Union of Mathematicians of Macedonia - ARMAGANKA

IX SEMINAR OF DIFFERENTIAL EQUATIONS AND ANALYSIS

and

1st CONGRESS OF DIFFERENTIAL EQUATIONS, MATHEMATICAL ANALYSIS AND APPLICATIONS

CODEMA 2020

Proceedings of the CODEMA 2020

Зборник на трудови од CODEMA 2020

Skopje, 2021

DETERMINATION OF CERTAIN PARAMETERS IN HYDROLOGY THROUGH STATISTICS

ISBN 978-608-4904-09-0 UDC:519.248:556.04/.06(497.7)

Violeta Gjeshovska, Daniel Velinov, Sasha Jadrovski

Abstract. Statistics uses predetermined methods for collecting and analyzing data, mainly based on probability theory. In hydrology they are used for the analysis of various hydrological and meteorological data for some past period, as well as for making decisions and conclusions about the regime in future.

The approach to statistical data processing does not have a single form in hydrology. The choice of the method depends on the accuracy or the level of confidence that this data should be analyzed i.e. interprets the course of an occurrence, or the conclusion we want to deduce for this phenomenon combined with other hydrological phenomena.

Because hydrological data are mostly limited data, statistics is the main discipline that allows obtaining complete data from the data and draws a conclusion concerning the characteristics of hydrological phenomena. The purpose of applying statistics to the hydrological variables that have been registered in the past is to determine the probability with which these variables would appear in the future.

In this paper, using old and new statistical methods, the unknown values of the annual average temperatures for five cities Skopje, Stip, D.Kapija, Prilep and Bitola in the period from 1925 to 2000 are determined. The obtained values lead us to some conclusions about their adaptability and confidentiality.

1. INTRODUCTION

Hydrology is a science for the water regime on earth surface, in atmosphere and into the soil. The processes of atmosphere's water transformation on earth surface, from the earth surface under the surface and vice versa, i.e. its mutual influence within natural environment, in every aggregate state, are main subject of investigation in this scientific field.

There are historical data for water behavior in the nature, i.e. measurements made with previously defined purpose and goal. These data form a sequence are used to obtain a certain rule for the observed phenomenon. For the measuring and observation of the parameters (biological, metrological etc.) it is formed a net of measuring stations equipped with specialized measuring equipment.

2010 *Mathematics Subject Classification.* Primary: 62P30, Secondary: 65C20, 65C60.

Key words and phrases. statistical analysis, correlation, coefficient of correlation, regression, hydrology.

The data from the monitoring net are not always available for some observed palaces or it may be available but with interceptions (deficiency) in the sequences.

Therefore, these data are not always enough; the sequences are too short or discontinuous for the analysis that we intend to do. So, in the hydrology, mathematical methods from the field of statistics and probability theory are used very often in order to fulfill (to add missing members) or continue the sequences of data. This can be done by searching a connection of at least two hydrological phenomena with measured data (sequences), such that one of them is sufficiently large and continuous.

The procedure of continuation of the sequences requests determining the strengthens of the connection between two or more hydrological phenomena (one is always with sufficiently long and continuous sequence of measured data).

The connection between hydrological and meteorological phenomena in hydrology is given with certain correlation (here under correlation we understand a stochastic connection of two or more variables).

Hydrological analysis and exact data for design of a hydrotechnical object can be made exclusively with sufficiently long and continuous sequence.

The main goal of this paper, in point of view of application, is to study the hydrological sequences of annual average temperatures at the measuring stations in Prilep, Bitola, Demir Kapiija, Stip and Skopje, to analyze the observed data, to make a decision which measuring station we should take for correlation, to fulfill the sequences, to determine and analyze the strength of the connections through coefficient of correlation and regression.

The subject of investigation in this paper is the quality of the data obtained from the measurable stations Prilep, Bitola, Demir Kapija, Stip and Skopje through statistical analysis.

