9, or even approximate to 1.0. The latter observation suggests that the k-means  
933 clustering method is predestined for the analysis of very large sets, and is more reliable  
934 in their case;

935 4) The developed complex cluster analysis methodology for the division of sets of storm  
936 rainfalls and identification of model hyetographs implemented in the case of all three  
937 sets of storm rainfalls from Kraków generated coherent final results. For each of the  
938 three sets, the optimum number of clusters was four, and the resulting averaged  
939 normalised cumulative hyetographs for particular clusters showed no mutual differences  
940 within the three analysed sets. The coherence of the obtained results also concerned the



941 frequency of storm rainfalls occurrence included to particular clusters. For all the three  
942 sets, storm rainfalls were distributed in proportions of approximately: 1/3, 1/3, 1/6, and  
943 1/6 for clusters No. 1, 2, 3, and 4, respectively. All the aforementioned observations  
944 suggest the possibility of development of model hyetographs based on multiannual  
945 records from suburban stations (e.g., from the rain gauge at the nearby airport), and like  
946 in the case of set No. 1, their application in practice throughout the city in the case of  
947 lack of the possibility of hyetographs development based on records from the territory  
948 of the city itself (based on set No. 2).

949 5) The obtained set of model hyetographs for Kraków does not include hyetographs with  
950 a shape corresponding to that of synthetic hyetographs developed based on IDF  
951 (Intensity-Duration-Frequency) or DDF (Depth-Duration-Frequency) models, adopted  
952 *a priori* for hydrodynamic modelling of drainage systems in Poland, such as: model  
953 rainfall according to Euler type II, block rainfall, or model rainfall according to DVWK.  
954 Nonetheless, the shapes of the developed hyetographs point to high similarity to classic  
955 medians (50th-percentile) of dimensionless hyetograph curves derived from point  
956 rainfall values derived by Huff (1967, 1990) and Pani and Haragan (1981). In results  
957 from Kraków, model hyetographs for subsequent clusters No. 3, 1, 2, and 4 correspond  
958 to medians (50th-percentile) of dimensionless hyetographs for: first-quartile storms,  
959 second-quartile storms, third-quartile storms, and fourth-quartile storms. Profiling of  
960 clusters results of storm rainfalls from Kraków also remain in complete accordance with  
961 earlier research by Huff (1990) according to which first-quartile storms (storms from  
962 cluster 3) usually correspond with time scales of about 6 hours or less, whereas second-  
963 quartile storms (storms from cluster 1) usually have longer durations within a range  
964 from 6 to 12 hours.

965



966 The development of a complex methodology of storm rainfall hyetographs analysis and  
967 its successful testing in a polygon of a large rain gauge network in Kraków offers the possibility  
968 of its implementation at a considerably broader scale in the scope of implementation of the  
969 WaterFolder Connect project. The practical objective here is to develop credible sets of local  
970 model hyetographs in a network of 100 rain gauges in Poland for the purpose of their later use  
971 in practice for supplying a digital platform dedicated for designing and modelling drainage  
972 systems throughout Poland. Further research, however, must be also undertaken due to new  
973 hypotheses that appeared as a result of the study. Firstly, the research hypothesis assuming the  
974 correlation of the division of storm rainfalls into particular clusters with the genesis of rainfalls  
975 needs to be verified. Moreover, in the context of the determined lack of variability of model  
976 hyetographs at the single urban precipitation field scale, it is important to verify the thesis on  
977 the regionalisation of study results possibility from more mutually distant rain gauges, and the  
978 practical application of common sets of model hyetographs in larger areas of the country. In the  
979 future, the developed methodology of hyetograph analysis could be also implemented in other  
980 countries around the globe.

981  
982 **Acknowledgments:** This work was carried out as part of the fourth edition of the programme  
983 implementation doctorate conducted by the Ministry of Science and Higher Education,  
984 Republic of Poland, and within the project entitled '*WaterFolder Connect – an integrated*  
985 *platform for design and modelling of drainage systems – POIR.01.01.01-00-0119/21*', financed  
986 by the National Centre for Research and Development under the Smart Growth Operational  
987 Programme 2014-2020, Priority axis: Support for R&D in enterprises; Measure: R&D projects  
988 in enterprises; Sub-measure: Industrial research and development conducted by enterprises.





989 Precipitation data were partially provided by the Municipal Water Supply and Sewerage  
990 Company (MWSSC) in Kraków, Poland.

991

## 992 **References**

993

- 994 1. Kundzewicz, Z., Licznar, P., 2021 (2022). Climate change adjustments in engineering  
995 design standards. European perspective. *Water Policy* (in print).
- 996 2. Nix, S.J., 1994. *Urban stormwater modelling and simulation*. Lewis Publishers, CRC  
997 Press.
- 998 3. de Lima, M.I.P., 1998. *Multifractals and the temporal structure of rainfall*. Doctoral  
999 dissertation, Wageningen Agricultural University, Wageningen.
- 1000 4. Deidda, R., Benzi, R., Siccardi, F., 1999. Multifractal modeling of anomalous scaling  
1001 laws in rainfall. *Water Resources Research* 35, 1853–1867.
- 1002 5. Licznar, P., 2009. *Generatory syntetycznych szeregów opadowych do modelowania*  
1003 *sieci kanalizacji deszczowych i ogólnospławnych*. Monografie LXXVII. UP we  
1004 Wrocławiu, pp 180.
- 1005 6. Licznar, P., 2008. *Obliczenia częstotliwości nadpiętrzenia sieci kanalizacji*  
1006 *deszczowej*. *Gaz, Woda i Technika Sanitarna* 7-8, 16-21.
- 1007 7. Gires, A., Onof C., Maksimovic C., Schertzer D., Tchiguirinskaia I., and Simoes N.,  
1008 2012. Quantifying the impact of small scale unmeasured rainfall variability on urban  
1009 hydrology through multifractal downscaling: a case study, *J. Hydrol.* 442–443, 117–  
010 128.

- 1011 8. Gires, A., Tchiguirinskaia, I., Schertzer, D., and Lovejoy, S., 2013. Multifractal analysis  
1012 of an urban hydrological model on a Seine-Saint-Denis study case, *Urban Water J.* 10,  
1013 195–208. *Remote Sensing of Environment* 147, 173-185.
- 1014 9. Zhou, Y., Smith, S.J., Elvidge, C.D., Zhao, K., Thomson, A., Imhoff, M., 2014. A  
1015 cluster-based method to map urban area from DMSP/OLS n
- 1016 10. Licznar, P., 2013. Stormwater reservoir dimensioning based on synthetic rainfall time  
1017 series, *Ochrona Srodowiska*, 35, 27–32.
- 1018 11. Jakubiak, P., Licznar, P., Malinowski, S., 2014. Rainfall estimates from radar vs.  
1019 raingauge measurements. Warsaw case study. *Environment Protection Engineering* 40,  
1020 162-170.
- 1021 12. Nayak, M.A., Ghosh, S., 2013. Prediction of extreme rainfall event using weather  
1022 pattern recognition and support vector machine classifier. *Theor Appl Climatol* 114,  
1023 583–603.
- 1024 13. Molnar, P., Lüscher, R., Hausherr, R., 2006. Impact of storm rainfall variability on  
1025 urban drainage system performance. *Proc. 7th International Workshop on Precipitation*  
1026 *in Urban Areas: Extreme Precipitation, Multisource Data Measurement and*  
1027 *Uncertainty*, 7-10 Dec. 2006, St. Moritz, Switzerland.
- 1028 14. Güntner, A., Olsson, J., Calver, A., Gannon, B., 2001. Cascade-based disaggregation of  
1029 continuous rainfall time series: the influence of climate. *Hydrology & Earth System*  
1030 *Sciences* 5, 145–164.
- 1031 15. Hingray, B. and Ben Haha, M., 2005. Statistical performances of various deterministic  
1032 and stochastic models for rainfall series disaggregation, *Atmos. Res.* 77, 152–175.
- 033 16. Licznar, P., Łomotowski, J., Rupp, D.E., 2011a. Random cascade driven rainfall  
034 disaggregation for urban hydrology: An evaluation of six models and a new generator.  
035 *Atmospheric Research* 99, 563-578.



- 1036 17. Licznar, P., Schmitt, T.G., Rupp, D.E., 2011b. Distributions of microcanonical cascade  
1037 weights of rainfall at small timescales. *Acta Geophysica* 59, 1013-1043.
- 1038 18. Over, T.M., Gupta, V.K., 1994. Statistical analysis of mesoscale rainfall: dependence  
1039 of a random cascade generator on large-scale forcing. *Journal of Applied Meteorology*  
1040 33, 1526–1542.
- 1041 19. Rupp, D.E., Licznar, P., Adamowski, W., Leśniewski, M., 2012. Multiplicative cascade  
1042 models for fine spatial downscaling of rainfall: parameterization with rain gauge data.  
1043 *Hydrol. Earth Syst. Sci.* 16, 671–684.
- 1044 20. Deidda, R., 2000. Rainfall downscaling in a space-time multifractal framework. *Water*  
1045 *Resources Research*. <https://doi.org/10.1029/2000WR900038>.
- 1046 21. Schmitt, T.G., 2000. Kommentar zum Arbeitsblatt A 118 Hydraulische Bemessung  
1047 und Nachweis von Entwässerungssystemen. DWA, Hennef.
- 1048 22. Sifalda, V., 1973. Entwicklung eines Berechnungsregens für die Bemessung von  
1049 Kanalnetzen. *GWF Wasser Abwasser* 114, 435–440.
- 1050 23. Yen, B.C., Chow, V.T., 1980. Design hyetographs for small drainage structures. *J.*  
1051 *Hydraul. Eng. Div. ASCE* 106, 1055–1976.
- 1052 24. Desbordes, M., 1978. Urban runoff and design storm modeling. In *Proceedings of the*  
1053 *First International Conference on Urban Drainage*, London, UK, April 1978; pp. 353–  
1054 361.
- 1055 25. Peyron, N., Nguyen, V.T.V., Rivard, G., 2002. An optimal design storm pattern for  
1056 urban runoff estimation in southern Québec. In *Proceedings of the 30th Annual*  
1057 *Conference of the Canadian Society for Civil Engineering*, Montréal, QC, Canada, 5–8  
1058 June 2002.
- 1059 26. Lee, K.T., Ho, J.Y., 2008. Design hyetograph for typhoon rainstorms in Taiwan. *J.*  
1060 *Hydrol. Eng.* 13, 647–651.



