Journal of Hydrology 525 (2015) 345–351

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

The suitability assessment of a generalized exponential distribution for the description of maximum precipitation amounts

Bartosz Kaźmierczak *, Andrzej Kotowski

Wroclaw University of Technology, Faculty of Environmental Engineering, pl. Grunwaldzki 9, 50-377 Wroclaw, Poland

ARTICLE INFO

Article history:
Received 8 October 2011
Received in revised form 9 March 2015
Accepted 30 March 2015
Available online 6 April 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Alin Andrei Carsteanu, Associate Editor

Keywords:
Rainfall
Probabilistic models
Generalized exponential distribution
Maximum likelihood method

SUMMARY

The paper is a methodological extension of the current description of maximum precipitation amounts on the basis of gamma, Gumbel, lognormal or Weibull distribution to newly developed theoretical distributions, namely, the two-(GED2) and three-parameter (GED3) generalized exponential distribution. The verification is carried out on the basis of meteorological data from the Wroclaw–Strachowice meteorological station of the Institute of Meteorology and Water Management from years 1960–2009.

1. Introduction

The primary form of the rainfall quantitative description are models based on the relation of: precipitation amount (h) or intensity (I) to its duration (t) and probability of exceeding (p). The relations are most often presented in the form of DDF type curves (Depth–Duration–Frequency) or IDF type curves (Intensity–Duration–Frequency) for various probabilities p or, alternatively, frequencies $C = 1/p$. Historically, the first precipitation models were created on the basis of rainfall height measurements by means of simple recorders, such as Hellmann's rain gauge. Talbot was the precursor in the field, who developed precipitation reference curves as early as in 1899. The modern precipitation models are based on data digitalization read from pluviograms in specified intervals of time (most frequently from 5 min to 72 h) for a real beginning and end of precipitation – calculated by means of a moving sum method.

Rainfall curves of an IDF or DDF type are artificial creations. They are the basis to form the so called block precipitation – of a constant intensity value, which, in turn, is the basis for storm water system or combined sewerage system dimensioning using the so called flow time methods (Schmitt, 2000). The European standard EN 752:2008 recommends the following rainfall occurrence frequencies C (repeatability in years) for the designing of drainage areas: once a year in case of rural areas and once in 2–10 years for municipal areas – respectively for spatial development types. It is currently recommended in such dimensioned, larger drainage systems (area $F > 2\ \text{km}^2$) to verify the frequency of outflows using hydrodynamic modeling – with different rainfall loading scenarios, varied both in time and space. The scenarios include actual measured series of local precipitations of many years, or a model rainfall, such as Euler's type II – created from IDF or DDF curves (Schmitt, 2000, Kotowski, 2011, Kotowski and Kaźmierczak, 2013), or random generated precipitations (Mehrotra and Sharma, 2007a,b, Licznar et al., 2011). The quoted standard restricts the frequency of outflows from sewerage systems (or the impossibility to collect storm water) to a seldom occurrence repeatability: once per 10 to 50 years.

Thus, a safe dimensioning of sewerage systems (networks and facilities) aims at ensuring a proper standard for area drainage, which is defined as a system adaptation to collect the maximum forecast precipitation water jets with a frequency equal to a permissible (socially accepted) frequency of outflows onto an area surface. Because of the uncertainty of current prognoses (Kaźmierczak and Kotowski, 2014; Kaźmierczak et al., 2014), sewerage systems should be tested in terms of current extreme precipitation with the frequency of once per 50 or 100 years (Siekmann and Pinnekamp, 2011; Staufer et al., 2010; Willems, 2011).

---

* Corresponding author.
E-mail addresses: bartosz.kazmierczak@pwr.edu.pl (B. Kaźmierczak), andrzej.kotowski@pwr.edu.pl (A. Kotowski).

http://dx.doi.org/10.1016/j.jhydrol.2015.03.063

Therefore, systematic precipitation investigations focused on the determination of empirical and probabilistic frequencies of maximum height (intensities) occurrences on a specific municipal drainage area has now become of utmost importance (Kotowski and Kaźmierczak, 2013).

The aim of the paper was to assess the usefulness of a newly developed generalized exponential distribution (Gupta and Kundu, 1999, 2000, 2007) to describe the rainfall maximum values. The study based on an example of measurement data recorded in Wroclaw (in Poland), compares the proposed distribution to lognormal, Gumbel, gamma and Weibull distributions commonly used in the description of the maximum rainfall.

2. Research material and investigation methods

Archival pluviograms from Wroclaw–Strachowice meteorological station of the Institute of Meteorology and Water Management from years 1960–2009 (50 years) constituted the research material. The station coordinates: 51–06 N, 16–54 E; terrain altitude: about 120 m above sea level. Precipitations were recorded by means of a float pluviograph until 2006, whereas the automatic rain gauge RG-50 SEBA featuring electronic recording was used from 2007 (Kotowski et al., 2010). The average annual precipitation amount at Wroclaw–Strachowice station in the period of 1960–2009 amounted 568 mm. The lowest annual sum amounted 380.8 mm in 1982, whereas the highest annual precipitation was recorded in the year 1977–776.2 mm. The number of precipitation days varied from 118 in the year 2003 to 190 in the year 1970 (on average 157 days).

