STATISTICAL ANALYSIS OF METEO-DROUGHTS FOR KIKINDA BASED ON THE CARPAT CLIM DATA

Ognjen Gabrić[1]
Jasna Plavšić [2]
UDK: 551.577.38:519.21
DOI: 10.14415/konferencijaGFS2019.065
Summary: *In order to overcome the lack of long-term meteorological dataset, European Commision launched the CARPATCLIM project – climate atlas of the Carpathian region with the daily basis meteorological data. This paper presents comparison of the results of stochastic analysis of meteorological droughts occurrence in a growing season based on the data from meteorological station of Kikinda and the data from CARPATCLIM database which covers the region of Kikinda meteorological station. In this paper, meteorological drought is defined as a product of drought length (period lasting at least 25 days with precipitation below 5 mm/day) and mean air temperature during the dry period. Stochastic model of the meteo-drought occurrence is based on the theory of extremes with a random number of random variables, allowing to make use of all significant observed meteo-droughts.*

Keywords: *carpatclim, drought, growing season, duration of drought, mean air temperature of meteo-drought, design meteo-drought, Kikinda*

1. INTRODUCTION

Historical data records are important for water resources management and modelling, climate change assessment, flood modelling and prediction, and other hydrological activities [1]. Hydrological data are costly, in terms of both effort and resources, to record and collect.

To overcome the lack of a dense long-term dataset based on in situ measurements over the Carpathians, the European Commission (EU) launched and financed the CARPATCLIM project in 2010 [2]. The creation of a gridded (0.1°×0.1°) Climate Atlas of the Carpathian Region on a daily basis, is completed by participants from nine countries and ten institutions (Hungarian Meteorological Service (leading organization); Central Institute for Meteorology and Geodynamics, Austria; Meteorological and Hydrological Service of Croatia; Czech Hydrometeorological Institute; Institute of Meteorology and Water Management - National Research Institute, Poland; National Institute for Research and Development in Environmental Protection of Romania Republic Hydrometeorological Service of Serbia; Slovak Hydrometeorological Institute; Ukrainian Research

[1] Ognjen Gabrić, PhD, CE, University of Novi Sad, Faculty of Civil Engineering Subotica, Kozaračka 2a, Subotica, e-mail: ogabric@gf.uns.ac.rs
[2] Jasna Plavšić, PhD, CE, University of Belgrade, Faculty of Civil Engineering Belgrade, Bulevar kralja Aleksandra 73, Belgrade, e-mail: jplavsic@grf.bg.ac.rs

Hydrometeorological Institute; Szent Istvan University, Hungary) [3 www]. Together with the European Union's Joint Research Center (JRC), Climate Atlas of the Carpathian Region, based on a robust dataset made of homogenized daily data of sixteen variables and many derived indicators was made [2]. More information about CARPATCLIM and daily data available can be found at the website http://www.carpatclim-eu.org/pages/home/.

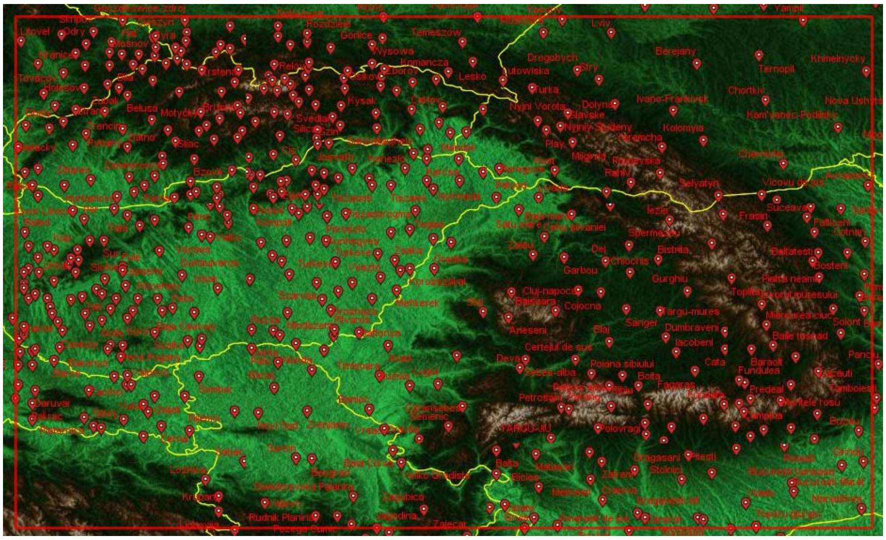

Figure 1: Region covered by CARPATCLIM project [3]