The obtained data from the previously mentioned measuring stations have gaps in measuring in different time intervals. These gaps will be fulfilled with certain data obtained with linear and nonlinear correlation. For that purpose, it is made a correlation with the nearest measuring stations. These correlations are tested in order to be obtained the correlation with the biggest coefficient of correlation. In continuation, analysis of the homogeneity of the sequences of data (Kolmogorov-Smirnov test) is made and then the standard statistical parameters as arithmetic mean, standard deviation, coefficients of asymmetry and variation, coefficients of correlation and regression are determined.

There are analyzed different type of correlations (linear and nonlinear). The simple regression, i.e. dependence between two variables x and y, one dependent variable (y) and one independent variable (x) and between three variables, one dependent variable (y) and two independent variables (x_1 and x_2). The statistical testing of the obtained results is made with the standard and recommended test in these types of analysis (normalized z-test, Student's t-test, Fisher's F-test and χ^2 – test).

Analysis presented in this paper has theoretical and practical value. There given methods for determining correlation and regression in hydrology. Use of quality data for sequence continuation is very important and significant in preparation of hydrological analysis.

Application of these statistical methods can be found in many areas as ecomony [22], industry [3], thermometric [18]-[20], sociology, education, mining, agriculture, meteorology etc. The application of the statistical methods in hydrology is pretty deep and development of this field is directly dependent from usage of these methods. Since hydrology is based on observed and measured data of the investigated phenomena, processing of the obtained data is made with some statistical methods.

In regression analysis we should know at the begging which of the phenomena are independent variables, respectively dependent variables. Here the main aim is to find the form (in mathematical sense) of the connection, i.e. the formula giving the dependence of observed phenomena.

Correlation and regression are methods for describing the dependence of two or more variables. Using these methods in any field of investigation, can be stated scientifically, with certain significance, new and important results in these fields. In every case, it is important, all the influencing parameters to be taken into consideration.

2. METHODOLOGY

Hydrological sequences are series of empiric data obtained with observations and measurements of certain hydrological phenomena as rains, water levels, flows etc. The well-known parameters of a hydrological sequence are arithmetic mean $\bar{x}$, module coefficient K_i, median m, mode μ, standard quadratic deviation σ, coefficient of variation C_v, whose explanations and formulas we are going to skip, since they are very well known and basic terms in this theory. We will shortly give the more important facts about the coefficient of asymmetry, recently used in hydrology.

Let we are comparing two sequences. These two sequences can have same deviations, but different signs. Therefore, a parameter determining the degree of symmetry is introduced, known as coefficient of asymmetry. This coefficient is given by the following formula

$$C_s = \frac{\sum_{i=1}^{i=n} (K_i - 1)^3}{nC_v^3} \;.$$

For accurate calculation of this expression, it is necessary for the sequence to have at least 60 members. For shorter sequences this coefficient can be calculated empirical by the following formula

$$C_s = 2C_v \,.$$

For sequences with rare occurrence, this coefficient is given by

$$C_S = \frac{3}{6} C_V .$$

From the distribution of Pirson, we have the following

$$C_S = \frac{2C_V}{1 - K_{\min}} ,$$

where $(K_{\min})$ is minimal module coefficient. Using the quantity of this coefficient, it can be said the following: if $(C_S = 0)$ the sequence is symmetric; if $(0 < C_S < 0{,}1)$ there is no asymmetry; if $(0{,}1 < C_S < 0{,}25)$ asymmetry of the sequence is small; if $(0{,}25 < C_S < 0{,}5)$ asymmetry of the sequence is medium and if $(C_S > 0{,}5)$ asymmetry of the sequence is large.

When we are dealing with two or more hydrological phenomena usually exists some connection, which can be very weak to very strong. As for an example, we can state the connection between rainfall and surface leakage in a certain catchment area, although these phenomena are random in space and time. Also, there is connection on the same hydrological phenomenon measured on two near measurable stations (for example, flow of a river, measured on different profiles). The strength of this connection will be called correlation. The main task of the correlation analysis is to find a way in which one the influence of the independent variable to dependent variable is given.

The dependence between dependent variable (Y) and independent variable (X) in hydrology can be functional $Y = f(X)$, correlative $Y[Y/X]$ and stochastic.

