ATMOSPHERIC SCIENCE LETTERS
Atmos. Sci. Let. **16**: 324–330 (2015)
Published online 6 February 2015 in Wiley Online Library
(wileyonlinelibrary.com) **DOI:** 10.1002/asl2.563

Royal Meteorological Society

Complexity analysis of spatial distribution of precipitation: an application to Bosnia and Herzegovina

Dragutin T. Mihailović,[1]* Nusret Drešković[2] and Gordan Mimić[3]

[1] *Faculty of Agriculture, Division of Meteorology, University of Novi Sad, Serbia*
[2] *Faculty of Sciences, Department of Geography, University of Sarajevo, Bosnia and Herzegovina*
[3] *Faculty of Sciences, Department of Physics, University of Novi Sad, Serbia*

**Correspondence to:*
D. T. Mihailović, Faculty of Agriculture, Division of Meteorology, University of Novi Sad, Novi Sad, Serbia.
E-mail: guto@polj.uns.ac.rs

Received: 28 May 2014
Revised: 12 November 2014
Accepted: 8 January 2015

Abstract

We have used the Kolmogorov complexity (KC) and three suggested complexity measures to describe the complexity of spatial distribution of precipitation in Bosnia and Herzegovina, for the period 1960–1984. In particular, we have examined the monthly precipitation amount time series from 23 stations and then calculated the KC using the Lempel–Ziv Algorithm (LZA), Kolmogorov complexity spectrum (KCS), Kolmogorov complexity spectrum highest value (KCM) and overall Kolmogorov complexity (OKC) values for each time series. Our results indicate that the difference in complexity of spatial distribution of precipitation may be attributed to influence of Adriatic Sea, relief and Pannonian Basin.

Keywords: **Bosnia and Herzegovina; precipitation data analysis; complexity; Kolmogorov complexity spectrum; overall Kolmogorov complexity**

1. Introduction

The term complexity is closely associated with the term precipitation, regardless of the context in which it is used, i.e. (1) analysis of precipitation events which can be surprisingly on a broad temporal and spatial scales (Millan *et al.*, 2011), (2) predictability of hourly and daily precipitation (Elsener and Tsonis, 1993; Silva *et al.*, 2006), (3) complexity analysis of precipitation in changing environment (Luan *et al.*, 2011), (4) statistical complexity of daily precipitation in reanalysis (Tatli, 2014), (5) complexity of a time series for multiple grid points to analyze, for example, precipitation under subtropical high conditions (Feng *et al.*, 2009), etc. However, the complexity of spatial distribution of precipitation in mountainous and other areas, as a result of influence of geographical factors such as sea, huge depressions and rivers, is not very well known and studied. This is partly because of the orographic complexity in these areas and partly because of the sparse distribution of pluviometric measurements (Prudhomme and Reed, 1999). In order to understand the complex dynamics of the spatial distribution of the precipitation, it is important to examine the complexity of the influence of the geographical factors on that distribution. The degree of complexity may also influence the predictability of precipitation in the region considered.

The purpose of this paper is to consider the complexity of the spatial distribution of precipitation in Bosnia and Herzegovina for the period 1960–1984, using the Kolmogorov complexity (KC) which is a measure of disorder in a time series. For complexity analysis of the monthly precipitation amount we have used (1) the Kolmogorov complexity spectrum (KCS), (3) Kolmogorov complexity spectrum highest value (KCM) and (3) the overall Kolmogorov complexity (OKC), which are introduced in this paper. To our knowledge, this is the first time that those measures are used in precipitation data analysis. Finally, we have defined the Precipitation Complexity Index (PCI) which clearly distinguishes degree of influence of geographical factors on spatial distribution of precipitation.

2. Study area and datasets

Bosnia and Herzegovina is located in the western Balkans, surrounded by Croatia to the north and south-west, Serbia to the east, and Montenegro to the south-east. It lies between latitudes 42° and 46°N, and longitudes 15° and 20°E. The country is mostly mountainous, encompassing the central Dinaric Alps. The northeastern parts reach into the Panonian basin, while in the south it borders the Adriatic Sea. Dinaric Alps generally run in east–west direction, and get higher toward the south. The highest point of the country is peak Maglić at 2 386 m, at the Montenegrin border, while the major mountains include Kozara, Grmeč, Vlašić, Čvrsnica, Prenj, Romanija, Jahorina, Bjelašnica and Treskavica (Figure 1a).