In order to test reliability of CARPATCLIM data for calculation of meteo-droughts, result of previously calculated meteo-droughts for meteorological station of Kikinda [4] were compared with the result obtained by CARPATCLIM data using the same methodology.

2. DEFINITION OF METEO-DROUGHTS

The meteo-drought, Z (day · ^{0}C), is defined as the product of drought duration, T (days), and mean air temperature over the drought duration y (^{0}C) [4]:

$$Z = T \cdot y \tag{1}$$

The drought duration is the duration of a dry spell in which days with precipitation smaller than 5 mm are treated as dry days because such a small amount of precipitation is insignificant for the roots of most crops. Drought duration of 25 and more days is adopted as a critical threshold, for which the distribution of meteo-droughts is estimated and design droughts with different probabilities of occurrence are derived.

According to this method, the stochastic process of the meteo-droughts is defined with 10 characteristic parameters [5]: meteo-drought Z, drought duration T, mean air temperature y throughout the drought duration, beginning of meteo-drought τ_b, end of meteo-drought τ_e, mid-point $\tau = (\tau_b + \tau_e)/2$, ordinal number of meteo-drought in growing season n ($n = 1, 2, 3,...$), total number of meteo-droughts in growing season k ($k = 0, 1, 2,...$), the greatest meteo-drought in growing season $sup\{Z_n\}$, time of occurrence of the greatest meteo-drought τ_{supZ}. Occurence of meteo-droughts in a growing season and their describing parameters are shown schematically in Figure 2.

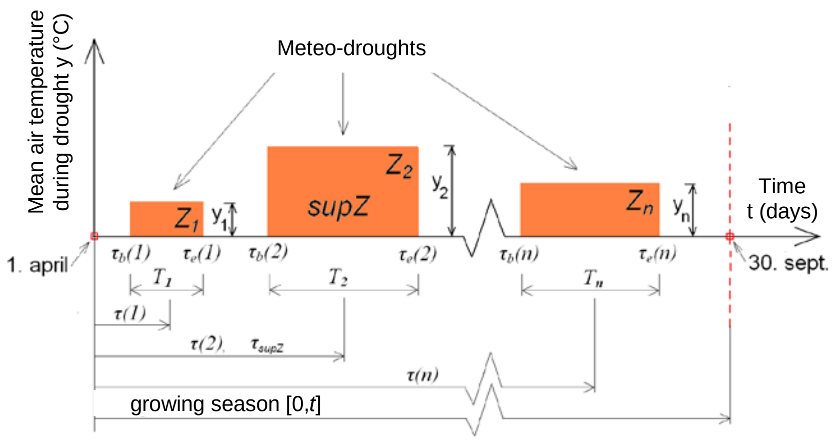

Figure 2: Occurrence of meteo-droughts during a growing season and their describing parameters [5]

3. METHODOLOGY

The methodology for developing distributions of the meteo-droughts includes the following steps: (1) preparatory analysis, in which statistical tests are applied to the available data sets to test whether they are suitable for statistical analysis, (2) identification of the distribution of meteo-droughts Z, (3) identification of distributions of drought duration T and mean temperature y, and (4) construction of design droughts.

