Investigation of precipitation trends in eastern Slovakia using singular spectrum analysis

Zeleňáková M.[1,*], Jothıprakash V.[2], Purcz P.[3], Unnıkrıshnan P.[2] And Hlavatá H.[4]

[1]Technical University of Košice, Institute of Environmental Engineering, Vysokoškolská 4, 042 00 Košice, Slovakia

[2]Indian Institute of Technology, Department of Civil Engineering, Bombay, Mumbai 400076, India

[3]Technical University of Košice, Institute of Construction Technology and Management, Vysokoškolská 4, 042 00 Košice, Slovakia,

[4]Slovak Hydrometeorological Institute, Ďumbierska 26, 041 17 Košice, Slovakia

*corresponding author:

e-mail: martina.zelenakova@tuke.sk

Abstract

Precipitation plays an important role in the water management of any country. Knowledge of the temporal and spatial distribution of precipitation is inevitable for proper integrated water resources management. In the present study, trends of precipitation time series from Košice climatic station situated in eastern Slovakia have been investigating. Analyzed data for period from 1951 to 2013 were provided by Slovak Hydrometeorological Institute. Trend analysis were done by Mann–Kendall (MK) test and the Singular Spectrum Analysis (SSA) method. Method of periodogram analysis has been used in order to select the components corresponding to trend in the grouping stage of SSA. SSA has been applied to daily precipitation time series in Kosice station. MK test is also used to detect trends in monthly series and the performance of SSA and MK test was compared. The result showed that the MK test could detect the presence of a positive or negative trend at a significant level, whereas the proposed SSA method could extract the nonlinear trend present in the series along with its shape. MK test does not prove any significant trends in Kosice climatic station during the evaluated period. The study proves the applicability of SSA for extracting nonlinear trends.

Keywords: precipitation, trend analysis, singular spectral analysis, Mann-Kendall test

1. Introduction

The main idea of trend analysis is to detect whether values of data are increasing, decreasing or trendless over time (Kisi, Ay, 2014). Detection of trend is a complex subject because of characteristics of data. Descriptive statistics of meteorological, hydrological and climatological data show variability in time. This variability may be cyclical with the seasons, steadily (a trend), sudden jumps or some other established variations. These characteristics are a possible existence of seasonality, skewness, serial correlation, non-normal data, ‘‘less-than’’ (censored) values, outliers and missing values (Hirsch and Slack, 1984; Hirsch *et al.*, 1991; Helsel and Hirsch, 1992; Mozejko, 2012; Kisi, Ay, 2014); therefore, some assumptions have been developed to overcome these possibilities while applying trend methods. A positive autocorrelation is one of these possibilities (pre-whitening), and causes more false detection of a significant trend than specified by the significance level (von Storch, 1995; Zhang and Zwiers, 2004). Some authors proposed using the pre-whitening method for hydro-climatic series. The pre-whitening method is a procedure for reduction of serial correlation within a given time series by adding white noise (serially independent) series to the original series. For instance, von Storch (1995) indicated that the existence of positive serial correlation increases the probability that the MK test detects trend when the given time series is trend free. He then proposed removal of the serial correlation through the pre-whitening procedure before the application of the MK test. But, Douglas *et al*. (2000) and Yue *et al*. (2002) explored the influence of the pre-whitening procedure and they found that removal of positive serial correlation by pre-whitening removes a portion of trend. The other possibility is the homogeneity tests such as Van Belle and Hughes (1984). This test is used to detect whether or not seasonal trends have the homogeneity feature (Yu *et al*., 1993).

Singular spectrum analysis (SSA) is a nonparametric, model-free technique for time series analysis, which decomposes the time series into small interpretable components such as trend, periodic component, cyclic component, and noise (Golyandina and Zhigljavsky, 2013; Hassani, 2007). Non-parametric models have also been developed and applied, largely following the work of Yakowitz (1973). The use of non-parametric techniques tends to be a more robust option when testing data that differ from “normality”. Furthermore, the use of nonparametric techniques is known to be more resilient to outliers (Hirsh *et al*, 1982).