- 1061 27. Keifer, C.J., Chu, H.H., 1957. Synthetic storm pattern for drainage design. *J. Hydr. Eng.*  
1062 Div. 83, 1–25.
- 1063 28. DVWK. Arbeitsanleitung zur Anwendung Niederschlag-Abflub-Modellen in kleinen  
1064 Einzugsgebieten. Regeln 113 (Teil II: Synthese); Verlag Paul Parey: Hamburg,  
1065 Germany, 1984.
- 1066 29. Huff, F.A., 1967. Time Distribution of Rainfall in Heavy Storms. *Water Resources*  
1067 *Research*, 3, 1007-1019.
- 1068 30. Huff, F.A., 1970. Time Distribution Characteristics of Rainfall Rates, *Water Resources*  
1069 *Research* 6, 447-454.
- 1070 31. Huff, F.A., 1990. Time Distributions of Heavy Rainstorms in Illinois; Circular. Illinois  
1071 StateWater Survey 173: Champaign, IL, USA, 1990; Available online:  
1072 <http://hdl.handle.net/2142/94492>
- 1073 32. Pani, E.A., Haragan, D.R., 1981. A comparison of Texas and Illinois temporal rainfall  
1074 distributions. In *Proceedings of the Fourth Conference on Hydrometeorology*, Reno,  
1075 NV, USA, 7–9 October 1981; American Meteorological Society: Boston, MA, USA,  
1076 1981; pp. 76–80.
- 1077 33. Bonta, J.V., Rao, A.R., 1987. Factors affecting development of Huff curves. *Trans.*  
1078 *ASAE* 30, 1689–1693.
- 1079 34. Bonta, J.V., 2004. Development and utility of Huff curves for disaggregating  
1080 precipitation amounts. *Appl. Eng. Agric.* 20, 641–652.
- 1081 35. Terranova, O.G., Iaquina, P., 2011. Temporal properties of rainfall events in Calabria  
1082 (southern Italy). *Nat. Hazard. Earth Syst.* 11, 751–757.
- 083 36. Elfeki, A.M., Ewea, H.A., Al-Amri, N.S., 2014. Development of storm hyetographs for  
084 flood forecasting in the Kingdom of Saudi Arabia. *Arab. J. Geosci.* 7, 4387–4398.



- 1085 37. Pan, C., Wang, X., Liu, L., Huang, H., Wang, D., 2017. Improvement to the Huff Curve  
1086 for Design Storms and Urban Flooding Simulations in Guangzhou, China. *Water* 9, 411.
- 1087 38. McCuen, R. H., 1986. *Hydrologic Analysis and Design*. Prentice Hall.
- 1088 39. Licznar, P., Szelaǳ B., 2014. Analiza zmiennoŹci czasowej opadów atmosferycznych w  
1089 Warszawie. *Ochrona Źrodowiska* 36, 23–28.
- 1090 40. Licznar, P., Burszta-Adamiak, E., Łomotowski, J., Stańczyk, J., 2017. Modern proposal  
1091 of methodology for retrieval of characteristic synthetic rainfall hyetographs. *E3S Web*  
1092 *of Conferences* 22, 00104 (2017) DOI: 10.1051/e3sconf/20172200104 (E3S Web of  
1093 Conferences, Volume 22 (2017), International Conference on Advances in Energy  
1094 Systems and Environmental Engineering (ASEE17), Wrocław, Poland, July 2-5, 2017,  
1095 B. Kaźmierczak, M. Kutylowska, K. Piekarska, H. Jouhara and J. Danielewicz (Eds.))
- 1096 41. Licznar, P., 2018. Analiza opadów atmosferycznych na potrzeby projektowania  
1097 systemów odwodnienia. *Monografie Komitetu Inżynierii Źrodowiska PAN* 137,  
1098 Wrocław 2018, pp 209.
- 1099 42. Wartalska, K., Kaźmierczak, B., Nowakowska, M., Kotowski, A., 2020. Analysis of  
1100 Hyetographs for Drainage System Modeling. *Water* 12, 149.
- 1101 43. Burszta-Adamiak, E., Licznar, P., Zaleski, J., 2019. Criteria for identifying maximum  
1102 rainfalls determined by the peaks-over-threshold (POT) method under the Polish Atlas  
1103 of Rainfalls Intensities (PANDa) project. *Meteorology, Hydrology and Water*  
1104 *Management* 7, 3-13.
- 1105 44. Licznar, P., Siekanowicz, K., Stach, A., Zaleski, J., 2020. Atlasy opadowe. W:  
1106 *Metodyka opracowania polskiego atlasu natężeń deszczów (PANDa) / pod red. Pawła*  
107 *Licznara i Janusza Zaleskiego*. *Warszaw: Instytut Meteorologii i Gospodarki Wodnej*.  
108 *Państwowy Instytut Badawczy* 99-137.



- 1109 45. Larose, D.T, 2005. Discovering knowledge in data. An Introduction to Data Mining.  
1110 John Wiley & Sons, Inc., Hoboken, New Jersey and Canada.
- 1111 46. Stanisław, A., 2007. Przystępny Kurs Statystyki z Zastosowaniem STATISTICA PL na  
1112 Przykładach z Medycyny. Tom I-III; StatSoft Polska Sp. z o.o.: Kraków, Poland.
- 1113 47. Licznar, P., De Michele, C., Adamowski, W., 2015. Precipitation variability within an  
1114 urban monitoring network via microcanonical cascade generators, Hydrol. Earth Syst.  
1115 Sci. 19, 485-506.
- 1116 48. Gordon, A.D., 1999. Classification. Chapman and Hall/CDC, London.
- 1117 49. Everitt, B.S., Landau, S., Leese, M., 2001. Cluster analysis. Edward Arnold, London.
- 1118 50. Dżimińska, P., Drzewiecki, S., Ruman, M., Kosek, K., Mikołajewski, K., Licznar, P.,  
1119 2021. The Use of Cluster Analysis to Evaluate the Impact of COVID-19 Pandemic on  
1120 Daily Water Demand Patterns. Sustainability 13, 5772.
- 1121 51. Walesiak, M., Gatnar, E., 2009. Analiza Skupień. Statystyczna Analiza Danych z  
1122 Wykorzystaniem Programu R; Wydawnictwo Naukowe PWN: Warszawa, Poland.