3.1. Tests for randomness and homogeneity

The sample for the statistical analysis is created by extracting the drought events and their parameters during the growing season in each year of observations. The threshold for the drought duration is set at 25 days. Each sample should represent a sample of independent and identically distributed variables and the appropriate statistical tests need to be undertaken to confirm these assumptions.
The samples are tested for homogeneity by two nonparametric tests (Mann-Whitney and Kolmogorov-Smirnov tests). The runs test is applied to test the randomness of sample data [6]. Both tests are applied at the 5% significance level.

3.2. Distribution of the meteo-droughts *F*(*z*)

Statistical analysis of the maximum meteo-droughts is performed by means of the peaks over threshold (POT) method [7], which considers all extreme values of the meteo-droughts that exceed a given threshold. Generally, this is an advantage of this method over the annual maxima method, with which only the annual maximum values are included in the statistical analysis and which neglects the fact that there can be more than one extreme value in some year that exceeds annual maxima in other years.
Statistical analysis with the POT method is performed in three steps:
1. fitting the distribution of the number of occurrences of meteo-droughts (i.e. the number of exceedances of meteo-drought magnitude over the chosen threshold) during the growing season,
2. fitting the distribution of the meteo-drought exceedances, and
3. combining the above two distributions into the distribution of the maximum meteo-drought in the growing season.

3.2.1. Distribution of the number of exceedances

The number of the exceedances of meteo-droughts, *k*, in the growing season depends on the selected threshold. Generally, the number of meteo-droughts decreases with increasing threshold and consequently the number of years in which no meteo-droughts occur in the growing season increases. After counting the exceedances of meteo-droughts over the threshold in each year, the mean number of exceedances $\bar{k}$ and its variance S_k^2 are calculated:

$$\bar{k} = \frac{1}{N}\sum_{i=1}^{N} k_i, \quad S_k^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(k_i - \bar{k}\right) \qquad (2)$$

where *N* is the number of years in the observation record, and k_i is the number of exceedances in the growing season of year *i*.
The number of exceedances is a discrete random variable. Theoretical distribution for the number of exceedances can be chosen based on the value of the dispersion index *I*, defined as:

$$I = \frac{S_k^2}{\bar{k}} \qquad (3)$$

For binomial distribution $I < 1$; for the Poisson distribution $I = 1$ (in practice the Poisson distribution is a good fit for $0.8 < I < 1.2$) and for negative binomial distribution $I > 1$.

3.2.2. Distribution of exceedances

The magnitude of meteo-droughts *Z* that exceed the chosen threshold Z_b is an exceedance and a continuous random variable defined as: $U = Z - Z_b$. Cumulative distribution function of the exceedances, *H*(*u*), is given with:

$$H(u) = H(Z - Z_b) = P\{U \leq u\} \tag{4}$$

Different theoretical distributions can be used to fit the distribution of exceedances. The most commonly used are the exponential, Weibull and general Pareto distributions. In this study, exponential and Weibull distributions are considered.
The empirical distribution of the exceedances is estimated using the Weibull plotting position formula. The goodness-of-fit between the theoretical and empirical distributions is evaluated on the basis of the Cramer-von Mieses test and the Kolmogorov-Smirnov test at the 5% significance level.

3.2.3. Distribution of the greatest meteo-drought in the growing season

The annual maximum meteo-drought over the growing season is a random variable *Z*, defined as the maximum of a random number *k* of exceedances *U* in the growing season:

$$Z = Z_b + max\{U_j; j = 1,2 \dots k\} \tag{5}$$

The cumulative distribution function of the annual maximum meteo-droughts, *F*(*z*), given with:

$$F(z) = P\{Z \leq z\} \tag{6}$$

is defined only for values above the threshold, i.e. $z > Z_b$. General expression for *F*(*z*) is derived by combining the distributions of the number and magnitude of exceedances:

$$F(z) = p_0 + \sum_{k=1}^{\infty} p_k \left[H(z - Z_b)\right]^k \tag{7}$$

Depending on the type of distribution for the number of exceedances and type of distribution for magnitude of exceedances, the above general expression can be simplified. For the Poisson-Weibull combination, the above equation reduces to:

$$F(z) = \exp\left\{-\lambda \exp\left[-\left(\frac{z-z_b}{\alpha}\right)^{\beta}\right]\right\} \tag{8}$$

and for the binomial-Weibull combination one obtains:

$$F(z) = \left[1 - p \exp\left\{-\left(\frac{z-z_b}{\alpha}\right)^{\beta}\right\}\right]^{a} \tag{9}$$

In the previous equations, λ, *p* and *a* are the parameters of the corresponding discrete distributions, and α and β are the parameters of the Weibull distribution. Equations for the combination of the same discrete distributions with the exponential distribution for the exceedances are obtained for $\alpha = 1$.