In the present study, the MK test and the method of SSA have been applied for Košice precipitation time series. Trend analyzes has been carried out for individual months by MK test and for daily rainfall series the trend have been extracted using the integrated SSA periodogram method. There were found no significant trends in monthly precipitation series.

The similar study was done by Unnikrishnan and Jothiprakash (2015). They applied SSA to extract the trends of two rainfall series from different locations having different time periods and time steps. Unnikrishnan and Jothiprakash (2015) analyzed monthly rainfall data of England and Wales precipitation from 1766 to 2002 in which no periodic component is prevailing and daily rainfall data of Koyna watershed, Maharashtra, India from 1961 to 2009, which shows a strong periodic component of 365 days.

The methods used in this study are particularly described in the following topics.

2. Material and methods

2.1. Data

In this study daily precipitation time series observed in Kosice climatic station for the period of 63 years (from 1st January 1951 to 31st December 2013) have been analysed. Figure 1 presents monthly average precipitation during the evaluated period.

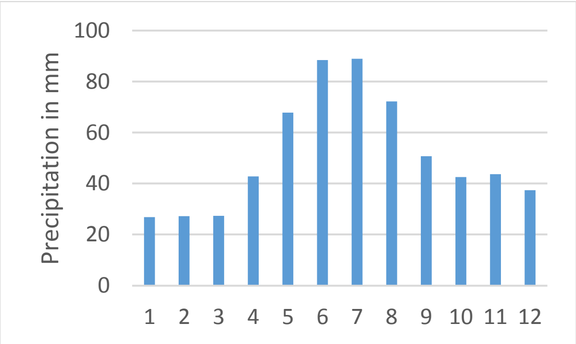

Figure 1. Monthly average precipitation in Košice station (1951–2013).

These data were provided by Slovak Hydrometeorological Institute.

2.2. Mann-Kendall test

For detecting trends in hydro-meteorological variables, various statistical methods have been developed and used over the years (Jha and Singh, 2013; Martinez *et al*, 2012; Modarres and Silva 2007; Sonali and Nagesh, 2013; Tabari and Talaee, 2011; Önöz, and Bayazit, 2003).

Parametric (distribution-dependent) and non-parametric (distribution-free) analysis methods are frequently used by many researchers for trend analysis. These methods have also been applied to different kinds of data such as annual, monthly, and seasonal time series in several applications. The Mann–Kendall (MK) test is suitable for cases where the trend may be assumed to be a monotonic and normal distribution and thus no seasonal or the other cycle is present in the data. An analysis technique for observed time series may be particularly significant if it is able to expose important characteristics of the time series predictability (Marques *et al*., 2006).

The non-parametric MK trend test (Kendall, 1975; Mann, 1945) is the most common of the various statistical procedures used to analyse time series datasets. The Mann-Kendall test follows statistics based on standard normal distribution (Z), using Eq. (1) (Kendall, 1975; Mann, 1945):

$$Z = \begin{cases} \frac{S-1}{\sqrt{Var(S)}} & if \quad S > 0 \\ 0 & if \quad S = 0 \\ \frac{S+1}{\sqrt{Var(S)}} & if \quad S < 0 \end{cases} \qquad (1)$$

in which,

$$S = \sum_{k=1}^{n-1} \sum_{k+i}^{n} \; sgn(x_j - x_k) \qquad (2)$$

$$sgn(x_j - x_k) = \begin{cases} +1 & if \left(x_j - x_k\right) > 0 \\ 0 & if \left(x_j - x_k\right) = 0 \\ -1 & if \left(x_j - x_k\right) < 0 \end{cases} \qquad (3)$$

$$Var(S) = [n(n-1)(2n+5) - \sum_{i=1}^{m} t_i(t_i - 1)(2t_i + 5)]/18 \qquad (4)$$

where n is the number of data points and m is the number of tied groups (a set of sample data having the same value).

According to this test, the null hypothesis H0 states that the depersonalized data (x_1, …, x_n) consist of a sample of n independent and identically distributed random variables. The alternative hypothesis H1 of a two-sided test is that the distributions of x_k and x_j are not identical for all k, $j \leq$ n with $k \neq j$.