If some certain value of the dependent variable is connected with many values of the independent variable, then this value is called a correlational value. Correlational dependences can be linear and nonlinear. Correlational dependences can be defined between dependent variable (Y) and one (X), or more independent variables (X_i). Concerning the number of variables, the correlational dependences are simple (correlation between two variables) and multiple (correlation between multiple variables). The simple correlations are formed with two random variables, one dependent Y and one independent variable X. The multiple correlation dependences are formed with one dependent variable Y and two or more independent variables X_i, where $i = 1, 2, ..., m$.

The correlation with one independent variable is given by

$$Y = a + b \cdot X ,$$

where $a = \overline{Y}$ and $b = \dfrac{S_{xy}}{S_{xx}}$.

where $\overline{X}$ is the average value of the random variable X, $\overline{Y}$ is the average value of the random variable Y, S_{xx} is the sum of the residual squares of the variable X , S_{yy} is the sum of the residual squares of the variable Y , S_{xy} is the covariance of the residuals of the random variables X and Y .

The correlation with two independent variables is given by

$$Y = a + b_1(X_1 - \overline{X_1}) + b_2(X_2 - \overline{X_2})$$

where a, b_1, b_2 are regression parameters found by the method of least squares, which is quite standard approach in regression analysis. The regression parameters are given by

$$a = \overline{Y}$$

$$b_1 = \frac{S_{x_2x_2} \cdot S_{x_1y} - S_{x_1x_2} \cdot S_{x_2y}}{S_{x_1x_1} \cdot S_{x_2x_2} - S_{x_1x_2}^2}$$

$$b_2 = \frac{S_{x_1x_1} \cdot S_{x_2y} - S_{x_1x_2} \cdot S_{x_1y}}{S_{x_1x_1} \cdot S_{x_2x_2} - S_{x_1x_2}^2},$$

where $\overline{X_1}$ is the average value of the random variable X_1, $\overline{X_2}$ is the average value of the random variable X_2, $\overline{Y}$ is the average value of the random variable Y , $S_{x_1x_1}$ is the variance of the random variable X_1, $S_{x_2x_2}$ is the variance of the random variable X_2 , S_{yy} is the variance of the random variable Y , S_{x_1y} is the covariance of the random variables X_1 and Y , S_{x_2y} is the covariance of the random variables X_2 and Y , $S_{x_1x_2}$ is the covariance of the random variables X_1 and X_2 .

The regression equation describes a way how the variable Y is dependent of one or more independent random variables X_i . In defining and obtaining a regression, as was mentioned before, the method of least squares, is one of the most investigated and applied techniques in applications.

The main difference between correlation and regression is that in correlation, you sample both measurement variables randomly from a population, while in regression you choose the values of the independent (X) variable.

The dependence between Y_i and X_i can be written as

$$y_i = a + b \cdot (x_i - \overline{X}) + e_i , \qquad i = 1, 2, ..., n$$

where $\overline{X}$ is the average value of the random variable X , e_i are the errors (errors of the model and errors of the measurement of the variable Y_i), a, b are parameters (coefficients) which can be determined with data obtained by the measurement, n is the number of measurements of X_i and Y_i .

The parameters a and b need to be determined from the variables X_i and Y_i . Usually, when we are dealing with applications, we have limited number of measured data n, and obtained parameters, denoted with a and $\hat{b}$, are their estimates, not exact values of the parameters a and b. Calculated value y_i, obtained by regression, by using exact values x_i and estimated parameters a and $\hat{b}$, is again just an estimate (most probable) value and is denoted by y_i. Therefore, the regression is given by

$$y_i = a + \hat{b} \cdot (x_i - \overline{X}), \quad \text{where } i = 1, 2, ..., n.$$

The estimation of the parameters a and $\hat{b}$ is made by the method of „least squares", which one guaranties minimal sum of the squares of the differences between real y_i and calculated y_i values of the dependent variable, i.e.

$$\sum_{i=1}^{n} (y_i - y_i)^2 \to \min .$$

In the case of nonlinear regression with two variables (three parameters) for sequences of the measured data (x_i) and (y_i) it is assumed that the function has a form $y = f(x)$ or $y = f(x, a, b, c, ...)$. Parameters of the functions are determined in a way that the squares of the differences of the observed points and ordinates of the curve are minimal. In order to get minimal value of the sum of the squares of the differences, it is needed all partial derivatives by all variables to be zero.