3.2.4. Empirical distribution of the greatest meteo-drought in the growing season

Over the record of N years, some years may not contain any meteo-droughts. The number of droughts within the growing season depends on the threshold set for the duration of dry spell. From the statistical point of view, a sample of N experiments in which only the values exceeding the given detection threshold are recorded, while the values below the threshold are not recorded, is referred to as the censored data sample. Distribution of the censored data sample is obtained from the total probability theorem taking into account the probability p_0 that a recorded value would be below the threshold. If N' out of N experiments yield data above the detection threshold, then:

$$p_0 = \frac{N-N'}{N} \tag{10}$$

If $G(x)$ denotes the conditional distribution obtained from N' data values above the detection threshold, the unconditional distribution is then given with:

$$F(x) = p_0 + (1-p_0)G(x) \tag{11}$$

In accordance with the above, empirical distribution of the greatest meteo-drought in the growing season is estimated like:

$$F_{emp} = p_0 + (1-p_0) \cdot F'_{emp} \tag{12}$$

where F'_{emp} is the empirical distribution of the observed droughts with duration of at least 25 days, described using the Weibull plotting position formula:

$$F'_{emp}(z_j) = \frac{j}{N'+1}, \; j = 1,2,\dots,N' \tag{13}$$

where N' is the number of uncensored data, i.e. the number of droughts with duration of at least 25 days, and j is the ordinal number of data in the ordered sample of N' data.
With defined empirical and theoretical distributions of the greatest meteo-drought, goodness-of-fit is tested using the same statistical tests used for fitting the distribution of the magnitude of exceedances.

3.3. Distributions of the drought duration *G*(*t*) and mean temperature *Φ*(*y*)

Drought duration, T, and mean air temperature, y, of a meteo-drought are also random variables. Their distributions are needed to construct design meteo-droughts for selected probability of occurrence. Analysis of maximum duration T or temperature y is performed here also by means of the peaks over threshold method, analogously to identifying the distribution of the maximum meteo-droughts Z in the growing season.

3.4. Deriving design meteo-droughts

Design meteo-droughts are meteo-droughts of a given probability of occurrence or return period n. In this paper, return periods of n = 10, 20, 50 and 100 are considered. Design meteo-droughts are constructed by finding pairs (Z_n, T_n) and (Z_n, y_n) through the following steps [2]:

1. Quantiles of n-year meteo-drought (Z_n), duration (T_n) and mean temperature during the drought (y_n) are estimated.
2. The following corresponding values are then calculated: $y_a = \frac{Z_n}{T_n}$ and $T_b = \frac{Z_n}{y_n}$. These values determine two rectangles (Figure 3), each having the area representing the n-year drought Z_n.
3. By averaging the sides of two rectangles, new rectangle ABCD is obtained representing the design n-year meteo-drought, with sides equal to:

$$T_r = \frac{1}{2y_a}[T_n y_a - T_b y_n + \sqrt{(T_b y_n - T_n y_a)^2 + 4y_a T_b Z_n}],\; y_r = \frac{Z_n}{T_r} \qquad (14)$$