The null hypothesis H0 (no trend) is accepted if $Z<Z_{\alpha/2}$ and rejected (hypothesis H1) if $Z>Z_{\alpha/2}$ where α is the significance level and $Z_{\alpha/2}$ the standard normal distribution for α/2. The applications carried out considered α=0.05 and, accordingly, $Z_{\alpha/2}$=1.645. Positive values of Z indicate increasing trends, while negative values of Z show decreasing trends.

In this study, a programme developed in Microsoft Excel was used to generate the algorithm of the non-parametric MK test.

2.3. Singular spectrum analysis

Singular Spectrum Analysis (SSA) is a non parametric method of time series analysis which utilizes singular value decomposition (SVD) for decomposing the time series into sum of eigen triples which include eigen function, principal component and square root of eigen values. The method was developed by Broomhead and King (1986) and is explained in detail in Golyandina and Zhigljavsky (2013). The two main stages in SSA are decomposition and reconstruction. The time series upon consideration is converted into a trajectory matrix based on the chosen window length and singular value decomposition is carried out. The end product of decomposition is eigen triples which include eigen functions, principal components and square root of eigen values. Selection of appropriate components from the decomposed ones is called grouping. Reconstruction of the time series is achieved by averaging the corresponding diagonals of the selected component matrices. The methodology of SSA can be simply explained as follows.

If $X = x_1, x_2, x_3, \ldots, x_N$ is the time series, the trajectory matrix Y can be expressed as:

$$Y = \begin{bmatrix} x_1 & x_2 & \cdots & x_L \\ x_2 & x_3 & \cdots & x_{L+1} \\ \cdots & \cdots & \cdots & \cdots \\ x_K & x_{K+1} & \cdots & x_N \end{bmatrix} \quad (5)$$

where L is window length, K is lag parameter and can be calculated using the relationship $K=N-L+1$ and N is the time series length.

$$Y = \sum_{i=1}^{d} Y_i \quad (6)$$

where

$$Y_i = U_i \sqrt{\chi_i} V_i \quad (7)$$

where χ is eigen value of U; V are eigen functions and principal components of the matrix YY^T, d is number of positive eigen values and i varies from 1 to d.

SSA decomposes the vector space of the trajectory matrix of the time series into a signal subspace and a noise subspace by singular value decomposition. Selection of appropriate components for grouping is very important for the accuracy of results as the time series is reconstructed from those selected components only. If proper grouping is not done, the signal subspace can still contain noise. The existing methods of grouping for trend extraction using SSA are either selecting some of the leading eigen functions or choosing slowly varying eigen functions, but these methods may not work well if the trend component is comparatively small. In the present study, in order to select corresponding trend components, the method based on periodogram analysis proposed by Alexandrov (2009) is integrated with basic SSA.

The Fourier representation of a time series $X = x_1, x_2, x_3, \ldots, x_N$ can be written as

$$x_n = c_0 + \sum (c_k \cos(2\pi nk/N) + \sin(2\pi nk/N)) + (-1)^n c_{N/2} \quad (8)$$

where $k \in N$, $0 \le n \le N\text{-}1$; and $c_{N/2}=0$ if N is an odd number. The perdiogram of X at the frequencies $\omega=k/N$ can be defined as

$$I_X^N(K/N) = N/2 \begin{Bmatrix} 2c_0^2, k=0 \\ c_k^2 + s_k^2, 0 \le k \le N/2 \\ 2c_{N/2}^2, k = N/2 \end{Bmatrix} \quad (9)$$

The eigen vectors that satisfy conditions can be selected for the extraction of trend.

In this study we used the software implementing SSA http://gistatgroup.com: general-purpose interactive 'Caterpillar'-SSA software (Windows) following the methodology described in Golyandina *et al.* (2001), Golyandina and Zhigljavsky (2013).

3. Results

3.1. Mann-Kendall test

The MK test has been applied to monthly precipitation series in Košice climatic station and the test statistic is tested for a confidence level of 95%. The results of the MK statistic – *Z*, for each month of year from January (1) to December (12) of the 1951–2013 are shown in Figure 2.