Datasets of monthly precipitation amounts for the period 1960–1984 were taken from the Annual Report of the Hydrometeorological Institute of Bosnia and Herzegovina. On that basis, we have obtained the map of spatial distribution of annual precipitation (Figure 1b) following Drešković and Đug (2012), who used the Ordinary Kriging interpolation procedure which is suitable for application, especially, in

© 2015 Royal Meteorological Society

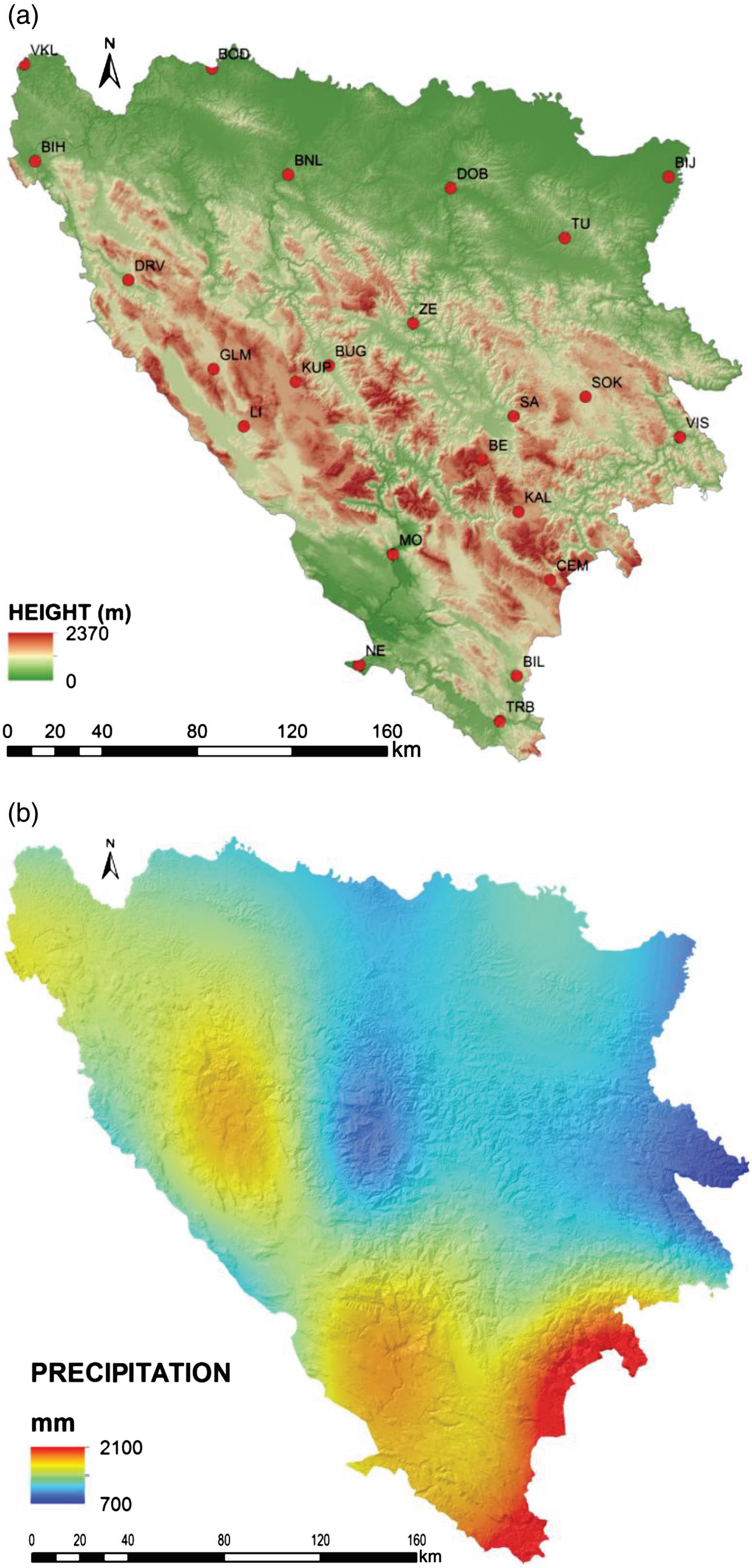

Figure 1. (a) Relief of Bosnia and Herzegovina with location of 23 meteorological stations used in the study (their abbreviations are given in Table 1) and (b) spatial distribution of annual precipitation amount obtained by the Ordinary Kriging interpolation procedure from the observations (Drešković and Đug, 2012).

© 2015 Royal Meteorological Society

mountain regions (Antonić et al., 2000; Ninyerola et al., 2000; Whiteman, 2000).

3. Measures based on the KC

KC is a measure which points out to the minimum length of a program such that a universal computer can generate a specific sequence. A good introduction to this complexity can be found in Li and Vitanyi (1997). On the basis of Kolmogorov's idea, Lempel and Ziv (1976) developed an algorithm (Lempel–Ziv Algorithm, LZA), which is used as a measure of disorder of finite sequences. For complexity analysis of the monthly precipitation amount in Bosnia and Herzegovina, besides the KC measure, we have used (1) the KCS, (2) KCM and (3) the OKC – measures which are introduced in this paper.

3.1. The KC

The KC analysis of a time series $\{x_i\}$, $i = 1, 2, 3, 4, \dots, N$ can be carried out as follows. First, encoding the time series by constructing a sequence S of the characters 0 and 1 written as $\{s(i)\}$, $i = 1,2,3,4, \dots, N$, according to the rule

$$s(i) = \begin{cases} 0 & x_i < x_* \\ 1 & x_i \geq x_* \end{cases} \tag{1}$$