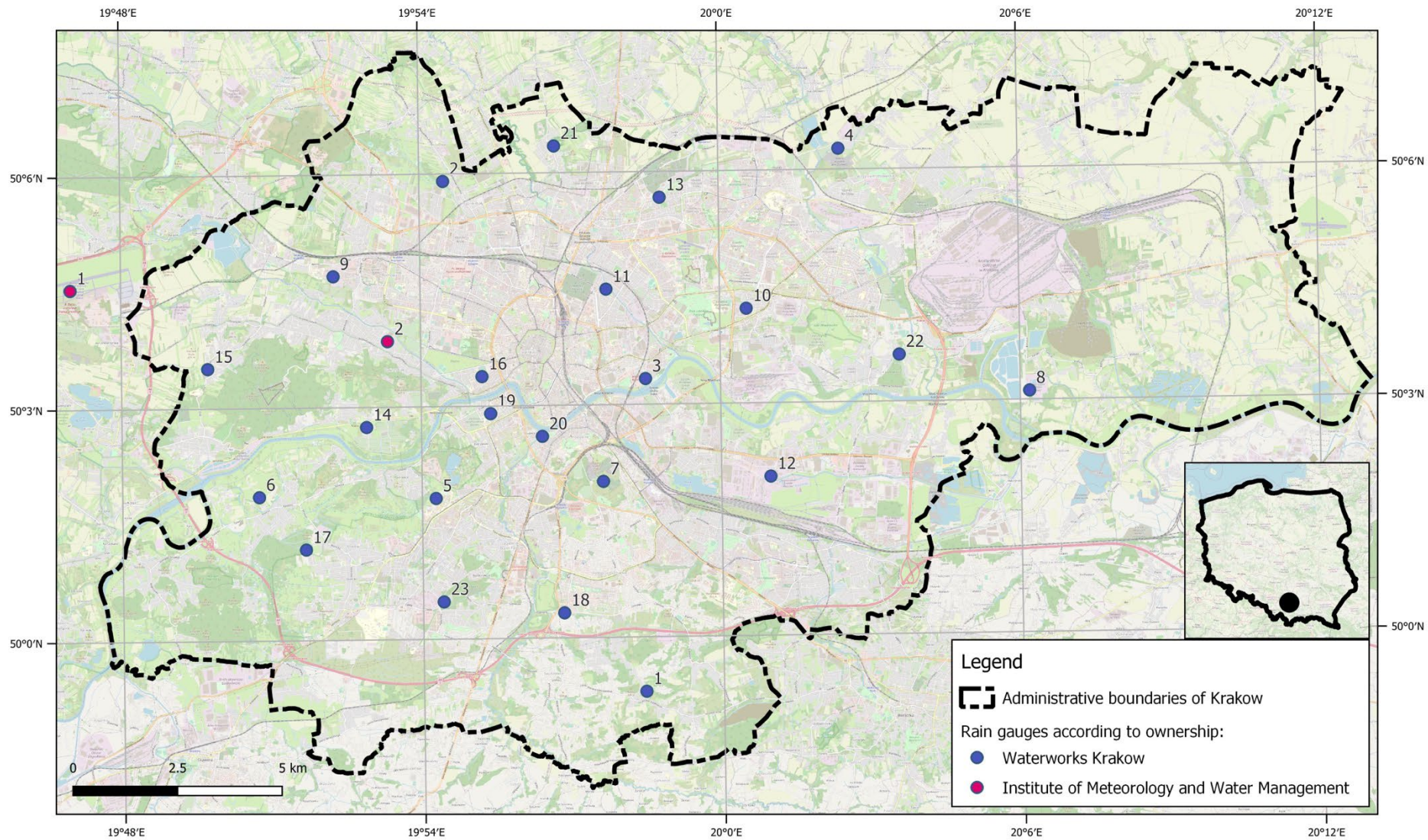
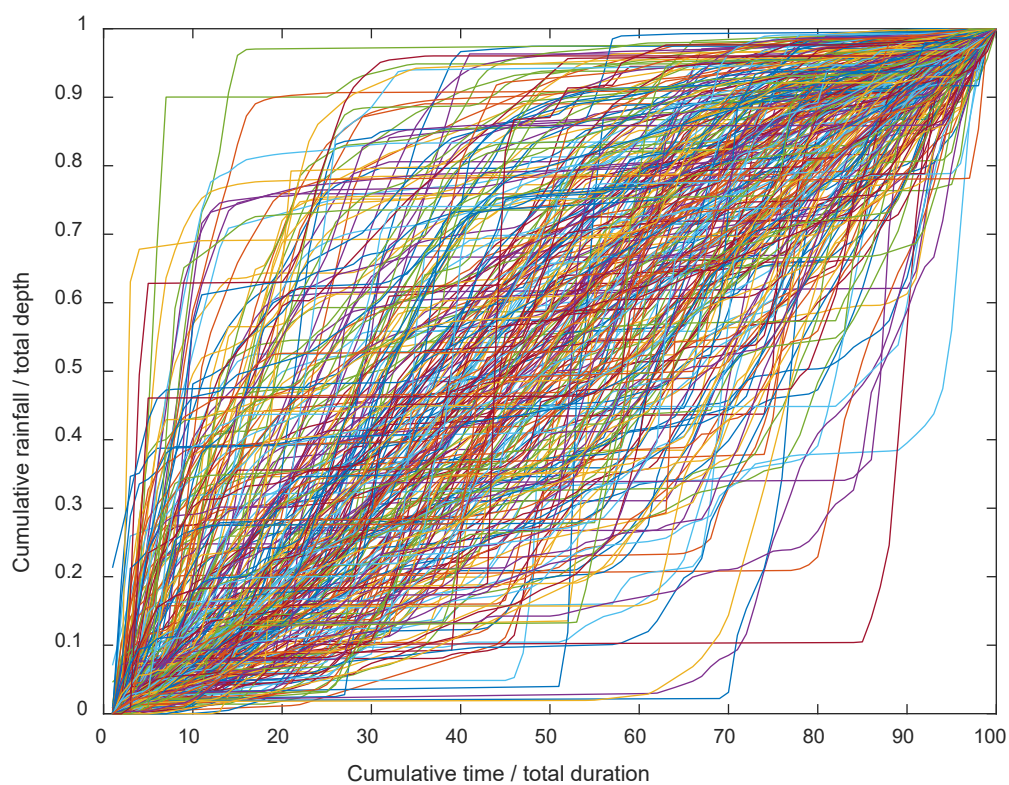
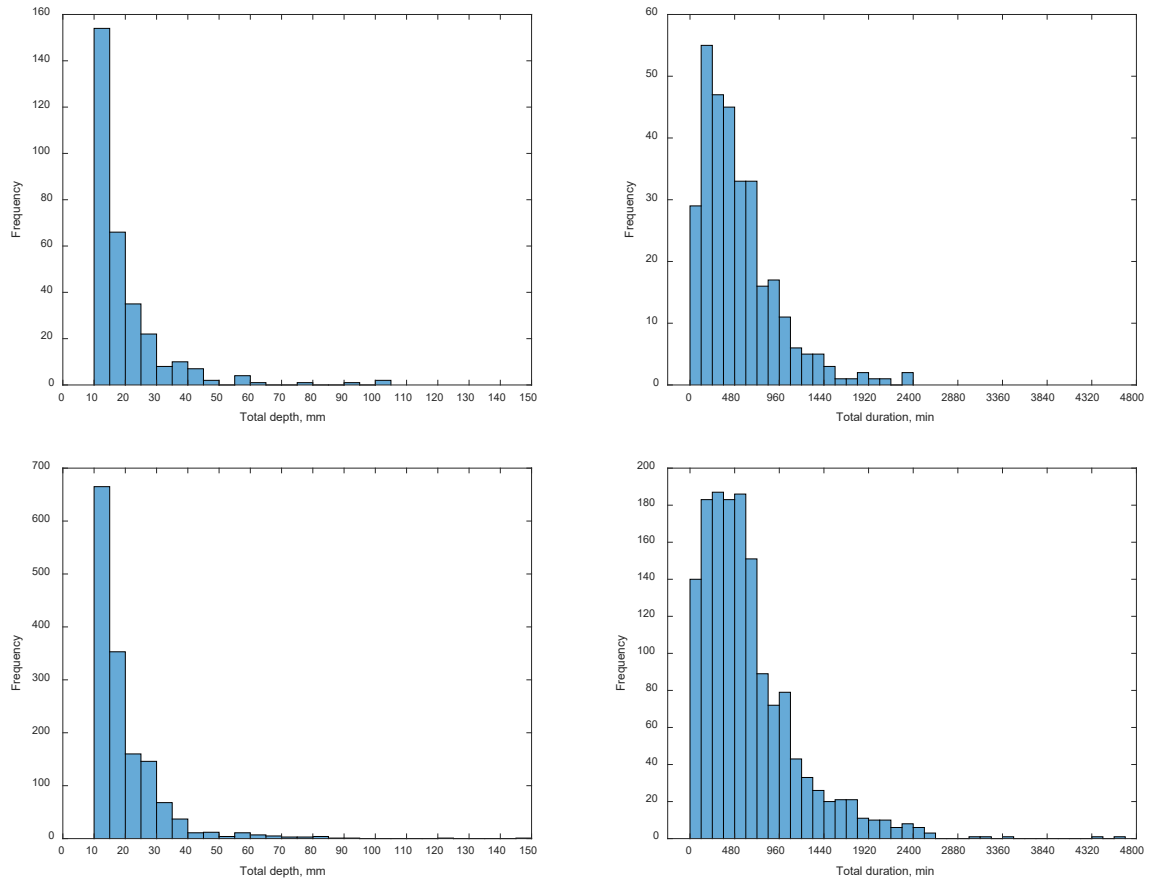


Figure 1. Location of rain gauges belonging to Waterworks Kraków and the Institute of Meteorology and Water Management

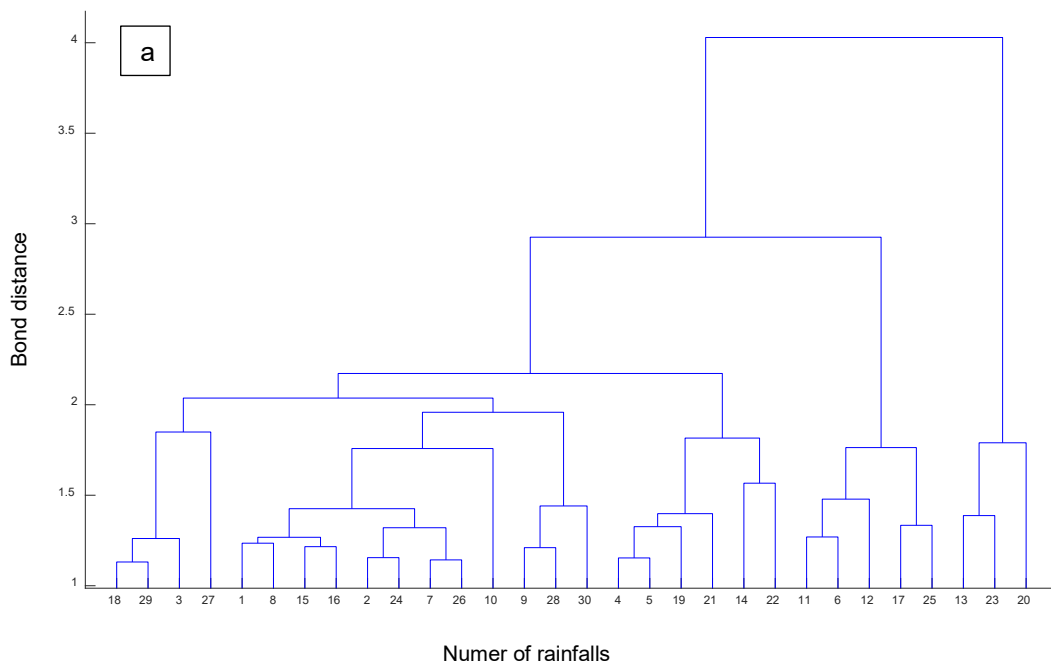
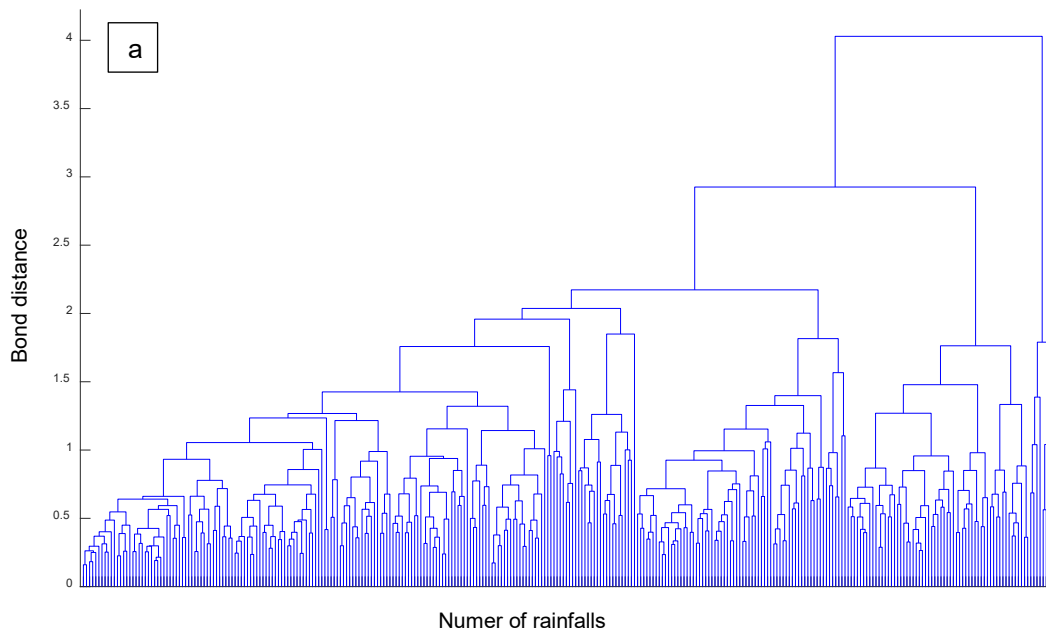