3. APPLICATIONS

The discussed techniques above, are going to be applied on hydrological sequences for average annual temperatures on the air for the measuring stations Prilep, Demir Kapija, Bitola, Stip and Skopje. For these measuring stations we have sequences for the period from 1925 to 2000. In these measuring stations there are interruptions in measuring in different time periods. For each sequence it is made statistical analysis by finding basic statistical parameters (see Table 1). The analysis on the homogeneity of the sequences of annual sums of the temperatures for the period $1925-2000$ for measuring stations Prilep, Bitola, Demir Kapija, Stip, using different tests for the hypothesis (normalized z-test, Student's t-test, Fisher's F - test) is given.

Table 1. Statistical parameters for hydrological sequences

Hydr. measure. /Station	**Prilep**	**D. Kapija**	**Bitola**	**Skopje**	**Stip**
Period	1925-2000	1925-2000	1925-2000	1925-2000	1925-2000
Missing data	1925-2000	1925-2000	1925-2000	1925-2000	1925-2000
Number of data- n	69	63	65	67	69
t_a	$11,389$	$13,808$	$11,281$	$12,317$	$12,877$

$t_{\min}$	10,125	12,558	10,333	10,825	11,200
$t_{\max}$	13,217	15,300	12,992	14,267	14,300
C_s	0,512	0,455	0,319	0,383	0,307
σ	0,721	0,676	0,565	0,648	0,677
C_v	0,063	0,457	0,319	0,053	0,053

3.1. Measuring station Prilep

The sequence of data for average air temperatures for city Prilep from 1925 till 2000 has 69 entries, and entries missing for 1925,1941,1943,1944,1946,1953 and 1954. According to Fisher's exact test, the sequence of annual sum temperatures for $1925-2000$ at the measuring station Prilep, can be accepted that the sequence is homogeneous with significance level of 5%, meaning that there is 5% risk of concluding that difference exists when there is no actual difference. To fill the gaps for this period, correlative connection with one variable and correlative connection with two variables with measuring stations Stip (X_1) and Demir Kapijia (X_2) are made.

In establishing correlative connections with two variables the order of including the variables is made accordingly the values of the coefficient of correlation between the dependent variable (Y) and the independent variable $(X)-(r_{xy})$. Since $r^2_{x_2y}=0{,}828>r^2_{x_1y}=0{,}822$, the first variable which will be included in the correlation is the independent variable (X_2), i.e. it will be used the data for Demir Kapija. The correlation equation with one independent variable (X_2), $y=a'+b\cdot X_2$ for the measuring station Prilep is the following $y=-0{,}82+0{,}88X_2$. The correlation equation with two variables (X_1) and (X_2), $y=a+b_1(X_1-\overline{X_1})+b_2(X_2-\overline{X_2})$ for the measuring station Prilep is the following $y=11{,}39+0{,}42(X_1-12{,}88)+0{,}5(X_2-13{,}81)$.

The regression equation with two variables (dependent variable Y (Prilep) and independent variable X (Demir Kapija) is given with linear and polynomial degree. The linear regression model between Prilep and Demir Kapija is given by $y=0{,}9003x-1{,}085$ and the polynomial regression with degree 2 is given by $y=0{,}0578x^2-0{,}7377x+10{,}307$. Since, these sequences are not continuous, the existing gaps are fulfilled only for the period where measuring data exists (1946,1953 and 1954).

For comparison, the obtained results are given in Table 2 and the graphical data are presented in annual average air temperatures for measuring station Prilep, where the missing data are fulfilled by linear and polynomial regression in Figure 1.

