GEOGRAPHICAL INSTITUTE "JOVAN CVIJIC" SASA COLLECTION OF PAPERS
№ 54 YEAR 2005

911.2:511.58(23)

Bosko Milovanovic*

STATISTICAL PROCEDURES APPLICATION AND RESULTS OF RESEARCH OF PRECIPITATION ON MOUNTAIN STARA PLANINA

It is justified belief, that scientific knowledge
differs from other types of knowledge
by method for outreaching it.
Zivan Ristic - About scientific research, method and knowledge

Abstract: In this paper, brief review of few statistical procedures, which can be used in climatological researches is given. For some of them, statistical formula is shown. As a numerical example, we have choose mean annual and decadal data for NAO index and amount of precipitation on wider area of mountain Stara planina in period 1951-2000. On a first place, homogeneity of precipitation data from 11 stations which are located on mentioned area is tested by usage of Man-Whitney's test. Fullfilment of lacking data is done with three different methods. By application of different statistical procedures we have try to determine which one of them is the most convenient in explanation of precipitation spatial distribution and connection with NAO index. Application of parametric and nonparametric procedures, shown that there is statistically significant correlation between NAO index and amount of precipitation on both, annual and decadal scale. Relation between amount of precipitation and slope exposition on mountain Stara planina is examined. Linear trend of precipitation is calculated and its statistical significance on a decadal scale in mentioned period is tested.

Key words: Parametric-nonparametric statistical procedures, NAO index, precipitation, slope exposition, mountain Stara planina, Serbia

Introduction

Mountain Stara planina is a part of a massive Carpatho-Balkanian massive. In Serbia (its south-eastern part) just a smaller part is located with area of 1802 km^2. As a morphological entity, on the west, Mt Stara planina is bordered with valleys of rivers Beli Timok, Trgoviski Timok and Visocica, and on the east with state border Serbia and Bulgaria. This territory is characterized with very sharp relief and significant differences in heights above sea level (from 42 m

* Bosko Milovanovanovic, Institute for nature protection of Serbia, Dr Ivana Ribara 91 Novi Beograd, bohumil@verat.net

a.s.l, or 132 m a.s.l. to 2169 m a.s.l). Net of meteorological stations is unequally distributed (the highest station Dojkinci is located on “just” 880 m a.s.l) and often with incomplete time series (Milovanovic, 2005).

The NAO index is defined as the standardized difference between the SLP (sea level pressure) average over the Azores and over Iceland. Actuality of researches concerning influence of NAO on climate of Europe, or some particular regions of Europe (Hurell, 1995), pointed us in examination of possible connection between this climate factor (and some climate modificators as well) and amount of precipitation on Mt Stara planina. Concerning fact that choice and quality of a methodological approach could affect quality of achieved results, basic characteristics and functions of used statistical procedures are shown.

Data

In this paper, data for mean yearly amount of precipitation from eleven stations belonging to Republic Hidrometeorological Institute of Serbia which are located on a wider area of Mt. Stara planina (Aldinac, Dimitrovgrad, Dojkinci, Kamenica Dimitrovgradska, Kalna. Negotin, Pirot, Smilovci, Visocka Rzana, Vitanovac, Zajecar - period 1951-2000) are used. Homogeneity of the data is examined with Man-Whitney nonparametric test. “Using a test based on the ranks of values from a time series has the benefit, that it is not greatly particularly adversely affected by outliers” (Peterson et al, 1998). All stations, except Aldinac and Dojkinci have relatively homogeneous data. Concerning fact that time series are complete on just five stations (Dimitrovgrad, Dojkinci, Kalna, Negotin, Zajecar) fulfillment of lacking data is done. On a first place, coefficient of correlation is calculated among each pair of stations on monthly level. Achieved results on a particular pair of stations for specific month are not obligatory used for the rest of station pairs. Fulfillment of lacking data is done with three different methods (method of reduction on same number of years, method of interpolation based on matrix – basic version with 8 elements of matrix and modified version with 6 elements of a matrix, and method of extrapolation based on an assumption of constant relation between shorter and longer time series).