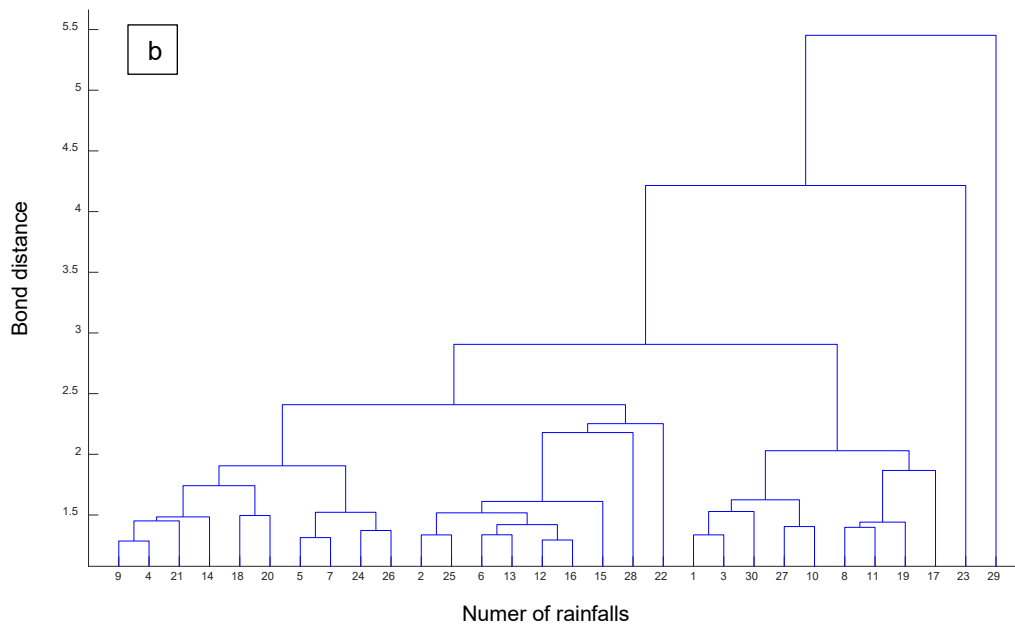
**Figure 2.** Cumulative dimensionless hyetographs of 313 storm rainfalls from Kraków (set No. 1)



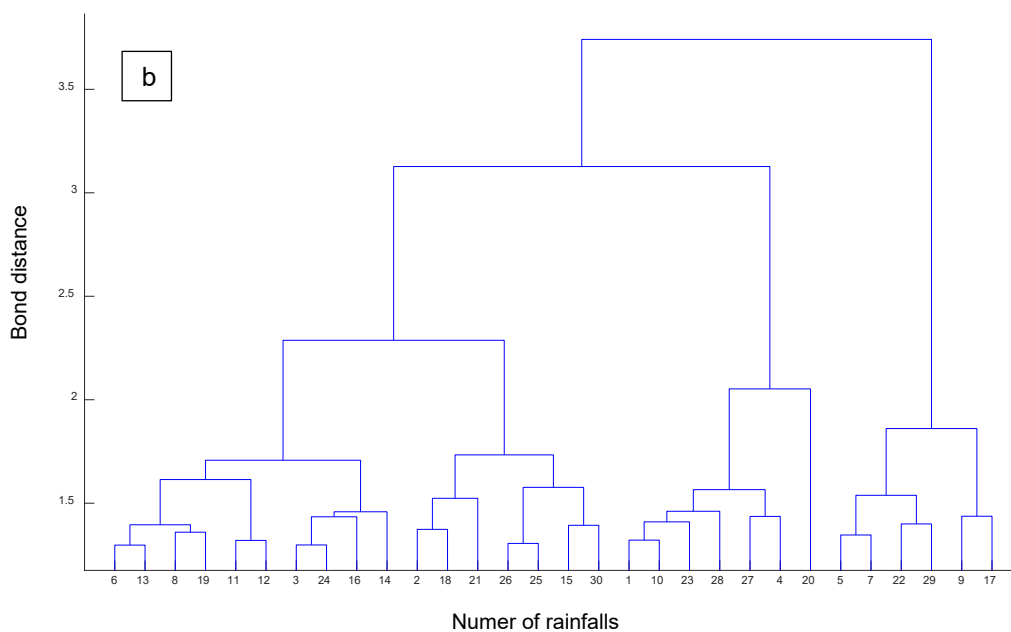
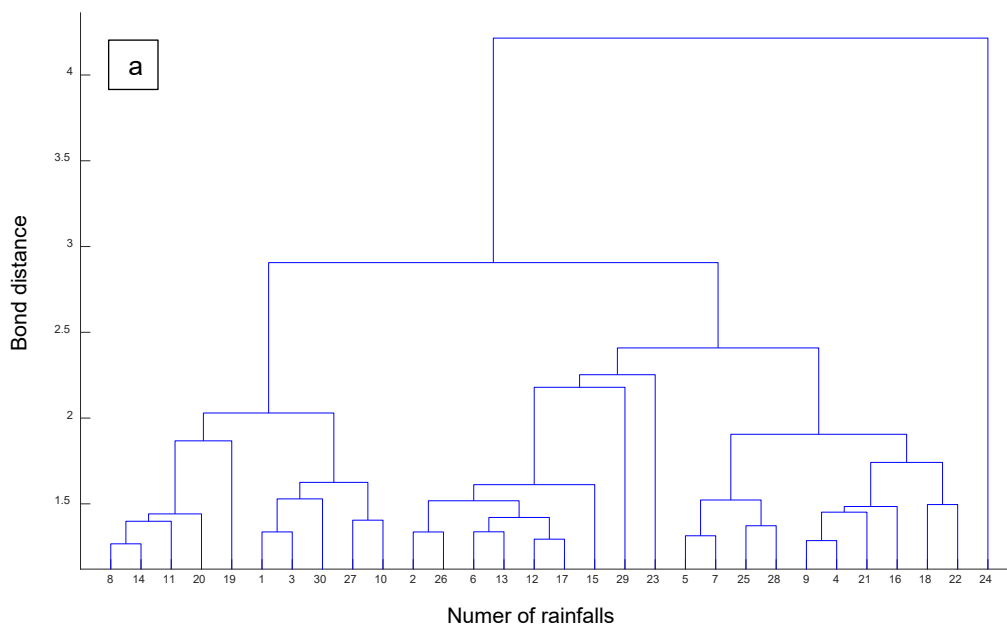


**Figure 3.** Histograms of rainfall depths and durations for the designated sets of rainfalls from Kraków, diagrams in the top row for set No. 1 (313 rainfalls), diagrams in the bottom row for set No. 2 (1493 rainfalls)