3.2. Measuring station Bitola

The sequence of data for average air temperatures for city Bitola from 1925 till 2000 has 65 entries, and entries missing for $1925-1927, 1934, 1939-1941$, $1944, 1945, 1952$ and 1953. According to Fisher's exact test, the sequence of annual sum temperatures for $1925-2000$ at the measuring station Bitola, it can be accepted that the sequence is homogeneous with significance level of 5%.

To fill the gaps for this period, correlative connection with one variable (X_1) Demir Kapija and correlative connection with two variables with measuring stations Demir Kapija (X_1) and Stip (X_2) are made.

Since $r^2_{x_1 y} = 0,823 > r^2_{x_2 y} = 0,820$, the first variable which will be included in the correlation is the independent variable (X_1), i.e. it will be used the data for Demir Kapija. The correlation equation with one independent variable (X_1), $y = a' + b \cdot X_1$ for the measuring station Bitola is the following $y = 1,79 + 0,69 X_1$. The correlation equation with two variables (X_1) and (X_2), $y = a + b_1(X_1 - \overline{X_1}) + b_2(X_2 - \overline{X_2})$ for the measuring station Bitola is the following $y = 11,28 + 0,37(X_1 - 13,81) + 0,34(X_2 - 12,88)$.

The regression equation with two variables (dependent variable Y (Bitola) and independent variable X (Stip) is given with linear and polynomial degree. The linear regression model between Bitola and Stip is given by $y = 0,8116x + 0,84$ and the polynomial regression with degree 2 is given by $y = 0,0199x^2 + 0,2946x + 4,1922$. Since, these sequences are not continuous, the existing gaps are fulfilled only for the period where measuring data exists ($1927, 1934, 1940, 1944, 1945, 1952$ and 1953). For comparison, the obtained results are given in Table 3 and the graphical data are presented in annual average air temperatures for measuring station Bitola, where the missing data are fulfilled by linear and polynomial regression in Figure 2.

3.3. Measuring station Demir Kapija

The sequence of data for average air temperatures for city Demir Kapija from 1925 till 2000 has 63 entries, and entries missing for $1925-1933, 1941$, $1943-1945$. According to Fisher's exact test, the sequence of annual sum temperatures for $1925-2000$ at the measuring station Demir Kapija, it can be accepted that the sequence is homogeneous with significance level of 5%.

To fill the gaps for this period, correlative connection with one variable (X_1) Stip and correlative connection with two variables with measuring stations Stip (X_1) and Skopje (X_2) are made.

Since $r^2_{x_1y} = 0,917 > r^2_{x_2y} = 0,841$, the first variable which will be included in the correlation is the independent variable (X_1), i.e. it will be used the data for Stip. The correlation equation with one independent variable (X_1), $y = a' + b \cdot X_1$ for the measuring station Demir Kapija is the following $y = 2,01 + 0,92X_1$. The correlation equation with two variables (X_1) and (X_2), $y = a + b_1(X_1 - \overline{X_1}) + b_2(X_2 - \overline{X_2})$ for the measuring station Demir Kapija is the following $y = 13,81 + 0,77(X_1 - 12,88) + 0,17(X_2 - 12,32)$.

The regression equation with two variables (dependent variable Y (Demir Kapija) and independent variable X (Stip) is given with linear and polynomial degree. The linear regression model between Demir Kapija and Stip is given by $y = 0,9604x + 1,4446$ and the polynomial regression with degree 2 is given by $y = 0,0958x^2 - 1,5208x + 17,473$. Since, these sequences are not continuous, the existing gaps are fulfilled only for the period where measuring data exists ($1927 - 1929, 1931, 1933, 1943 - 1945$).

For comparison, the obtained results are given in Table 4 and the graphical data are presented in annual average air temperatures for measuring station Demir Kapija, where the missing data are fulfilled by linear and polynomial regression in Figure 3.

3.4. Measuring station Stip

The sequence of data for average air temperatures for city Stip from 1925 till 2000 has 70 entries, and entries missing for $1925, 1926, 1930, 1932, 1936, 1941$ and 1945. According to Fisher's exact test, the sequence of annual sum temperatures for $1925 - 2000$ at the measuring station Stip, it can be accepted that the sequence is homogeneous with significance level of 5%.