That combined approach in fulfillment of lacking data is chosen because it has been proven to be more precise. Radovanovic, Milovanovic (2003) quote: “Each of applied and tested methods showed its advantages and disadvantages. The method of reduction demonstrated a great applicability and satisfying accuracy. The method of interpolation showed certain limitations in the possibilities of its application, but also in the accuracy of the data processed. The method of

Table 1. Mean yearly values of NAO index and amount of precipitation on a wider area of Mt Stara planina (1951-2000)

Year	Precipitation (mm)	NAO[1] index	Year	Precipitation (mm)	NAO index
1951	711.0	-0.07	**1976**	861.2	-0.07
1952	721.3	-0.37	**1977**	709.4	-0.21
1953	640.4	0.40	**1978**	717.5	0.21
1954	854.6	0.51	**1979**	689.6	0.19
1955	880.5	-0.64	**1980**	807.6	-0.37
1956	702.3	0.17	**1981**	770.1	-0.09
1957	887.5	-0.02	**1982**	662.9	0.67
1958	536.1	0.12	**1983**	570.6	0.34
1959	682.3	0.49	**1984**	493.0	0.26
1960	720.4	-0.30	**1985**	629.8	-0.47
1961	626.0	1.05	**1986**	618.0	0.56
1962	806.6	-0.13	**1987**	729.0	-0.51
1963	735.6	-0.39	**1988**	577.7	-0.32
1964	731.5	0.24	**1989**	627.6	0.57
1965	660.3	-0.23	**1990**	543.3	1.23
1966	819.2	-0.22	**1991**	648.5	0.34
1967	629.1	0.56	**1992**	544.9	1.11
1968	641.0	-0.62	**1993**	486.9	0.12
1969	736.2	-0.44	**1994**	518.1	0.51
1970	822.7	0.18	**1995**	822.2	-0.61
1971	627.7	-0.55	**1996**	745.0	-1.01
1972	728.7	-0.04	**1997**	665.2	-0.18
1973	702.7	-0.09	**1998**	715.3	0.26
1974	774.5	0.59	**1999**	741.6	0.05
1975	768.4	0.05	**2000**	393.8	0.04

extrapolation proved to have the greatest limitations in its application, but also the highest degree of accuracy. It can be concluded, that in depending of our needs and the completeness of the series of data, the combination of the methods presented in the paper seems to be an adequate approach to data processing".

On an observed area, great variability of a precipitation is spotted. In a period 1951-2000 the largest amount of precipitation was on a station Dojkinci (880 m a.s.l – 865.8 mm). On the other hand, on a station Pirot which is located on 370

[1] NAO index data are taken from www.cru.uea.ac.uk/ftpdata/nao.dat

m a.s.l (i.e. higher altitude than Negotin and Zajecar) there are just 583.3 mm. Such an unequal distribution of precipitation on a relatively small territory is very often in mountain terrains. “In some cases, very low mountain riffs, in complex mountain systems, could be cause of “rain shadows”. Weaken air masses, could lost their humidity on such low barriers (Radovanovic, 2001)”. Mean yearly amount of precipitation from mentioned stations are averaged, so one time series for wider area of Mt Stara planina is calculated. Such approach is chosen because of relatively small area on which precipitation are analyzed, lacking data in time series and inhomogeneities in data base.

Methods and results

In a quantifying of relation between two variables *x, y* (it is not confirmation of a cause-consequence relation), correlation analyzes are used very often. When there is a measurement on interval scale and assumption of normal distribution is satisfied, it is possible to calculate Pearson’s coefficient of correlation *r*. Coefficient of correlation between NAO index and amount of precipitation on a Mt Stara planina on a yearly level (table 1) is -0.37. By application of correlation analyzes on the NAO index data and amount of precipitation from 10 stations in Bulgaria (period 1961-2000), Nikolova (2004) calculated similar values for r (from -0.13 to -0.47).

Testing of null hypothesis (there is no connection between NAO index and amount of precipitation) and by calculation of t-statistic, information about probability that value of -0.37 happened by accident is achieved.