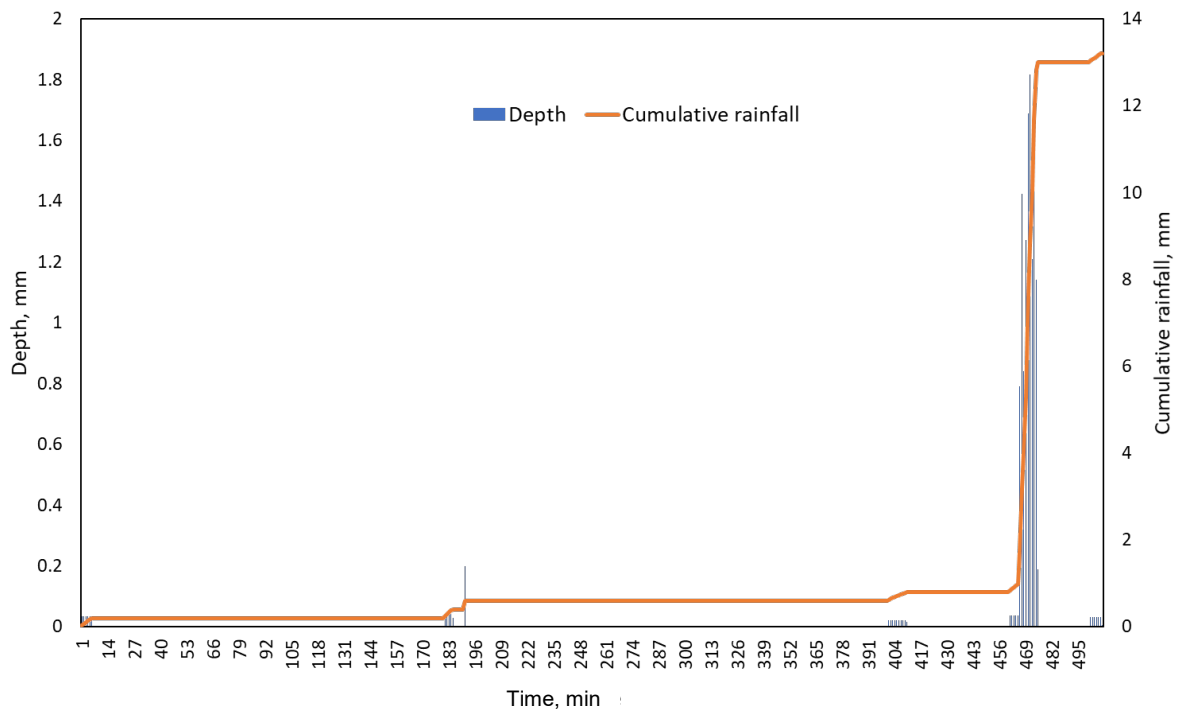


**Figure 4 a.** Dendrograms obtained for set No. 1 composed of 313 dimensionless cumulative rainfall hyetographs from Kraków (top panel). Below find the same dendrogram prepared for the reduced number of 30 leaf nodes (bottom panel). In the diagrams, vertical axes show bond distances for particular rainfalls and rainfall clusters. The horizontal axis of the dendrogram on the bottom panel shows numbers of rainfalls in particular clusters; **b.** Dendrogram obtained for set No. 2 composed of 1494 dimensionless cumulative rainfall hyetographs from Kraków. The dendrogram was prepared for the reduced number of 30 leaf nodes. The vertical axis of the diagram shows bond distances for particular rainfall clusters, and the horizontal axis shows their numbers

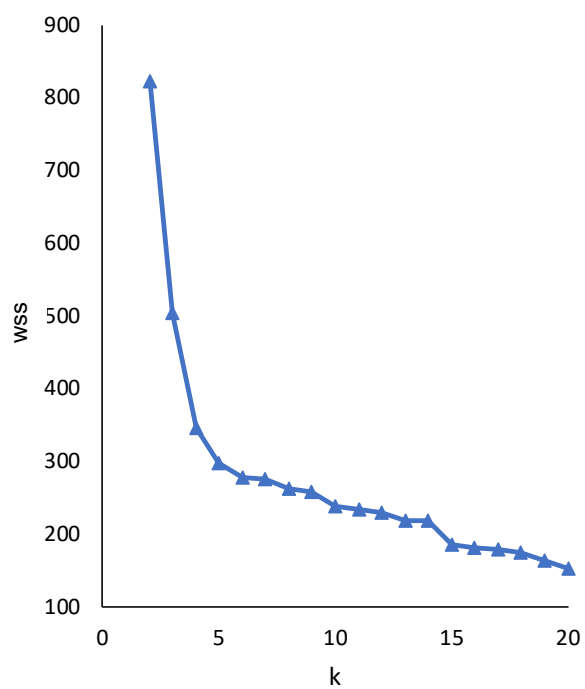
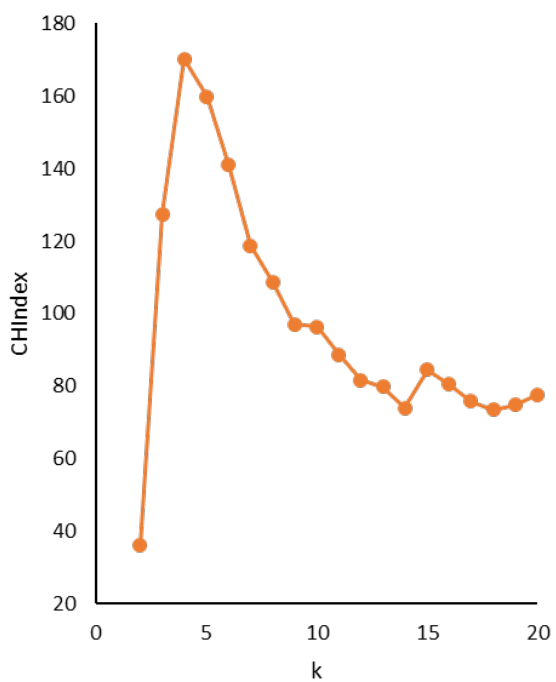


**Figure 5 a.** Dendrogram obtained for adjusted set No. 2 composed of 1493 dimensionless cumulative rainfall hietographs from Kraków. The dendrogram was prepared for the reduced number of 30 leaf nodes. The vertical axis of the diagram shows bond distances for particular rainfall clusters, and the horizontal axis shows their numbers; **b.** Dendrogram obtained for adjusted set No. 3 composed of 1806 dimensionless cumulative rainfall hietographs from Kraków. The dendrogram was prepared for the reduced number of 30 leaf nodes. The vertical

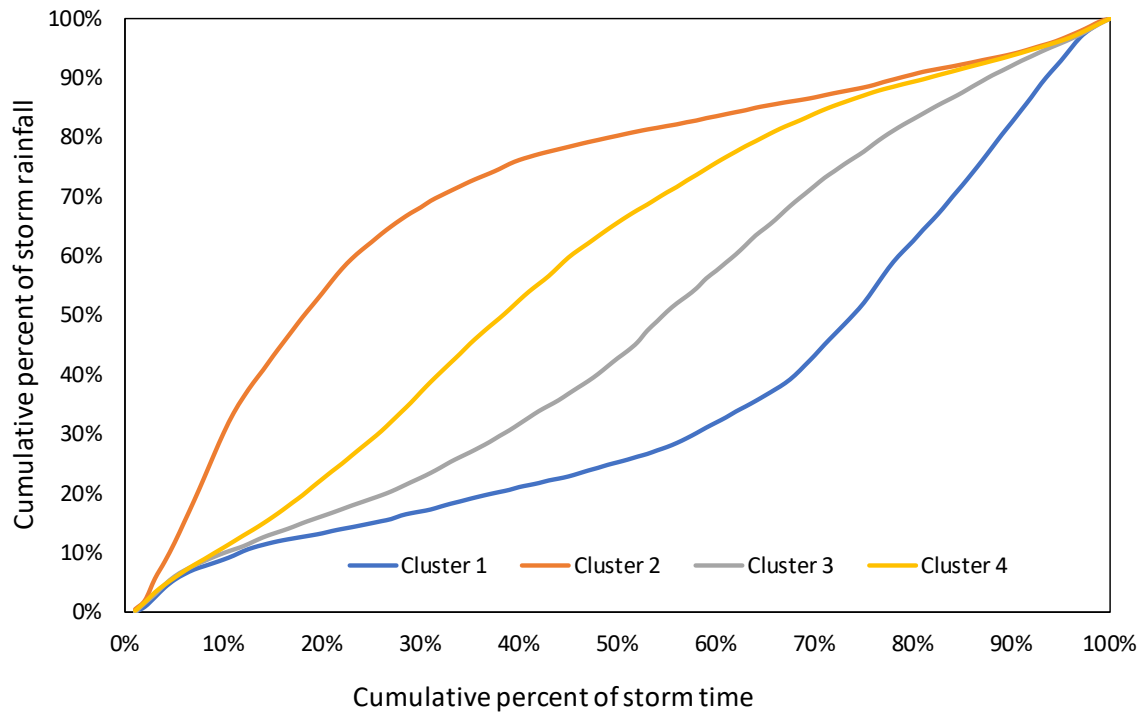
axis of the diagram shows bond distances for particular rainfall clusters, and the horizontal axis shows their numbers