Here x_* is a chosen threshold. We use the mean value of the time series to be the threshold. The mean value of the time series has often been used as the threshold (Zhang et al., 2001). Also, other encoding schemes can be used in dependance on the application (Radhakrishnan et al., 2000; Small, 2005; Mihailović et al., 2014). Second step is to calculate the complexity counter $c(N)$. The $c(N)$ is defined as the minimum number of distinct patterns contained in a given character sequence (Ferenets et al., 2006). The complexity counter $c(N)$ is a function of the length of the sequence N. The value of $c(N)$ is approaching an ultimate value $b(N)$ as N approaching infinite, i.e.

$$c(N) = O(b(N)), \quad b(N) = \frac{N}{\log_2 N} \tag{2}$$

Finally, we calculate the normalized complexity measure $C_k(N)$, which is defined as

$$C_k(N) = \frac{c(N)}{b(N)} = c(N)\frac{\log_2 N}{N} \tag{3}$$

The $C_k(N)$ is a parameter to represent the information quantity contained in a time series, and it is to be a 0 for a periodic or regular time series and to be a 1 for a random time series, if N is large enough. For a non-linear time series, $C_k(N)$ is to be between 0 and 1.

3.2. The KCS

The time series obtained either by a measuring procedure or as an output from an atmospheric model, for complexity analysis, has to be transformed on the following way. First, we create a normalized time series $\{x_i\}$, $i = 1, 2, 3, 4, \dots, N$ by the transformation $x_i = (X_i - X_{\min})/(X_{\max} - X_{\min})$, where $\{X_i\}$ is a time series obtained either by a measuring procedure (in our case that is precipitation time series) or from an atmospheric model, where $X_{\max} = \max\{X_i\}$ and $X_{\min} = \min\{X_i\}$. Then, we make a transformation into a finite symbol string by comparison with series of thresholds $\{x_{t,i}\}$, $i = 1, 2, 3, 4, \dots, N$, where each element is equal to the corresponding element in the considered time series $\{x_i\}$, $i = 1, 2, 3, 4, \dots, N$. The original signal samples are converted into a 0–1 sequences $\left\{S_i^{(k)}\right\}$, $i = 1, 2, 3, 4, \dots, N$, $k = 1, 2, 3, 4, \dots, N$ defined by comparison with a threshold $x_{t,k}$,

$$S_i^{(k)} = \begin{cases} 0 & x_i < x_{t,k} \\ 1 & x_i \geq x_{t,k} \end{cases} \tag{4}$$