By usage of t-distribution table, we can conclude that probability that null hypothesis is true one is very small (<0.01), so we can reject it and accept alternative hypothesis that there is statistically significant correlation between NAO index and amount of precipitation on Mt Stara planina in period 1951-2000.

Very often it is necessary to aggregate data (i.e. removing the “noise” from relation, or because of purifying the signal we want to determine). Relying on yearly data, we have calculated decadal values for NAO index and amount of precipitation on Mt Stara planina. Time series are now consisted of just five elements (table 2). Coefficient of correlation is now significantly stronger -0.81, but for mentioned number of elements (n=5, df=3) in time series it can not satisfied t-test on a 95% level of confidence.

Table 2: Decadal values of NAO index and amount of precipitation on a wider area of Mt Stara planina

Decade	NAO index	Amount of precipitation on wider area of Mt Stara planina (mm)
1951-1960	0.03	733.6
1961-1970	0.00	721.6
1971-1980	-0.03	738.7
1981-1990	0.22	622.2
1991-2000	0.06	628.2

When there is nonlinear relation between variables, it is possible to transform one or both of variables, or apply some of nonparametric procedures.

Such nonparametric procedures (sometimes called methods independent from distribution), are very useful when the data are far away from normal distribution. Spearman's procedure for correlation is very useful and widely spread nonparametric alternative for Pearson's coefficient of correlation. Basically, information we get is very similar to one we get in usage of Pearson's procedure (perfect positive correlation has value of + 1, perfect negative correlation has a value of -1, values between those two means that correlation is not perfect). This procedure is based on ranks and it is very convenient for the calculation of correlations when the variables are measured on ordinal level (or interval level, but assumption of normal distribution is not satisfied). Process of ranking could be done by calculation of differences between ranks of variables which are joined to particular cases (in descending or ascending order). According to Harnett, Murphy (1975) if squared differences are marked with d_i for every i pair, then:

$$\sum_{i=1}^{n} d_i^{\,2}$$

Value of r_s is calculated

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^{\,2}}{n^3 - n}$$

Applying this equitation on the data from table 2, as a result for r_s, we calculated -0.90. Checking this value in table for Spearman's rank correlation show us that this value is statistically significant and that there is statistically significant relation between decadal values of NAO index and amount of precipitation on wider area of Mt Stara planina.

Table 3: Mean yearly amount of precipitation on a groups of station with different slope exposition (according to Ducic et al, 2003)

	Slope exposition						
	W	SW	S	SE	NE	N	NW
Number of stations	2	11	3	3	4	2	3
Average heights above sea level	430.0	503.2	461.7	600.3	354.3	190.0	436.7
Average amount of precipitation	769.2	649.2	645.8	691.0	668.9	627.7	758.8
Average amount of precipitation on a same height above sea level	959.6	819.2	829.4	827.0	886.5	903.1	949.2

By ranking every group of stations (table 4) they have calculated Pearson's coefficient of correlation (0.97). "It can be concluded that correlation between slope exposition and received amount of precipitation is highly significant and satisfied Student's test on confidence level of 99 %" (Ducic et al, 2003).

Also, we have tried to determine, does the slope exposition (beside atmospheric circulation) could affect precipitation distribution on the Mt Stara planina. We based that research on the results of Ducic et al, (2003) and Milovanovic (2005). Mentioned authors, by analyzing data from 28 meteorological stations on a wider area of Mt Stara planine (period 1961-2000) have determined amount of precipitation on a group of stations with different slope exposition. By usage of pluviometric gradient, they reduced it on a same level above sea level (table 3).

Table 4: Coefficient of correlation between groups of stations with different slope exposition and received amount of precipitation (according to Ducic et al, 2003)

Groups of stations slope exposition	Given value	Amount of precipitation (reduced on 1000 m a.s.l.)
W	7	959.6
NW	6	949.2
N	5	903.1
NE	4	886.5
SE	3	827.0
S	2	829.4
SW	1	819.2
Coefficient of correlation	0.97	

Concerning fact that values for slope exposition are measured on nominal level (seven categories) we have applied chi-squared (goodness of fit) test on the same data. Harnett, Murphy quote: "When the set of outcomes can be divided into two categories, then the appropriate test statistic is binominal variable. When more than two categories or classes of outcomes are involved, then the appropriate statistic is the chi-square variable".