**Figure 6.** Hyetograph of a rainfall event recorded on station Kraków-Płaszów in 2014

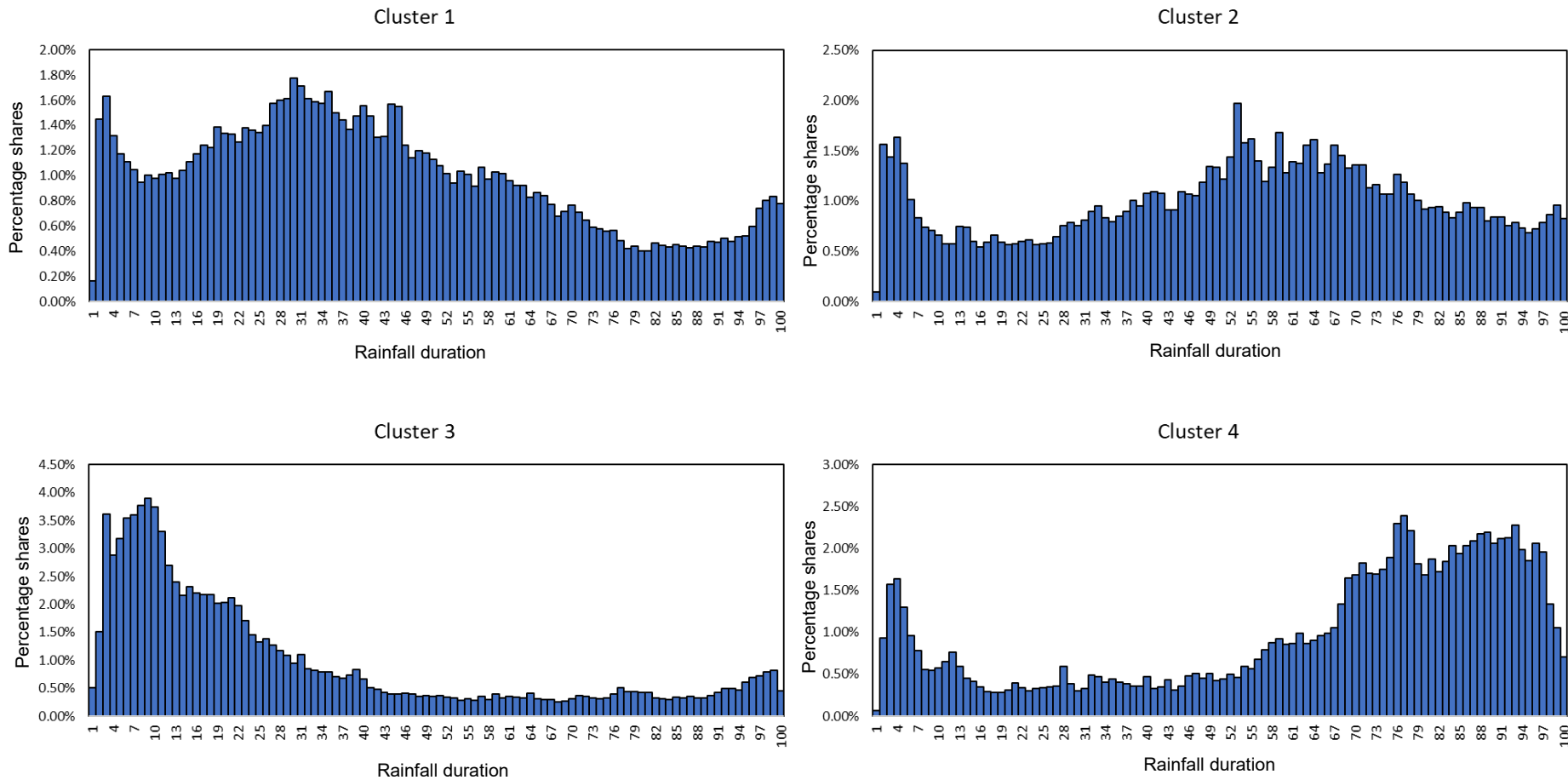


**Figure 7.** Value of the CHIndex and total within sum of squares (wss), for a set 1 of 313 rainfalls from Kraków, depending on the adopted number of clusters  $k$

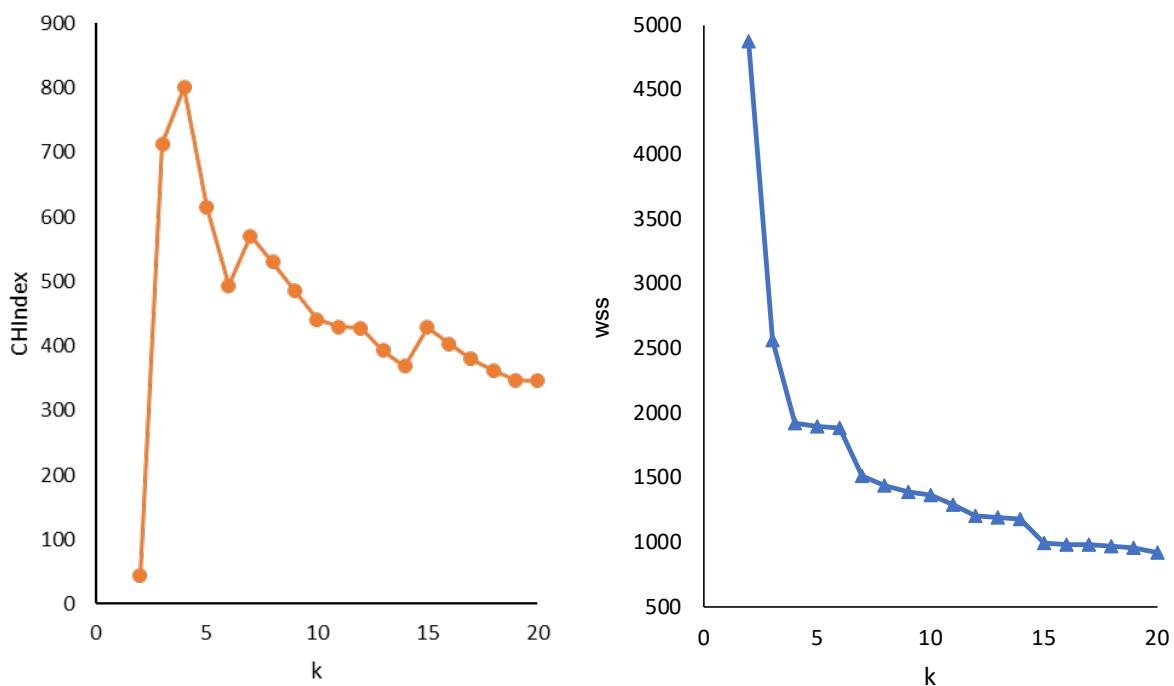


**Figure 8.** Diagrams of averaged dimensionless cumulative rainfall hyetographs for four clusters designated by means of k-means clustering for Kraków based on set No. 1 for 313 storm rainfalls

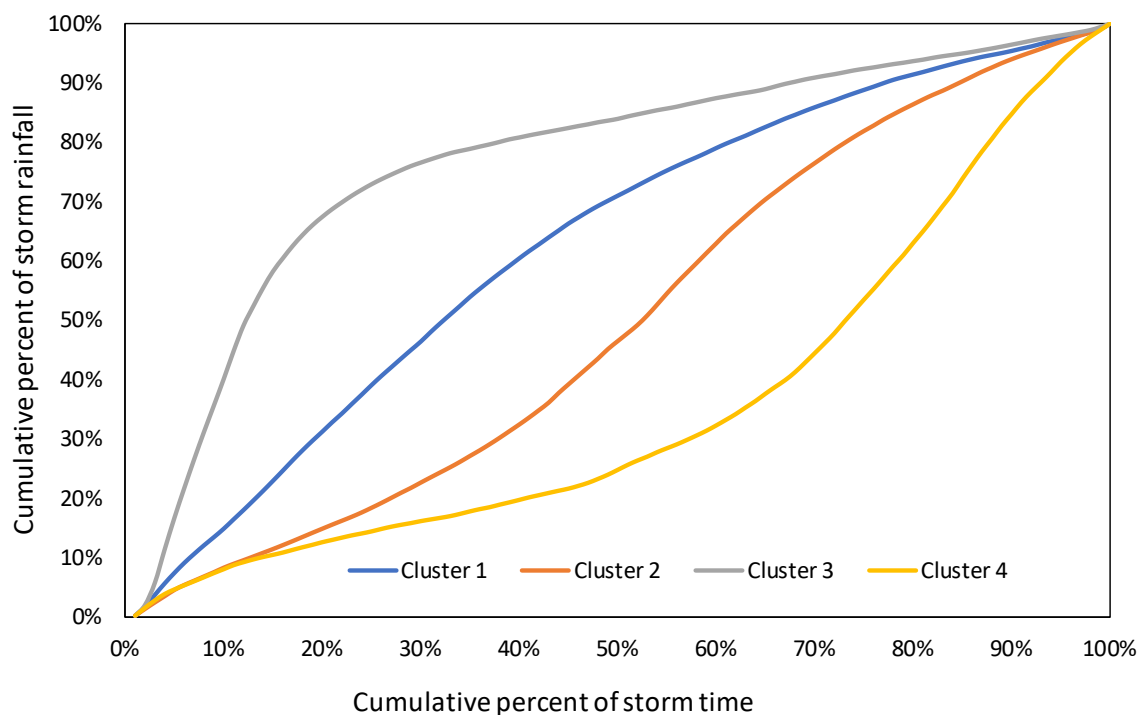




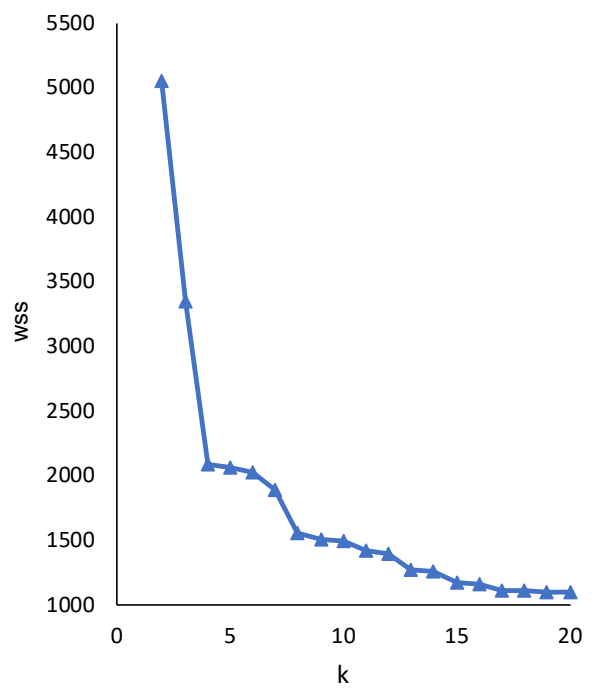
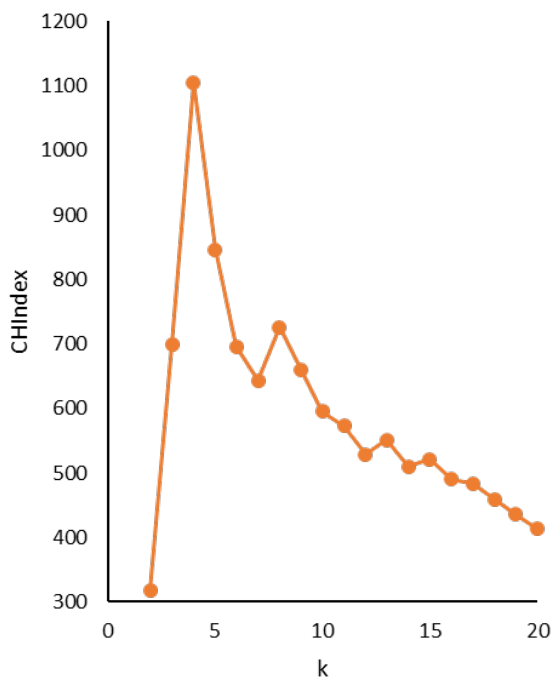
**Figure 9.** Model dimensionless rainfall hyetographs developed by means of the k-means clustering method for set No. 1 for 313 rainfalls from Kraków. The horizontal axis shows percent increase in rainfall duration, and the vertical axis shows percent shares in total precipitation amount