After we apply the LZA on each element of series $\left\{S_i^{(k)}\right\}$, we get the KCS $\left\{K_i^C\right\}$, $i = 1, 2, 3, 4, \dots, N$. We introduce this sequence to explore the range of amplitudes in a time series representing a process, for which that process has the highest complexity. This sequence we call the KCS. The highest value K_m^C in this spectrum, i.e. $K_m^C = \max\left\{K_i^C\right\}$, we call the KCM which can take value up to 1.2 (Hu et al., 2006).

3.3. The overall KC

Sometime, the KC complexity is not able to discern between time series with different Kolmogorov spectra of complexity. From this reason, in the precipitation time series analysis, we use an overall KC measure K_O^C (OKC) defined as

$$K_O^C = \frac{1}{x_m}\int_X K_s^C \, dx \tag{5}$$

where K_s^C is the spectrum of the KC, x_m is a highest value of the physical quantity in a time series, while dx and X are differential and domain of that quantity, over which this integral takes values, respectively (Mihailović et al., 2013). Because K_s^C is given as the sequence $\left\{K_i^C\right\}$, $i = 1, 2, 3, 4, \dots, N$, it is calculated numerically as

$$K_O^C = \frac{1}{x_m}\left[K_1^C\left(x_2 - x_1\right) + \frac{1}{2}\sum_{i=1}^{N-1} K_i^C\left(x_{i+1} - x_{i-1}\right) + K_N^C\left(x_N - x_{N-1}\right)\right] \tag{6}$$

The K_O^C takes value on the interval $(0, K_u)$, where K_u can also takes value larger than 1. This measure can make distinction between different time series having close values of the KC and KCM.

© 2015 Royal Meteorological Society

Table 1. Kolmogorov complexity (KC), Kolmogorov complexity spectrum highest value (KCM) and overall Kolmogorov complexity (OKC) values for the monthly precipitation amount time series of 23 places in Bosnia and Herzegovina for the period 1960–1984.

Place	Abb.	Altitude (m)	Latitude (N)	Longitude (E)	KC	KCM	OKC
Sarajevo	SA	637	43°52′	18°26′	1.070	1.152	0.557
Bihać	BIH	246	44°49′	15°53′	1.152	1.152	0.511
Bosanska Dubica	BOD	100	45°11′	16°49′	1.152	1.152	0.511
Banja Luka	BNL	153	44°47′	17°13′	1.042	1.179	0.438
Doboj	DOB	146	44°44′	18°06′	1.125	1.152	0.424
Tuzla	TU	305	44°33′	18°42′	1.097	1.152	0.440
Bijeljina	BIJ	90	44°46′	19°16′	1.070	1.125	0.499
Drvar	DRV	485	44°23′	16°24′	1.070	1.152	0.553
Bugojno	BUG	562	44°04′	17°28′	1.015	1.152	0.480
Zenica	ZE	344	44°13′	17°54′	1.042	1.070	0.431
Kupres	KUP	1190	44°00′	17°17′	1.070	1.179	0.553
Livno	LI	724	43°50′	17°01′	1.097	1.125	0.545
Bjelašnica	BE	2067	43°43′	18°16′	1.015	1.042	0.475
Kalinovik	KAL	1073	43°31′	18°27′	1.125	1.152	0.441
Sokolac	SOK	872	43°57′	18°49′	1.015	1.097	0.443
Mostar	MO	99	43°21′	17°48′	1.097	1.097	0.558
Bileća	BIL	491	42°53′	18°27′	1.015	1.097	0.526
Čemerno	CEM	1305	43°14′	18°36′	1.015	1.152	0.513
Glamoč	GLM	1031	44°03′	16°52′	1.073	1.128	0.461
Velika Kladuša	VKL	157	45°12′	15°49′	1.125	1.152	0.450
Neum	NE	130	42°56′	17°36′	1.015	1.070	0.560
Višegrad	VIS	364	43°47′	19°18′	1.125	1.125	0.434
Trebinje	TRB	276	42°43′	18°21′	0.960	1.070	0.526