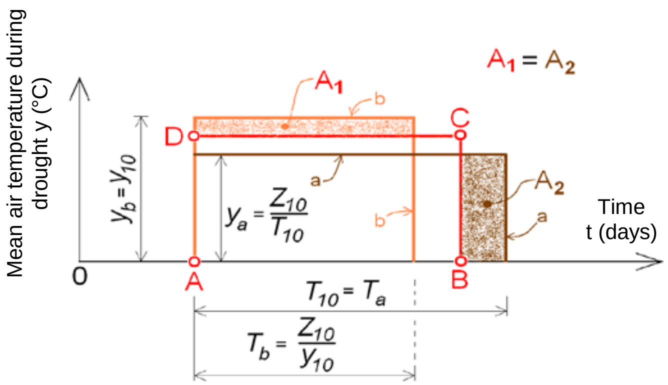

Figure 3: Construction of the design meteo-drought illustrated for the example of the 10-year drought as proposed by Zelenhasić [5]

4. RESULTS

Mean daily temperature and daily precipitation data from meteorological station of Kikinda (from 1961-2005), are compared with daily temperature and precipitation data from CARPATCLIM database for the same period and station. Differences between observed precipitation and temperature data from meteorological station and CARPATCLIM database are shown in Figure 4.
Total number of meteo-droughts based on the data from meteorological station was 62, while there was 52 meteo-droughts based on the data from CARPATCLIM.
For the homogeneity testing, samples were split in two subsamples of similar size. Both tests have shown that the null hypothesis of homogeneity of the samples cannot be rejected at 5% significance level (Table 1).

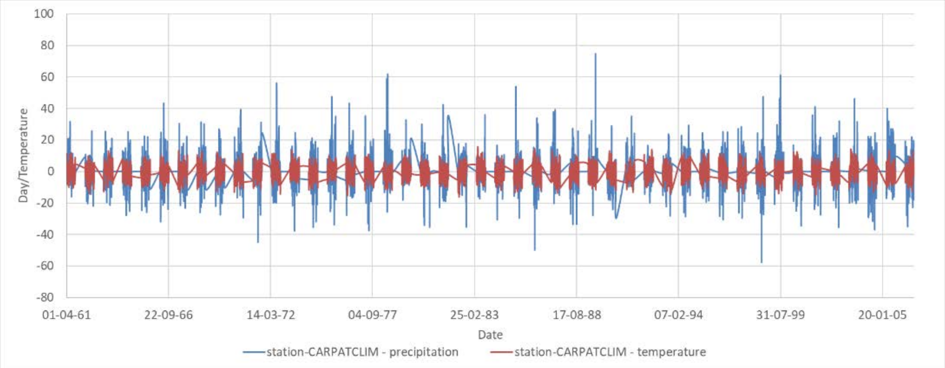

Figure 4: Differences between observed precipitation and temperature data

Table 1: Homogeneity test results

Test	Test statistic	Kikinda - CARPATCLIM	Kikinda - station
Mann-Whitney	U	-1.62	-1.94
	$u_{0.025}$	1.96	1.96
	$u_{0.975}$	-1.96	-1.96
	hypothesis accepted	H_0	H_0
Kolmogorov Smirnov	D_{max}	0.132	0.339
	D_{kr}	0.36	0.400
	hypotesis accepted	H_0	H_0

The runs test has also confirmed the assumption of randomness of data. The result of this test, at 5% significance level, is shown in Table 2.

Table 2: Runs test results

	Test statistic	Kikinda - CARPATCLIM	Kikinda - station
Runs test	k	-0.08	-0.13
	$k_{0.05}$	1.96	1.96
	$k_{0.95}$	-1.96	-1.96
	hypothesis accepted	H_0	H_0

The number of meteo-droughts greater than chosen threshold during the growing season represents a discrete variable. Dispersion index for all samples in this study was smaller than 1, so that the Poisson or binomial distribution were adopted as the distribution of the number of exceedances. The binomial distribution showed better fit with the empirical distributions of the observed samples. The results of the goodness-of-fit tests are shown in Table 3.

The best fit between the empirical and theoretical distributions was obtained for meteo-droughts exceeding threshold of Z_b = 450 day°C. Goodness-of-fit testing is carried out by the Cramer-von Mieses and Kolmogorov-Smirnov tests at 5% significance level. The results are shown in Table 4.