To fill the gaps for this period, correlative connection with one variable (X_1) Demir Kapija and correlative connection with two variables with measuring stations Demir Kapija (X_1) and Skopje (X_2) are made.

Since $r^2_{x_1y} = 0,917 > r^2_{x_2y} = 0,873$, the first variable which will be included in the correlation is the independent variable (X_1), i.e. it will be used the data for Stip. The correlation equation with one independent variable (X_1), $y = a' + b \cdot X_1$ for the measuring station Stip is the following $y = 0,19 + 0,92X_1$. The correlation equation with two variables (X_1) and (X_2), $y = a + b_1(X_1 - \overline{X_1}) + b_2(X_2 - \overline{X_2})$ for the measuring station Stip is the following $y = 12,88 + 0,63(X_1 - 13,81) + 0,36(X_2 - 12,32)$.

The regression equation with two variables (dependent variable Y (Stip) and independent variable X (Demir Kapija) is given with linear and polynomial degree. The linear regression model between Stip and Demir Kapija is given by $y = 0,934x - 0,026$ and the polynomial regression with degree 2 is given by $y = -0,0134x^2 + 1,308x - 2,6298$. Since, these sequences are not continuous, the existing gaps are fulfilled only for the period where measuring data exists (1936 and 1947).

For comparison, the obtained results are given in Table 5 and the graphical data are presented in annual average air temperatures for measuring station Stip, where the missing data are fulfilled by linear and polynomial regression in Figure 4.

3.5. Measuring station Skopje

The sequence of data for average air temperatures for city Skopje from 1925 till 2000 has 67 entries, and entries missing for 1934,1941 and 1943–1949. According to Fisher's exact test, the sequence of annual sum temperatures for 1925–2000 at the measuring station Skopje, it can be accepted that the sequence is homogeneous with significance level of 5%.

To fill the gaps for this period, correlative connection with one variable (X_2) Demir Kapija and correlative connection with two variables with measuring stations Stip (X_1) and Demir Kapija (X_2) are made.

Since $r^2_{x_2 y} = 0,873 > r^2_{x_1 y} = 0,841$, the first variable which will be included in the correlation is the independent variable (X_2), i.e. it will be used the data for Demir Kapija. The correlation equation with one independent variable (X_2), $y = a' + b \cdot X_2$ for the measuring station Skopje is the following $y = 1,54 + 0,83X_2$. The correlation equation with two variables (X_1) and (X_2), $y = a + b_1(X_1 - \overline{X_1}) + b_2(X_2 - \overline{X_2})$ for the measuring station Skopje is the following $y = 12,32 + 0,24(X_1 - 13,81) + 0,62(X_2 - 12,88)$.

The regression equation with two variables (dependent variable Y (Skopje) and independent variable X (Stip) is given with linear and logarithmic function. The linear regression model between Skopje and Stip is given by $y = 0,8509x + 1,3671$ and the logarithmic regression is given by $y = 10,976 \ln x - 15,708$. Since, these sequences are not continuous, the existing gaps are fulfilled only for the period where measuring data exists (1934,1943–1946,1948 and 1949).

For comparison, the obtained results are given in Table 6 and the graphical data are presented in annual average air temperatures for measuring station Skopje, where the missing data are fulfilled by linear and polynomial regression in Figure 5.

Table 2 – Comparison between obtained results for Prilep with correlation and regression

	Period	1946	1953	1954
Correlation	$y=-0,82+0,88X_2$			
	D. Kapija $=x_2$	15,0	13,5	13,5
	Prilep $=y$	12,4	11,1	11,1
	$y=11,39+0,42(X_1-12,88)+0,50(X_2-13,81)$			
	Stip $=x_1$	14,0	12,6	12,5
	D. Kapija $=x_2$	15,0	13,5	13,5
	Prilep $=y$	12,5	11,1	11,1
	Period	1946	1953	1954
Regression	$y=0,9003x-1,085$			
	D. Kapija $=x$	15,0	13,5	13,5
	Prilep $=y$	12,4	11,1	11,1
	$y=0,0587x^2-0,7377x+10,307$			
	D. Kapija $=x$	15,0	13,5	13,5
	Prilep $=y$	12,5	11,0	11,0