4. Results and comments

Using the calculation procedure outlined in Section 3.1, we have computed the KC values, for all 23 precipitation time series, on the basis of the LZC algorithm. The calculations are carried out for the entire time interval 1960–1984. From Table 1, it is seen that calculated KC values are ranged between 0.960 (TRE) and 1.152 (BOD, BIH). If the KC value of a time series is close to 1, then it is associated with a stochastic process (Ferreira *et al.*, 2003). Otherwise, if it is near 0, then it is associated with a simple deterministic process like a periodic motion. From the same table, it can be seen that all considered time series have the KC values close to 1, i.e. what indicates about their high randomness since that the KC measure can be also considered as a measure of randomness. The spatial distribution of the KC of the monthly precipitation amount time series in Bosnia and Herzegovina for the period 1960–1984 is visualized, using values from Table 1, as a map (Figure 2). This figure shows that the highest values of the KC are (1) in the lowland and hilly regions (north-west, north-east and central east areas) and partly (2) in a belt extending from the high mountain region (south-east) toward the hilly and lowland areas of the country (south-west) as it is seen from Figure 1. The lowest values of the KC capture the central part of the country; small ones are on the south and area next to the central east. From such distribution of the KC, only conclusions we can underline are (1) all monthly precipitation amount time series evidently include high degree of randomness and (2) the presence of stochastic component in them is enhanced in the areas next to the Pannonian Basin and central east part, next to the Drina River, extending as a belt toward south-west (under the influence of the Adriatic Sea). After this analysis, no more specific conclusions can be outlined about influence of Adriatic Sea, relief and Pannonian Basin on the spatial distribution of the precipitation in Bosnia and Herzegovina. The KC measure is used in analysis of environmental time series, such as river flow and non-ionized radiation, that cannot be done by traditional mathematical statistics (Mihailović *et al.*, 2013; Mihailović *et al.*, 2014).

An understanding of complexity in monthly precipitation amount time series may provide insight into its complex dynamics. However, the KC measure is not able to recognize as distinct between time series with different amplitude variations and similar random components. From that reason, we use the KCS that can be considered as a novel method in quantifying amplitude and complexity variations in time series (Mihailović *et al.*, 2013).

We consider the KCS spectra of the monthly precipitation amount time series for Tuzla (TU) and Trebinje (TRB). The first place is a typically continental one, while TRB is close to the Adriatic Sea (Figure 1a). They have different KC values (0.960 and 1.097). Because that the KC measure for TU (1.097) is greater than for TRB (0.960), it seems that the TU monthly precipitation amount time series is more complex than TRB time series. We have obtained the KCS for both places following procedure given in Section 3.2. Results of calculations, i.e. corresponding spectra, are depicted in Figure 3. From this figure, it can be observed that the KCM for TU (1.152) and TRB (1.070) are close to each other (Figure 3 and Table 1). Because the KCM is a finer measure for complexity of a time series than commonly used the KC one (Mihailović *et al.*, 2013), we