Table 3: Chi-square goodness-of-fit test for the number of exceedances (α = 5%)

Distribution	Test statistic	Kikinda - CARPATCLIM	Kikinda - station
Poisson	χ^2	8.933	16.588
	χ^2_{kr}	7.815	5.99
	hypothesis accepted	H_a	H_a
Binomial	χ^2	7.982	1.619
	χ^2_{kr}	9.488	5.99
	hypothesis accepted	H_0	H_0

Table 4: Goodness of fit results

Test	Distribution	Test statistic	Kikinda - CARPATCLIM	Kikinda - station
Cramer-von Meises		$N\omega^2_{kr}$	0.462	0.462
	exponential	$N\omega^2$	0.046	0.039
		accepted	H_0	H_0
	Weibull	$N\omega^2$	0.056	0.081
		accepted	H_0	H_0
Kolmogorov -Smirnov		D_{kr}	0.205	0.192
	exponential	D_{max}	0.078	0.063
		accepted	H_0	H_0
	Weibull	D_{max}	0.086	0.097
		accepted	H_0	H_0

The distribution of the maximum meteo-drought is obtained by combining the distributions of number of exceedance and level of exceedance over the threshold. In this paper combinations of the binomial with Weibull and binomial with exponential distribution are used. The derived distribution of the maximum of meteo-drought is compared to the observed annual maxima, while there were also years with no meteo-droughts lasting at least 25 days. Therefore, the model of conditional probability was used to establish the empirical distribution of the annual maximum meteo-droughts. Distributions of maximum meteo-droughts are shown in Figure 5.

A necessary condition for the construction of design meteo-drought is finding the distribution of dry period duration *(T)* and the distribution of the mean air temperature *(y)*. The POT method is applied in the same manner as for the meteo-droughts. Adopted threshold for the duration is T_b = 24 days and for the mean air temperature y_b = 9.5°C (Figures 6 and 7).

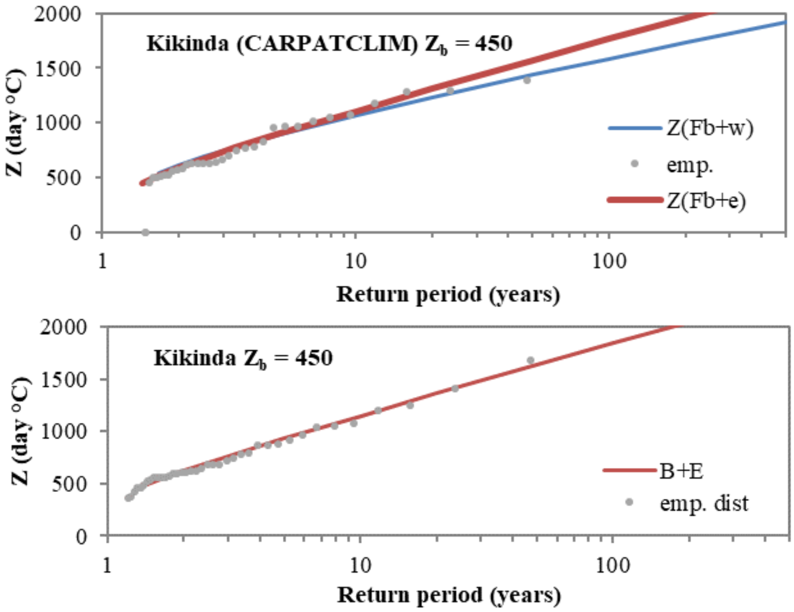

Figure 5: Distributions of maximum meteo-droughts for Kikinda based on CARPATCLIM data (top) and station data (bottom)