Table 3 – Comparison between obtained results for Bitola with correlation and regression

	Period	1927	1934	1939	1940	1944	1945	1952	1953
Correlation	$y=1,79+0,69X_2$								
	D.Kapija		15,2	14,6	12,7			15,3	13,5
	Bitola $=y$		12,2	11,8	10,5			12,3	11,1
	$y=11,28+0,37(X_1-13,81)+0,34(X_2-12,88)$								
	D.Kapija $=x_1$		15,2	14,6	12,7			15,3	13,5
	Stip $=x_2$		13,9	13,3	11,2			14,2	12,6
	Bitola $=y$		12,2	11,7	10,3			12,3	11,1
Regression	$y=0,8116x+0,84$								
	Stip $=x$	14,3	13,9	13,3	11,2	12,8	13,1	14,2	12,6
	Bitola $=y$	12,4	12,1	11,6	9,9	11,2	11,5	12,4	11,1
	$y=0,0199x^2+0,2946x+4,1922$								
	Stip $=x$	14,3	13,9	13,3	11,2	12,8	13,1	14,2	12,6
	Bitola $=y$	12,5	12,1	11,6	10,0	11,2	11,5	12,4	11,1

Table 4 – Comparison between obtained results for D. Kapija with correlation and regression

	Period	1927	1928	1929	1931	1933	1943	1944	1945
Correlation	$y = 2,01 + 0,92X_1$								
	Stip = x_1	14,3	13,6	12,8	12,6	11,8	13,9	12,8	13,1
	D.Kapija = y	15,1	14,5	13,7	13,6	12,8	14,7	13,7	14,0
	$y = 13,81 + 0,77(X_1 - 12,88) + 0,17(X_2 - 12,32)$								
	Stip = x_1	14,3	13,6	12,8	12,6	11,8			
	Skopje = x_2	13,6	12,9	11,6	12,3	11,8			
	D.Kapija = y	15,1	14,5	13,6	13,6	12,9			
Regression	$y = 0,9604x + 1,4446$								
	Stip = x	14,3	13,6	12,8	12,6	11,8	13,9	12,8	13,1
	D.Kapija = y	15,2	14,5	13,7	13,5	12,8	14,8	13,7	14,0
	$y = 0,0958x^2 - 1,5208x + 17,473$								
	Stip = x	14,3	13,6	12,8	12,6	11,8	13,9	12,8	13,1
	D.Kapija = y	15,3	14,5	13,7	13,5	12,9	14,8	13,7	14,0

Table 5 – Comparison between obtained results for Stip with correlation and regression

	Period	1936	1947
Correlation	$y = 0,19 + 0,92X_1$		
	D. Kapija = x_1	14,8	14,4
	Stip = y	13,8	13,4
	$y = 12,88 + 0,63(X_1 - 13,81) + 0,36(X_2 - 12,32)$		
	D. Kapija = x_1	14,8	
	Skopje = x_2	13,0	
	Stip = y	13,7	
	Period	1936	1947
Regression	$y = 0,934x - 0,026$		
	D. Kapija = x	14,8	14,4
	Stip = y	13,8	13,4
	$y = -0,0134x^2 + 1,308x - 2,6298$		
	D. Kapija = x	14,8	14,4
	Stip = y	13,8	13,4

Table 6 – Comparison between obtained results for Skopje with correlation and regression

	Period	1934	1943	1944	1945	1946	1948	1949
Correlation	$y = 1,54 + 0,83X_2$							
	Stip $= x_2$	13,9	13,9	12,8	13,1	14,0	12,6	12,6
	Skopje	13,2	13,2	12,3	12,5	13,3	12,1	12,1
	$y = 12,32 + 0,24(X_1 - 13,81) + 0,62(X_2 - 12,88)$							
	D.Kapija $= x_2$	15,2				15,0	13,8	13,6
	Stip $= x_1$	13,9				14,0	12,6	12,6
	Skopje $= y$	13,3				13,3	12,1	12,1
Regression	$y = 0,8509x + 1,3671$							
	Stip$= x$	13,9	13,9	12,8	13,1	14,0	12,6	12,6
	Skopje $= y$	13,2	13,2	12,3	12,5	13,3	12,1	12,1
	$y = 10,976 \ln x - 15,708$							
	Stip $= x$	13,9	13,9	12,8	13,1	14,0	12,6	12,6
	Skopje $= y$	13,2	13,2	12,3	12,5	13,3	12,1	12,1