© 2015 Royal Meteorological Society

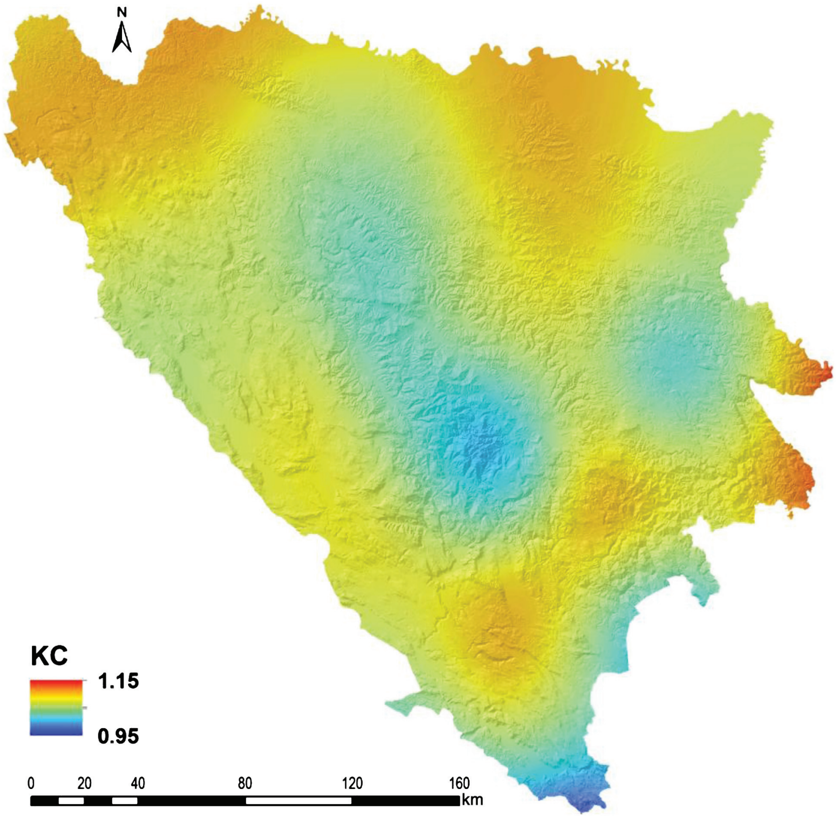

Figure 2. Spatial distribution of the Kolomogorov complexity (KC) of the monthly precipitation amount time series in Bosnia and Herzegovina for the period 1960–1984.

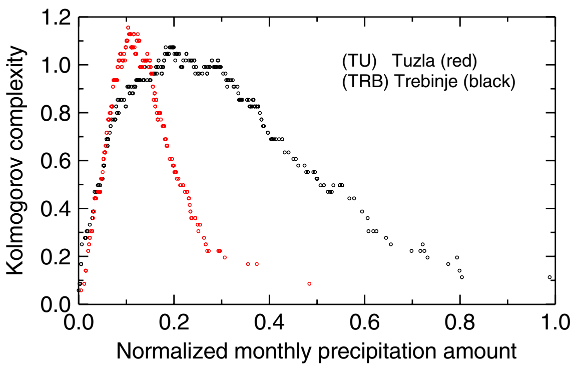

Figure 3. The Kolomogorov complexity spectrum (KCS) of the monthly precipitation amount time series in Tuzla (TU) and Trebinje (TRB) for the period 1960–1984. All monthly precipitation amounts are normalized on the highest value in the period indicated (Trebinje, 591 mm).

can note that these time series are more similar in their behavior. Thus, we still cannot distinguish complexities, which we analyze. However, from Figure 3, it is seen that KCS spectra of TU and TRB time series differ significantly. First, the KCS of TRB time series captures a broader range of amplitudes than TU one. Second, the maximum of the TU spectra is shifted toward lower amplitudes of the monthly precipitation amount. These two features point out that the TRB time series is more complex than TU one. The shape of KCS curve depends on variability of time series amplitudes what cannot be captured by the KC and KCM measures.

We have calculated the OKC, which is an integral measure of the complexity over the whole spectrum including all amplitudes, i.e. TRB (0.526) and TU (0.440) what can be seen in Table 1. Thus, while the differences in the KCM, for those places, are slightly different, on the other hand the differences in the OKC are larger (about 19.5%). Therefore, we have to include this additional information about complexity that is not contained in KC and KCM. In conclusion, we can say that the TRB time series is more complex than TU time series. The complexity measure OKC in the precipitation data analysis we will call the PCI. Finally, we have computed the PCI index for all 23 places in Bosnia and Herzegovina. Figure 4 depicts a spatial distribution of the PCI index of these places for the period 1960–1984. This map enhances two regions: (1) with higher PCI index which is strongly influenced by the presence of the Adriatic Sea (blue colored) and (2) with lower PCI index, which corresponds with the vicinity of the Pannonian Basin (green colored). Between them is a transition belt (light blue colored) with the PCI values indicating on the mixed influences of the above mentioned geographical factors and relief. This index quantitatively indicates that the difference in complexity of spatial distribution of precipitation in Bosnia and Herzegovina is caused by the influence of (1) Adriatic Sea (close to the Mediterranean Sea), (2) relief and (3) the Pannonian Basin (Figure 1) on its climatic regime and accordingly to spatial and temporal distribution of precipitation. Although situated close to the