Precipitation amounts were determined for the following 16 intervals of durations (t): 5, 10, 15, 30, 45, 60, 90, 120, 180, 360, 720, 1080, 1440, 2160, 2880 and 4320 min. From the period of 50 years of observations the 50 highest precipitation amounts were selected (for each of the 16 intervals). The period of analysis of archival pluviograms was limited to the months from May to October (V–X). At 63 meteorological stations in Poland in the 30-year period 1961–1990, the largest depth of daily rainfall occurred from November to April only occasionally, on average 2 times per 30 years and were much lower than the average of the highest daily rainfall (Bogdanowicz and Stachý, 1998).

To determine the probability distribution of a random variable, the empirical non-exceeding probability should be assigned to particular sample elements (Table 1):

$$p(m,n) = \frac{m}{n+1} \tag{1}$$

where m – a number of element in an nonincreasing sequence: $m = 1, \ldots, 50$; n – the sample size ($n = 50$).

Measurement data was depicted by four commonly used theoretical distributions, i.e. the gamma, Gumbel, lognormal and Weibull (Di Baldassarre et al., 2006; Ben-Zvi, 2009; Brath et al., 2003; Kotowski and Kaźmierczak, 2013; Kottegoda et al., 2000; Overeem et al., 2008), and also generalized exponential distribution (Gupta and Kundu, 1999) previously unexploited in precipitation data description. Generalized exponential distribution, was proposed as an alternative to most widely used distributions, characterized by distribution function as follows – in the case of two-parameter distribution:

$$F(x;\alpha,\lambda) = \left(1 - e^{-\lambda x}\right)^{\alpha} \tag{2}$$

where α – the shape parameter, $\alpha > 0$, λ – the scale parameter, $\lambda > 0$.

The density function assumes the form of a formula:

$$f(x;\alpha,\lambda) = \alpha\lambda\left(1 - e^{-\lambda x}\right)^{\alpha-1} e^{-\lambda x} \tag{3}$$

Quantiles of a random variable for the GED2 described by the equation:

$$x = -\frac{1}{\lambda}\ln\left(1 - (1-p)^{\frac{1}{\alpha}}\right) \tag{4}$$

The log-likelihood function for a two-parameter generalized exponential distribution (GED2) assumed the form (Gupta and Kundu, 2000, 2007):

$$\ln L(\alpha,\lambda) = n\ln\alpha + n\ln\lambda - \lambda\sum_{i=1}^{n} x_i + (\alpha-1)\sum_{i=1}^{n}\ln\left(1 - e^{-\lambda x_i}\right) \tag{5}$$

where x_i – ordered random sample ($x_1 \geqslant x_2 \geqslant \ldots \geqslant x_n$).

Considering the lower limit μ ($\mu < x$), the log-likelihood function for a three-parameter generalized exponential distribution (GED3) assumes the form of:

$$\ln L(\alpha,\lambda,\mu) = n\ln\alpha + n\ln\lambda - \sum_{i=1}^{n}\lambda(x_i - \mu) + (\alpha-1)\sum_{i=1}^{n}\ln\left(1 - e^{-(x_i-\mu)\lambda}\right) \tag{6}$$

Quantiles of a random variable for the GED3 described by the equation:

$$x = \mu - \frac{1}{\lambda}\ln\left(1 - (1-p)^{\frac{1}{\alpha}}\right) \tag{7}$$

Estimators of the parameter.

The parameter estimates in formulas (4) and (7) were determined by numerical maximization of the log-likelihood function (5) and (8), taking into account the range of variability of parameters.

The Anderson–Darling test (Hongjoon et al., 2012) was carried out in order to compare the fit of the theoretical distributions to empirical data. The statistic of the Anderson–Darling test is determined from the formula:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln F(X_i) + \ln\left(1 - F(X_{n-i+1})\right)\right] \tag{8}$$

where X_i – ith element in an nondecreasing sequence, $F(x)$ – is the cumulative distribution function of the specified distribution.

The hypothesis regarding the distributional form is rejected at the chosen significance level if the test statistic, A^2, is greater than the critical value.

Table 1

Selected time series of maximum precipitation amounts in Wroclaw (in an nonincreasing sequence).

m	p (m,n)	t, min															
		5	10	15	30	45	60	90	120	180	360	720	1080	1440	2160	2880	4320
1	0.020	13.1	18.7	24.7	32.9	34.7	35.3	42.7	57.7	61.9	63.1	64.2	72.9	80.1	92.6	103.9	116.9
2	0.039	11.6	18.0	22.8	30.3	34.7	35.3	37.7	41.5	42.8	50.4	64.2	71.5	77.9	92.5	103.2	111.6
5	0.098	9.9	15.7	20.1	28.2	32.1	34.7	35.4	36.2	38.4	43.9	54.2	69.1	72.2	85.4	94.5	101.9
10	0.196	9.3	13.8	17.7	26.7	28.8	30.5	33.9	35.4	35.7	38.7	49.2	57.4	65.0	73.1	76.2	87.5
20	0.392	8.3	12.3	14.6	19.7	22.8	24.6	28.7	29.5	32.4	36.3	41.5	48.9	51.5	57.0	61.1	68.2
50	0.980	6.4	8.9	10.1	13.7	14.8	15.3	16.3	17.9	20.0	26.2	32.0	36.5	39.9	45.2	48.1	49.0