None from the investigated months presents significant trends in precipitation. The obvious is almost significant negative trend in precipitation time series in Košice in November and March. The result of the MK statistic – Z = 0.001416, for daily series of the 1951–2013, which presents also not significant trend at confidence level of 95%.

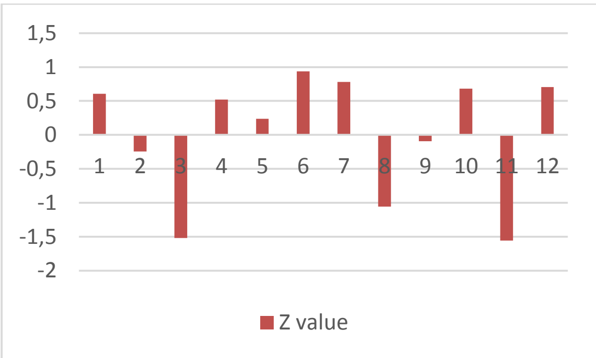

Figure 2. MK test statistic, *Z*, of various months for monthly Košice series (1951–2013).

3.2. Singular spectrum analysis

Daily time-step precipitation data observed for a period of 63 years over the Košice catchment area have been analysed using SSA. In this case, the trend has been analysed continuously from 1st January 1951 to 31st December 2013. There is a periodic component of 365 days in the data. The window length is selected as same as the period, 365 days. After singular value decomposition, eigenfunctions, principal components, and eigenvalues are generated. The first eigenfunction of 12.42%, is the leading eigenfunction. Thus, the first eigenfunction was used for extracting the trend. The extracted trend along with the observed precipitation and trend alone are shown in Figure 3.

The trend shows low values, implying that there is no relevant increase or decrease of average rainfall over a long period of time and the series is almost stationary.

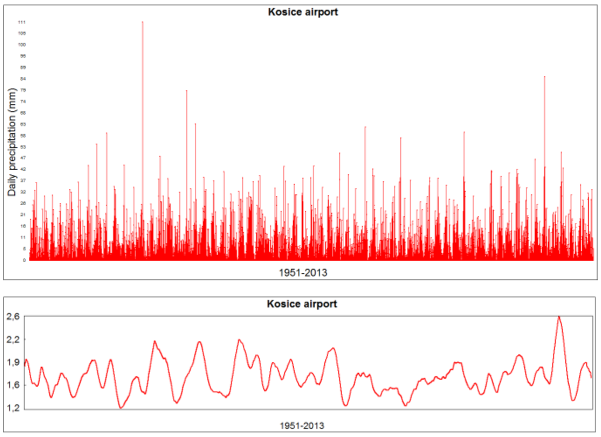

Figure 3. Extracted trend for Košice daily precipitation series.

4. Conclusion

Extraction of precipitation trends in Košice climatic station, Slovakia, for a time period of 1951–2013 by using SSA is the focus of the present study. SSA has been applied to daily and nonlinear trends were extracted. The Mann–Kendall test was also used to detect trends of Košice climatic station monthly precipitation series and compared with the the method of SSA. The comparison of the SSA method with the MK test showed that the MK test could detect the presence of a positive or negative trend for a particular confidence level over time (yet not significant). However the SSA method could extract the nonlinear trend with its shape so that small-scale variations in trend also can be studied. Results showed very small variation, implying very little change in the pattern of rainfall. The results of the present study for both the data sets provided ample opportunity on applicability of SSA and periodogram analysis in hydrologic time series (for different timescales) for extracting nonlinear trend in a successful manner.

Acknowledgement

This work has been supported by the Slovak Research and Development Agency bilateral project between Slovakia and Portugal SK-PT-2015-0007.

References

Alexandrov, T. (2009), A method of trend extraction using Singular Spectrum Analysis, *Revstat - Statistical Journal*, **7**(1), 1–22.

Broomhead, D. and King, G. P. (1986). Extracting Qualitative Dynamics from Experimental Data, *Physica D: Nonlinear Phenomena*, **20**, 217–236.

Douglas, E.M., Vogel, R.M. and Kroll, C.N. (2000), Trends in floods and low flows in the United States: impact of spatial correlation. *Journal of Hydrology* **240** (1–2), 90–105.

Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman&Hall/CRC.

Golyandina, N. and Zhigljavsky, A. (2013), *Singular Spectrum Analysis for Time Series*. Springer Briefs in Statistics. Springer.

Hassani, H. (2007), Singular Spectrum Analysis: Methodology and Comparison, *Journal of Data Science*, **5**, 239-257.

Hirsch, R.M., Alexander, R.B. and Smith, R.A. (1991), Selection of methods for the detection and estimation of trends in water quality, *Water Resources Research*, **27** (5), 803–813.

Hirsch, R.M. and Slack, J.R. (1984), A nonparametric trend test for seasonal data with serial dependence, , **20** (1), 727–732.

Hirsch R.M., Slack J.R. and Smith R.A. (1982), Techniques of trend analysis for monthly water quality data, *Water Resources Research*, **18**, 107–121.

Helsel, D.R. and Hirsch, R.M. (1992), *Statistical Methods in Water Resources*. Elsevier, p. 522.

Jha, M.K. and Singh, A.K. (2013), Trend analysis of extreme runoff events in major river basins of Peninsular Malaysia, *Int. J. Water*, 7, 142–158.

Kisi, O. and Ay, M. (2014), Comparison of Mann–Kendall and innovative trend method for water quality parameters of the Kizilirmak River, Turkey, *Journal of Hydrology*, **513**, 362–375.

Martinez, J.C., Maleski, J.J. and Miller F.M. (2012), Trends in precipitation and temperature in Florida, USA, *Journal of Hydrology*, **452–453**, 259–281.

Marques, C.F., Ferreira, J., Rocha, A., Castanheira, J.M., Melo-Gonçalves, P., Vaz, N. and Dias, J.M. (2006), Singular Spectrum Analysis and Forecasting of Hydrological Time Series, *Physics and Chemistry of the Earth*, **31**(18), 1172–1179.

Modarres, R. and Silva, V.P.R. (2007), Rainfall trends in arid and semi-arid regions of Iran, *Journal of Arid Environments*, **70**, 344-355.

Mozejko, J. (2012), *Detecting and estimating trends of water quality parameters, water quality monitoring and assessment.* InTech. (Chapter 4).

Önöz, B., and Bayazit, M. (2003), The power of statistical tests for trend detection, Turkish *J. Eng. Env. Sci.*, **27** 247–251.

Sonali, P. and Nagesh, K.D. (2013), Review of trend detection methods and their application to detect temperature changes in India, *Journal of Hydrology*, **476**, 212–227.

Tabari, H. and Talaee, P.H. (2011), Temporal variability of precipitation over Iran: 1966-2005, *Journal of Hydrology*, **396**, 313-320.

Unnikrishnan, P. and Jothiprakash, V. (2015). Extraction of Nonlinear Rainfall Trends Using Singular Spectrum Analysis, *Journal of Hydrologic Engineering*, **20**(15), 10 5015007-1-5015007–15.

von Storch, H. (1995), Misuses of statistical analysis in climate research. In: Storch, H.V., Navarra, A. (Eds.), *Analysis of Climate Variability: Applications of Statistical Techniques*. Springer-Verlag, Berlin, pp. 11–26.

Van Belle, G. and Hughes, J.P., (1984), Nonparametric tests for trend in water quality, *Water Resources Research*, **20** (1), 127–136.

Yakowitz, S. (1973), A stochastic model for daily river flows in an arid region, *Water Resour. Res.*, **9** (5), 1271–1285.

Yu, Y.S., Zou, S. and Whittemore, D. (1993), Non-parametric trend analysis of water quality data of rivers in Kansas, *Journal of Hydrology*, **150** (1), 61–80.

Yue, S., Pilon, P. and Caradias, G. (2002), Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series, *Journal of Hydrology*, 259 (1–4), 254–271.

Zhang, X. and Zwiers, F.W. (2004), Comment on ''Applicability of pre-whitening to eliminate the influence of serial correlation on the Mann–Kendall test'' by Sheng Yue and Chun Yuan Wang, *Water Resources Research*, **40**, W03805. http://dx.doi.org/10.1029/2003WR002073.