© 2015 Royal Meteorological Society

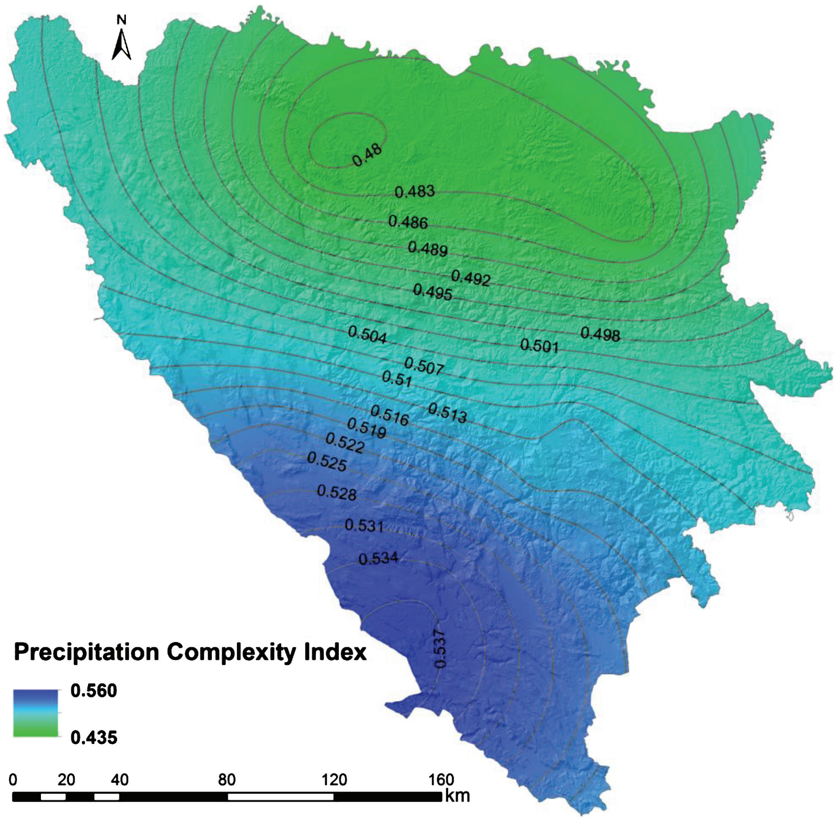

Figure 4. Spatial distribution of the overall Kolomogorov complexity (OKC) of the monthly precipitation amount time series in Bosnia and Herzegovina for the period 1960–1984.

Mediterranean Sea, Bosnia and Herzegovina is largely cut off from its climatic influence by the Dinaric Alps. Thus, amount of precipitation is highest in the south western part extending to the northwest along the Adriatic coast (Figure 1). From Figure 3, it is seen that the PCI index has the hugest value in this area. It means that randomness is present in the precipitation time series, in this region, resulting in a higher complexity. Next to the north and northeast amount of precipitation decreases, thus in their time series randomness is less present (lower PCI index).

5. Conclusions

In this paper, we have studied the complexity of spatial distribution of precipitation in Bosnia and Herzegovina using monthly precipitation time series from 23 meteorological stations. We have calculated the KC using the LZA. We have used complexity measures, i.e. KCS and OKC values in analysis of the monthly precipitation time series. We have defined the complexity precipitation index (PCI). This index quantitatively indicates that the difference in complexity of spatial distribution of precipitation may be attributed to influence of Adriatic Sea, relief and Pannonian Basin.

Acknowledgements

This paper was realized as a part of the project 'Studying climate change and its influence on the environment: impacts, adaptation and mitigation' (III 43007) financed by the Ministry of Education and Science of the Republic of Serbia within the framework of integrated and interdisciplinary research for the period 2011–2014.

References

Antonić O, Križan J, Marki A, Bukovec D. 2000. Spatio-temporal interpolation of climatic variables over large region of complex terrain using neural networks. *Ecological Modelling* **138**: 255–263, doi: 10.1016/S0304-3800(00)00406-3.