Value for chi-squared value is calculated:

$$\chi^2{}_{(c-1)d.f} = \sum_{i=i}^{c} \frac{(O_i - E_i)^2}{E_i}$$

Where:

c-number of categories
O_i – number of observed frequencies,
E_i - number of expected frequencies,

The null and alternative hypothesis are:

H_o: E_1=E_2…E_7=882,
H_a: Frequencies are not all equal

Value of chi-squared statistic is 23.51. Checking this value in table for chi-squared distribution, it can be concluded that on a level of significance of

$\alpha<0.005$ it is justified to reject null hypothesis (slope exposition does not affect received amount of precipitation), and accept alternative hypothesis (there is significant influence of slope exposition on received amount of precipitation).

After we have determined statistically significant correlation (by application of Spearman's procedure) between decadal values of NAO index and amount of precipitation on area of Mt Stara planina, we have tried to answer the question is there any trend in analyzed data and characteristics of the changes of observed climatological element in period 1951-2000. We have applied Mann-Kendall's test. According to Salmi et al (2002) this test is very convenient when monotonic trend (without seasonal or cyclic variations) exist in the data, while the Sen's method is very useful in slope estimation. Those methods could be applied on the time series with lacking data, and on the time series independent from distribution. Also, they are not sensitive on errors and outliers.

Sen's slope estimation is calculated[2]

$$Q_i = \frac{x_j - x_k}{j - k}$$

Where j>k

Sen's slope estimation Q, shows changes per time unit, and represent median N of Q_i values. N is calculated:

$$N = \frac{n(n-1)}{2}$$

Application of this equitation on the data from table 2, gives the result matrix (table 5).

Table 5. Matrix of calculated values Q_i –Sen's slope estimations

-12.06	2.54	-37.14	-26.37
	17.14	-49.68	-31.14
		-116.50	-55.28
			5.94

[2]http://www.fmi.fi/organisation/kontakt_11.html

Median from table 5 is 28.75 mm per decade.

Calculation of Mann-Kendall's test statistic S (for time series with less than 10 elements) is used for testing the statistical significance of trend. It is calculated as:

$$\sum (xj - xk) \begin{Bmatrix} 1 \text{ if } x_j - x_k > 0 \\ 0 \text{ if } x_j - x_k = 0 \\ -1 \text{ if } x_j - x_k < 0 \end{Bmatrix}$$

This one, and Vilkokson's test as well are used in researches of frequencies of extreme temperature events in Central and South-eastern Europe by Domonkos et al (2003) and particularly Szinell (2003) with detailed explanations.

Application of equitation for S statistic on the data for decadal amount of precipitations (table 2) shows value of -4. According to that (table 6) it can be concluded that there is no statistically significant trend in the decadal amount of precipitation on a wider area of Mt Stara planina (we can accept null hypothesis because there is negative trend, and value for S is higher than critical value for S_{max}).

Table 6. Mann-Kendall S statistic table

Number of elements N	Range of S	S_{max} α=0,2
4	-6 to 6	-4
5	-10 to 10	-5
6	-15 to 15	-6
7	-21 to 21	-7
8	-28 to 28	-8
9	-36 to 36	-9
10	-45 to 45	-10

Summary

The use of statistical procedures in climatology is necessity, from which it is possible to extract detailed and precise information about certain climatological elements, or relation between them. Brief review of basic characteristics and purposes of certain statistical procedures, which is given in this paper, can be used for easier application in climatological researches. Homogeneity of mean annual precipitation data examination (by Man-Whitney's nonparametric test) shown, that of eleven observed stations which are located on a wider area of Mt Stara planina, nine of them have relatively homogeneous data (exceptions are stations Aldinac and Dojkinci). Fulfillment of lacking data is done with three different methods (method of reduction on same number of years, method of interpolation based on matrix – basic version with 8 elements of matrix and modified version with 6 elements of a matrix, and method of extrapolation based on an assumption of constant relation between shorter and longer time series). Such approach is chosen because of better accuracy of metadata. Because of relatively small area, relatively inhomogeneous time series, data from eleven stations are averaged, so one time series (on a yearly and decadal scale) is calculated for whole area of Mt Stara planina.