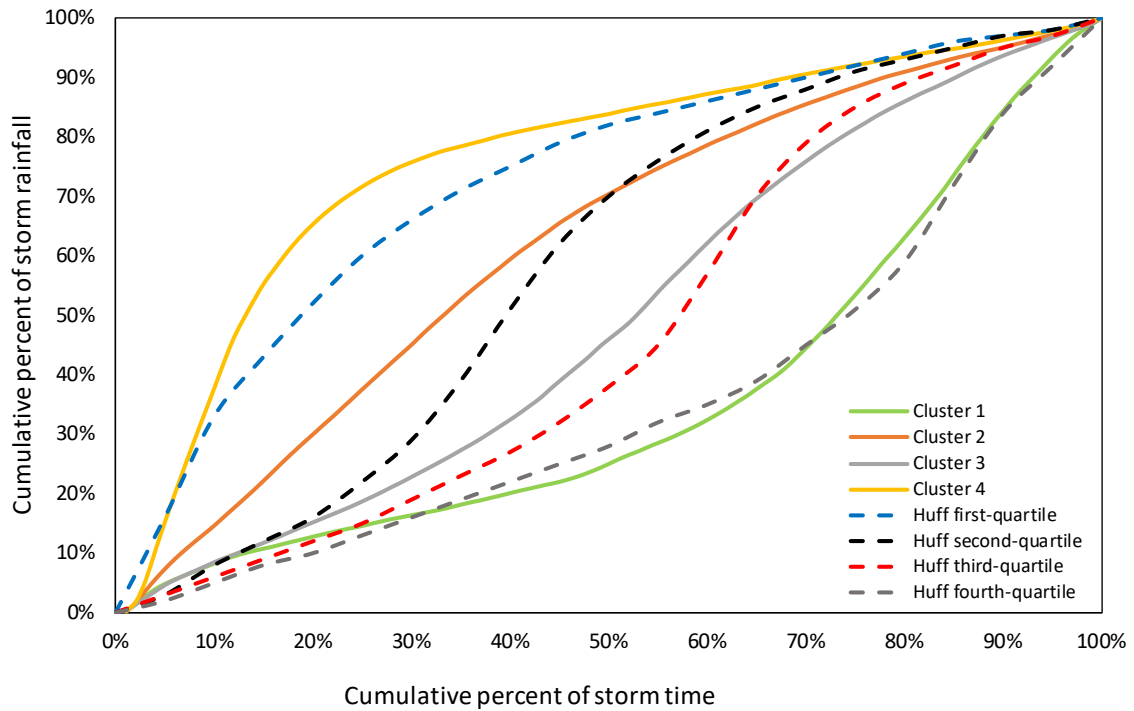
**Figure 10.** Value of the CHIndex and total within sum of squares (wss), for set No. 2 of 1493 rainfalls from Kraków, depending on the adopted number of clusters k



**Figure 11.** Diagrams of averaged dimensionless cumulative rainfall hyetographs for four clusters designated by means of the k-means clustering method for Kraków based on set No. 2 for 1493 storm rainfalls

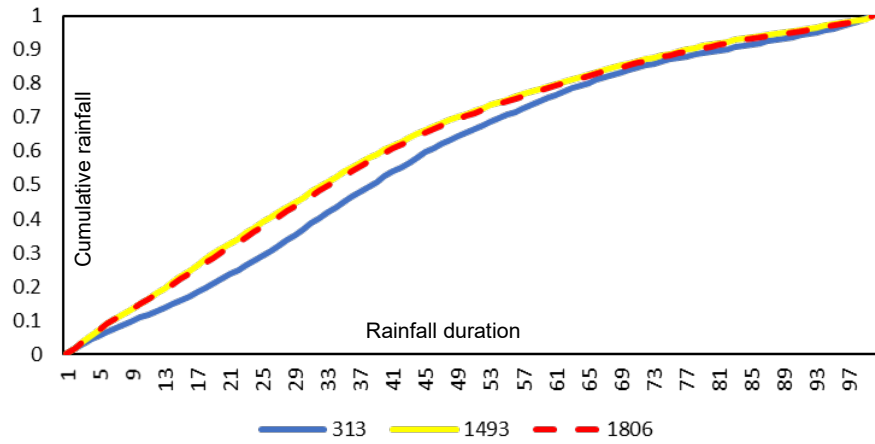


**Figure 12.** Value of the CHIndex and total within sum of squares (wss), for a set 3 of 1806 rainfalls from Kraków, depending on the adopted number  $k$  of clusters

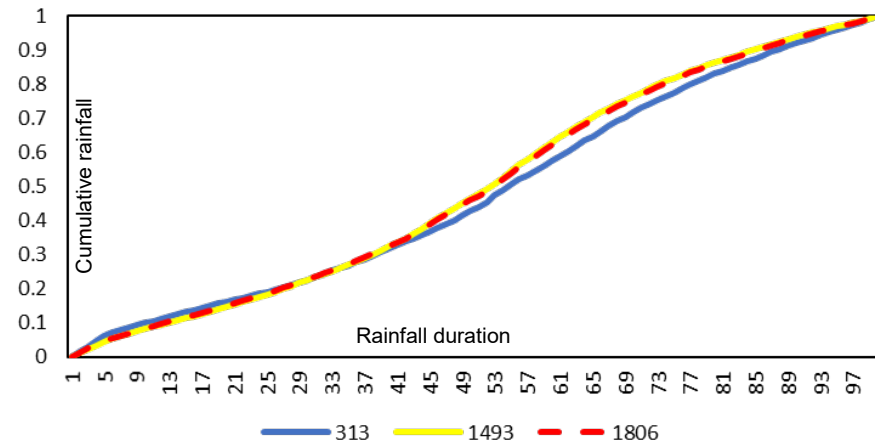


**Figure 13.** Diagrams of averaged dimensionless cumulative rainfall hyetographs for four clusters designated by means of the k-means clustering method for Kraków based on set No. 3 for 1806 storm rainfalls. For comparison, the diagram also shows Median Time Distributions of Heavy Storm Rainfall at a Point developed by Huff (1990)

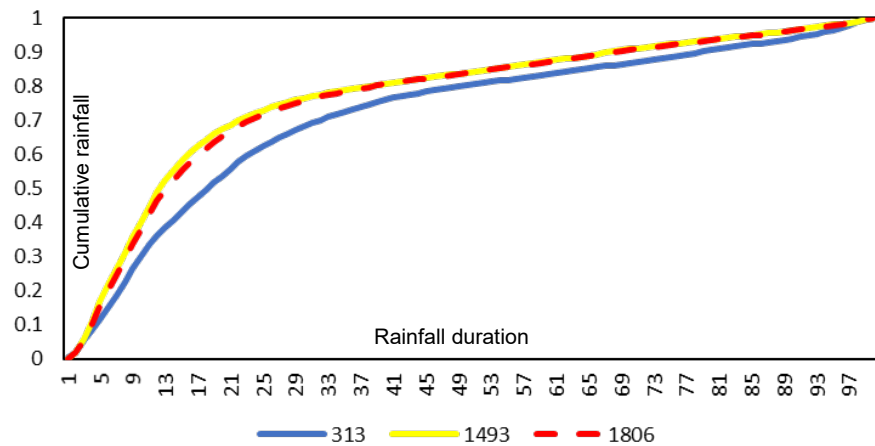
Cluster 1



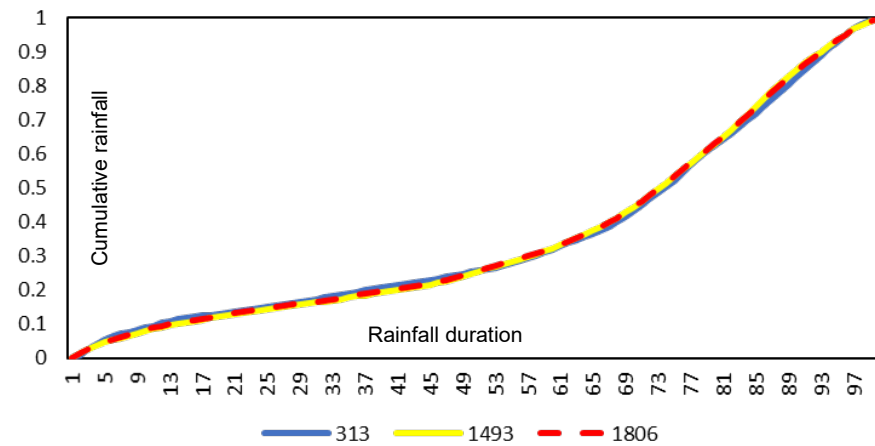
Cluster 2



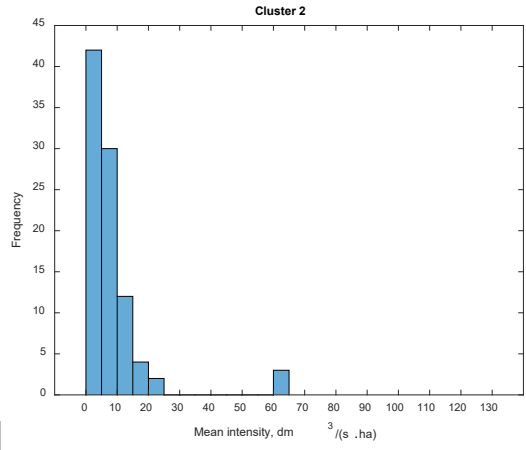
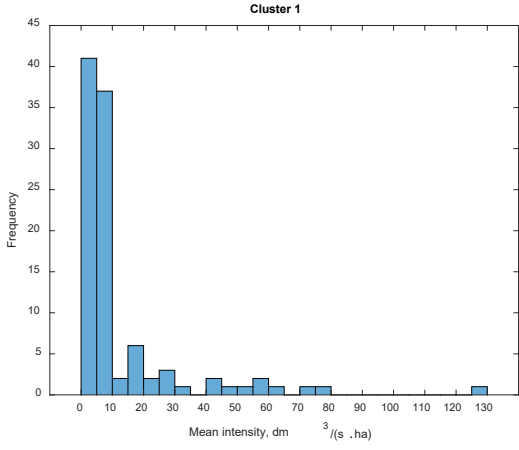
Cluster 3