Drešković N, Đug S. 2012. Applying the inverse distance weighting and Kriging methods of the spatial interpolation on the mapping the annual precipitation in Bosnia and Herzegovina. *International Environmental Modelling and Software Society (iEMSs), International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet*. Sixth Biennial Meeting, Leipzig, Germany. R Seppelt, AA Voinov, S Lange, D Bankamp (eds). http://www.iemss.org/society/index.php/iemss-2012-proceedings (accessed 25 April 2014).

Elsener JB, Tsonis AA. 1993. Complexity and predictability of hourly precipitation. *Journal of Atmospheric Sciences* **50**: 400–405.

Feng S, Liu Q, Fu Y-F. 2009. Precipitation under subtropical high conditions: evidence and implications. *Atmospheric and Oceanic Science Letters* **2**: 244–249.

Ferenets R, Lipping T, Anier A, Jantti V, Melto S, Hovilehto S. 2006. Comparison of entropy and complexity measures for the assessment of depth of sedation. *IEEE Transactions on Biomedical Engineering* **53**: 1067–1077.

Ferreira FF, Francisco G, Machado BS, Murugnandam P. 2003. Time series analysis for minority game simulations of financial markets. *Physica A* **321**: 619–632.

Hu J, Gao J, Principe JC. 2006. Analysis of biomedical signals by the Lempel-Ziv complexity: the effect of finite data size. *IEEE Transactions on Biomedical Engineering* **53**: 2606–2609.

© 2015 Royal Meteorological Society

Lempel A, Ziv J. 1976. On the complexity of finite sequences. *IEEE Transactions on Information Theory* **22**: 75–81.

Li M, Vitanyi P. 1997. *An Introduction to Kolmogorov Complexity and its Applications*, 2nd edn. Springer-Verlag: Berlin.

Luan Q-H, Wang H, Xia D-Z. 2011. Complexity analysis of precipitation in changing environment in Chien River Basin, China. *Water Science and Engineering* **4**(2): 133–142.

Mihailović DT, Malinović-Miličević S, Arsenić I, Drešković N, Bukosa B. 2013. Kolmogorov complexity spectrum for use in analysis of UV-B radiation time series. *Modern Physics Letters B* **27**: 1350194.

Mihailović DT, Nikolić-Đorić E, Drešković N, Mimić G. 2014. Complexity analysis of the turbulent environmental fluid flow time series. *Physica A* **395**: 96–104.

Millan H, Rodriguez J, Ghanbarian-Alavijeh B, Biondi R, Llerena G. 2011. Temporal complexity of daily precipitation records from different atmospheric environments: Chaotic and Lévy stable parameters. *Atmospheric Research* **101**: 879–892.

Ninyerola M, Pons X, Roure MJ. 2000. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *International Journal of Climatology* **20**: 1823–1841.

Prudhomme C, Reed DW. 1999. Mapping extreme rainfall in a mountainous region using geostatistical techniques: a case study in Scotland. *International Journal of Climatology* **19**: 1337–1356.

Radhakrishnan N, Wilson JD, Loizou PC. 2000. An alternative partitioning technique to quantify the regularity of complex time series. *International Journal of Bifurcation and Chaos* **10**: 1773–1779.

Silva MES, Carvalho LMV, da Silva Dias MAF, Xavier T d MBS. 2006. Complexity and predictability of daily precipitation in a semi-arid region: an application to Cear´a, Brazil. *Nonlinear Processes in Geophysics* **13**: 651–659.

Small M. 2005. *Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance*. World Scientific: Singapore.

Tatli H. 2014. Statistical complexity in daily precipitation of NCEP/NCAR reanalysis over the Mediterranean Basin. *International Journal of Climatology* **34**: 155–161.

Whiteman CD. 2000. *Mountain Meteorology Fundamentals and Aplications*. Oxford University Press: Oxford.

Zhang XS, Roy RJ, Jensen EW. 2001. EEG complexity as a measure of depth of anesthesia for patients. *IEEE Transactions on Biomedical Engineering* **48**: 1424–1433.

© 2015 Royal Meteorological Society