In numerical example, first we have calculated Pearson's coefficient of correlation (value for r = -0.37, which is statistically significant on 99% confidence level) among NAO index and annual values of amount of precipitation on mountain Stara planina. Application of same method on decadal scale, despite the growth of correlation coefficient (r = -0.81), shown that for n=5, df=3 there is no statistically significant signal on 95% confidence level. As a nonparametric alternative, we used Spearman's procedure, and on the decadal scale we found statistically significant signal. Usage of chi-square test on a very high confidence level (above 99%) confirm influence of slope exposition on received amount of precipitation on a wider area of Mt Stara planina. Sen's slope estimation shown that change per decade of amount of precipitation on mountain Stara planina in period 1951-2000 is -28.75 mm. Application of Mann-Kendall's nonparametric test shown that observed decadal changes are not statistically significant in mentioned period.

References

Domonkos P, Kysel J, Piotrowicz K, Petrovic P, Likso T. (2003): Variability of extreme temperature events in South-Central Europe during the 20th century and its relationship with large scale circulation. *International Journal of Climatology* 23, p. 987-1010.

Ducic V, Radovanovic M, Milovanovic B. (2003): Regional disposition of precipitation on Stara planina to dependence from exposition and elevation. *Bulletin of Geographical Faculty* LI, Belgrade, p. 39-55.

Milovanovic B. (2005): Climate regionalization of mountain Stara planina. Master thesis, Faculty of Geography, Belgrade.

Nikolova N. (2004): Rainfall variability in Bulgaria and its relation with North Atlantic Oscillation. BALWIOS-2004, Ohrid, FY Republic of Macedonia, 25-29 May 2004 (http://balwois.mpl.ird.fr/balwois/administration/full_paper/ffp-1o-047.pdf).

Peterson T, Easterling D, Karl T, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Boehm R, Gullett D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salinger J, Førland E, Bauer I. H, Alexandersson H, Jones P, Parker D. (1998): Homogeneity adjustment of in situ atmospheric climate data – a review. *International Journal of Climatology* 18, p. 1493-1517.

Radovanovic M. (2001): Influence of relief and atmospheric circulation on climate differentiation in Serbia. PhD thesis, Faculty of Geography, Belgrade.

Radovanovic M, Milovanovic B. (2003): Methods to Complete the Missing Data on Precipitation in the Mountains of Serbia – Testing and Application. Studia Geograficzne 75, No 2542, p. 593-598, International conference “Man and climate in the 20th Century”, Wroclaw, 13–15 June 2002. – Institute of Geography, University of Wroclaw and Polish Geophisycal Society of Marshall of the Voivod of Lower Silesia, Wroclaw.

Rimbu N, Boronean C, Buta C, Dima M. (2002): Decadal variability of the Danube river flow in the lower basin and its relation with the North Atlantic Oscillation. International *Journal of Climatology* 22, p. 1169-1179.

Ristic Z. (2006): About scientific research, method and knowledge. Institute of Pedagogy, Belgrade.

Salmi T, Määttä A, Anttila P, Airola T, Amnell T. (2002): Detecting trends of annual values of atmospheric pollutants by the Mann-Kendall test and Sen's slope estimates-the excel template application makesens. Finish Meteorological Institute, Helsinki, Finland http://www.fmi.fi/organisation/kontakt_11.html.

Szinell C, Bussay A, Szentimrey T. (1998): Drought tendencies in Hungary, International *Journal of Climatology* 18, p. 1479–1491.

Harnett D, Murphy J. (1975): Introductory statistical analysis. Addison-Wesley Publishing company, INC, Reading, Massachusetts.

Hurrell J. (1995) Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science, no. 269*, 676-679.