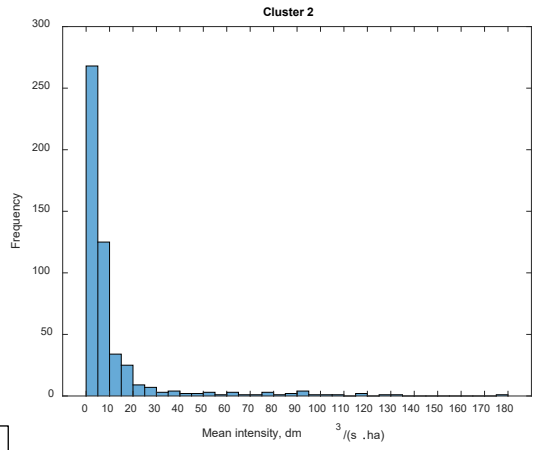
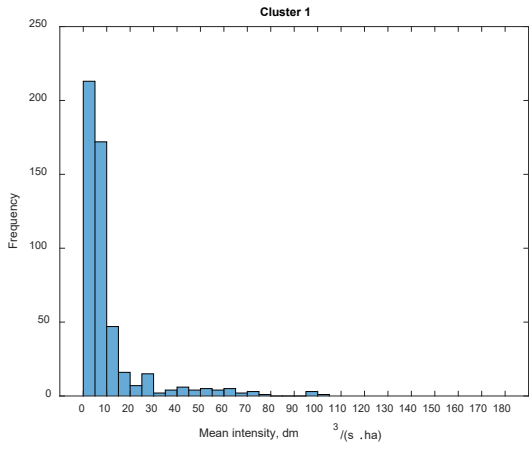
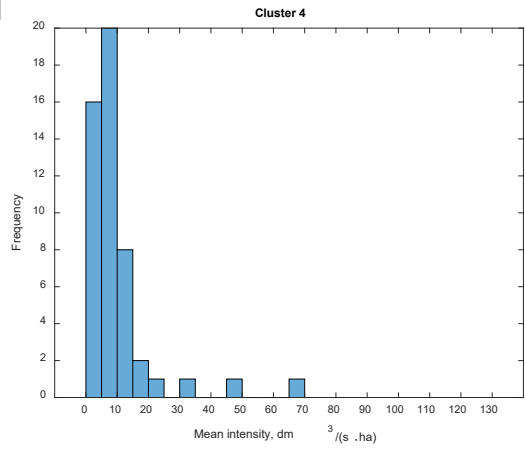
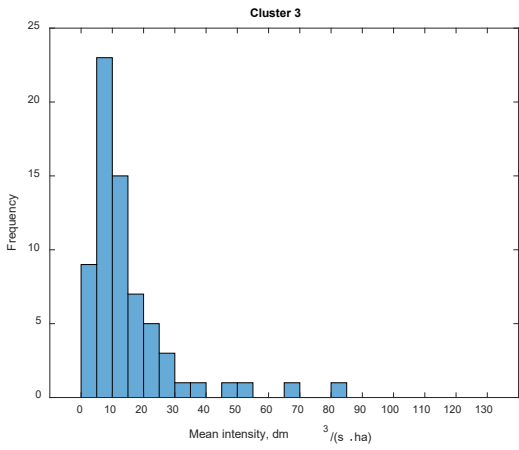
Cluster 4



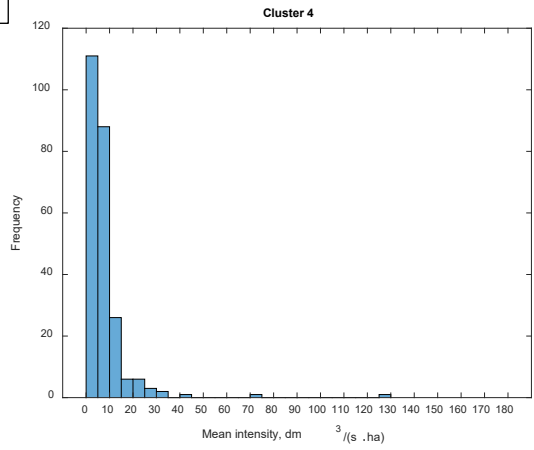
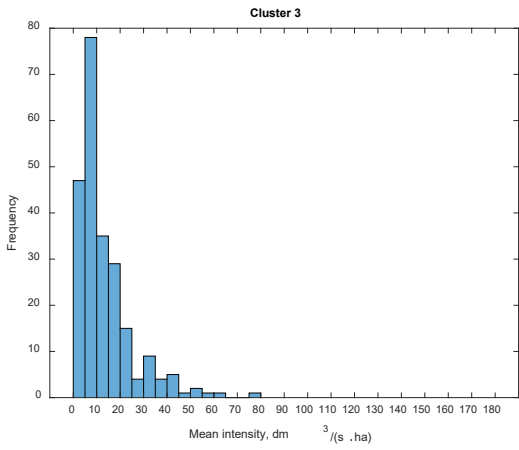
**Figure 14.** Comparison of averaged dimensionless cumulative rainfall hyetographs for four clusters designated by means of the k-means clustering method based on sets No. 1, 2, and 3: 313, 1493, and 1806 storm rainfalls recorded in Kraków. The horizontal axes present percent increase in rainfall duration, and the vertical axes show cumulative rainfall depth



a



b





**Figure 15 a.** Histograms of mean rainfall intensities for four designated clusters in rainfall set No. 1 from Kraków (313 rainfalls); **b.** Histograms of mean rainfall intensities for four designated clusters in rainfall set No. 2 from Kraków (1493 rainfalls)

**Table 1.** List of rain gauges belonging to sets No. 1 and 2 with characteristics of designated storms

No.	Rain gauge location	Observation period	Number of storm rainfalls	Minimum duration, min.	Maximum duration, min.	Minimum precipitation amount, mm	Maximum precipitation amount, mm
Set No. 1							
1	Kraków – Balice	1986 – 2006	313	24	2329	10.0	105.0
2	Kraków – Wola Justowska	2007 – 2015					
Set No. 2							
1	Bełzy	2016 ÷ 2018	36	89	1639	10.6	74.0
2	Chabrowa	2014 ÷ 2018	41	22	2206	10.2	52.8
3	Miedziana	2010 ÷ 2018	128	20	4359	10.15	121.6
4	Jeziorany	2013 ÷ 2018	86	59	2615	10.0	81.2
5	Kampus UJ	2014 ÷ 2018	67	14	1940	10.2	67.6
6	Kostrze	2013 ÷ 2016	33	49	2377	10.4	43.8
7	Krzemionki	2015 ÷ 2018	59	39	1927	10.0	68.2
8	Kujawy	2013 ÷ 2016	40	32	2258	10.2	93.6
9	Lindego	2008 ÷ 2018	106	21	3015	10.0	60.0
10	Narciarska	2014 ÷ 2018	49	17	2324	10.0	76.0
11	Olsza	2011 ÷ 2012	11	38	1065	10.4	19.4
12	Płaszów	2011 ÷ 2018	100	26	2499	10.0	69.6
13	Reduta	2015 ÷ 2018	49	47	2517	10.0	62.4
14	Rybna	2015 ÷ 2018	43	22	1697	10.2	61.8
15	Rzepichy	2012 ÷ 2018	80	22	2500	10.0	57.2
16	Senatorska	2013 ÷ 2018	81	34	2134	10.2	67.2
17	Skotniki	2009	5	67	839	10.2	24.2
18	Stojałowskiego	2011 ÷ 2018	92	32	3239	10.2	64.8
19	Szwedzka	2008 ÷ 2018	49	21	4576	10.0	145.4
20	Wilga	2013 ÷ 2018	102	25	2518	10.0	73.6
21	Witkowice	2013 ÷ 2018	60	31	2121	10.2	86.0
22	Żagłowa	2012 ÷ 2018	86	22	2405	10.0	82.4



23	Zawiła	2013 ÷ 2018	91	65	2595	10.2	82.2
----	--------	-------------	----	----	------	------	------

**Table 2.** Mean values of total depths, durations, and intensities of storm rainfalls included in particular clusters for sets No. 1 and No. 2 from Kraków

Cluster	Total depth, mm	Total duration, min	Mean intensity, dm <sup>3</sup> /(s·ha)
Set No. 1			
1	20.0	577	13.0
2	19.3	598	8.5
3	19.5	348	14.7
4	18.8	505	9.7
Set No. 2			
1	19.6	665	10.8
2	20.4	712	11.3
3	17.5	365	14.2
4	19.9	668	7.7