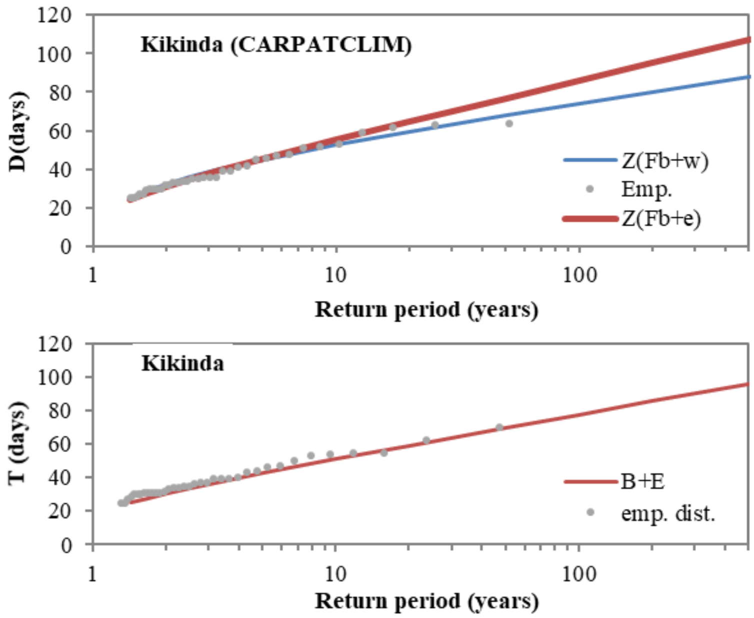

Figure 6. Distributions of maximum meteo-drought duration for Kikinda based on CARPATCLIM data (top) and station data (bottom)

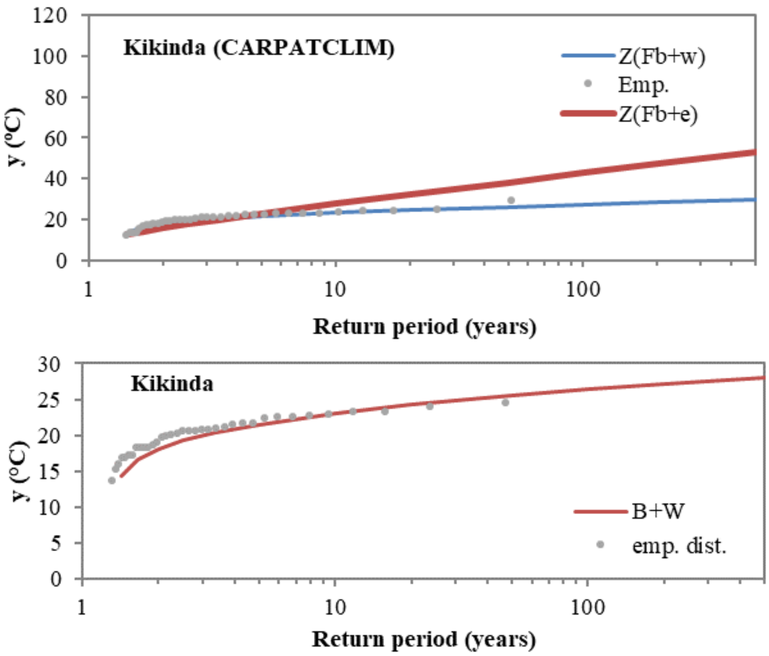

Figure 7. Distributions of mean air temperatures during meteo-drought for Kikinda based on CARPATCLIM data (top) and station data (bottom)

Based on the adopted theoretical distributions of meteo-droughts, dry period duration times and mean air temperatures, design meteo-droughts for return periods of 10, 20, 50 and 100 years are determined (Table 5).

Table 5. Design meteo-droughts

		Kikinda - CARPATCLIM			Kikinda - station		
F	R_p	$T_{calc.}$ (days)	$Y_{calc.}$ (day °C)	$Z_{calc.}$ (°C)	$T_{calc.}$ (days)	$Y_{calc.}$ (°C)	$Z_{calc.}$ (day °C)
0.9	10.0	49	22.0	1068.6	47	20.0	932.9
0.95	20.0	54	22.9	1231.0	54	21.3	1148.6
0.98	50.0	60	23.9	1435.1	61	22.1	1358.9
0.99	100.0	65	24.5	1584.3	69	23.5	1633.5

5. DISCUSSION AND CONSLUSIONS

This paper analyzes the phenomenon of meteo-drought occurrence in the growing season for the meteorological station in Kikinda based on the two sets of data – data from meteorological station of Kikinda and data from CARPATCLIM database. For determination of the theoretical distributions of meteo-drought, dry period duration and the mean air temperature, peak over threshold method was applied. Combination of binomial and Weibull or exponential distribution are most suitable for describing this phenomenon. Based on the adopted theoretical distributions of meteo-droughts, dry period duration and mean air temperatures, design meteo-droughts for return periods of 10, 20, 50 and 100 years are determined.

Although there are differences between two datasets in terms of input data (which are due to spatial averaging over the one pixel of the CARPATCLIM dataset), there is a good agreement between theoretical distributions of the meteo-droughts, duration of meteo droughts and average air temperature during meteo-droughts obtained from the two data sets.

Design meteo-droughts and adopted distribution of meteo-droughts, represents the model that in a relatively simple way describes the process of occurrence of meteo-droughts over the locality of Sremska Mitrovica and Kikinda. In order to get complete picture of the occurrence and distribution of meteo-drought over the region of Vojvodina, it would be useful to use spatial data instead of point (station) data. For this purpose, it is necessary to compare the results based on two data sets for other meteorological stations in Vojvodina in order to ensure validity of the results with the spatial data.

ACKNOWLEDGEMENT

This study was funded by the Serbian Ministry for Science; project TR37010 “Stormwater Drainage Systems as Part of Urban and Traffic Infrastructure”.

References

[1] Worl Meteorological Organisation, *Guidelines for Hydrological Data Rescue*, WMO-No 1146, 2014.
[2] http://www.carpatclim-eu.org/pages/about/
[3] http://www.carpatclim-eu.org/pages/atlas/
[4] Gabrić O, Janjikopanji Đ, Plavšić J, Fabian Đ, *Statistical Analysis of Meteo-droughts for Sremska Mitrovica and Kikinda*, Conference proceedings Contemporary achievements in civil engineering 2017, 21.04.2017, Subotica,
[5] Đ. Fabian and E. Zelenhasic, “Modelling of Meteo-Droughts,” Water Resources Management, vol. 30, no. 9, pp. 3229–3246, Jul. 2016.
[6] D. Sheskin, *Handbook of parametric and nonparametric statistical procedures*. Boca Raton: Chapman & Hall/CRC, 2004.
[7] Plavšić J., *Analiza rizika od poplava pomoću prekidnih slučajnih procesa* (Flood risk analysis by discrete stochastic processes), PhD thesis, Faculty of Civil Engineering, University of Belgrade, 2005.

STATISTIČKA ANALIZA METEOROLOŠKIH SUŠA ZA KIKINDU NA BAZI PODATAKA IZ CARPATCLIM BAZE

Rezime: *Kako bi se prevazišao problem u nedostatku dugoročnih meteoroloških podataka, Evropska komisija je pokrenula pojekat CARPATCLIM – klimatski atlas karpatskog regiona sa dnevnim meteorološkim podacima. U radu su upoređene meteorološke suše sračunate za meteorološku stanicu Kikinda na bazi podataka sa same stanice i podataka iz CARPATCLIMA za region Kikinde. Meteorološka suša je definisana kao proizvod dužine trajanja suše (period duži od 25 dana sa padavinama manjim od 5 mm/dan) i prosečne temperature vazduha tokom trajanja suše. Za formiranje stohastičkog modela pojave maksimalnih meteo-suša korišćena je teorija ekstrema sa slučajnim brojem slučajnih promenljivih, koja omogućava da se u analizu uključe sve značajne osmotrene meteo-suše.*

Ključne reči: *carpatclim, suša, vegetaciona sezona, beskišni period, prosečna temperatura vazduha beskišnog perioda, računske meteo-suše, stohastički model, Kikinda*