Figure 1 – Average yearly air temperatures in Prilep with results obtained linear and nonlinear correlation and regression

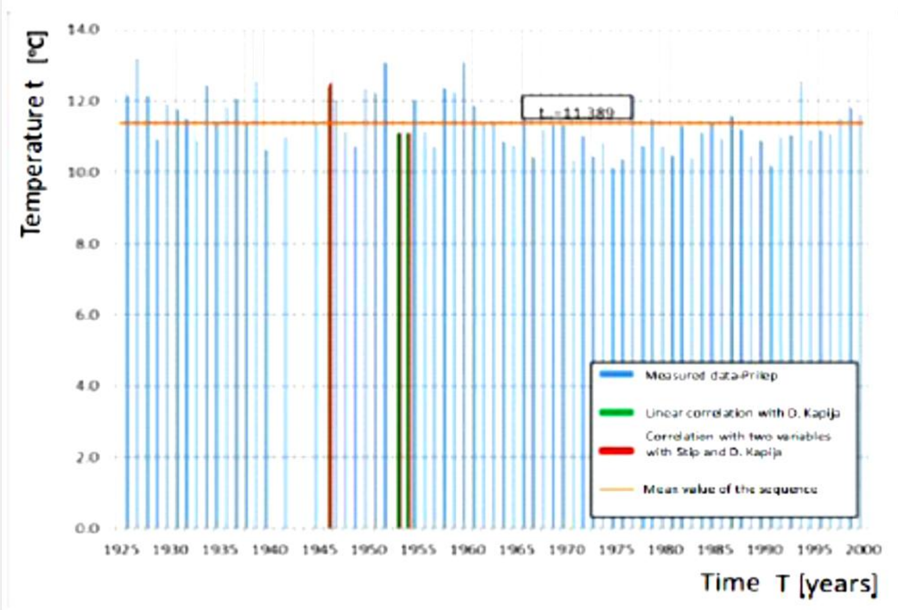


Figure 2 - Average yearly air temperatures in Bitola with results obtained linear and nonlinear correlation and regression

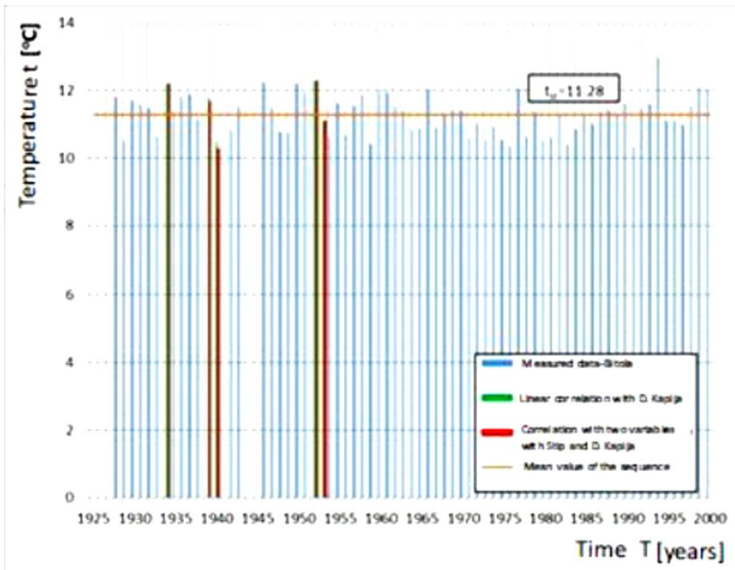


Figure 3 - Average yearly air temperatures in D. Kapija with results obtained linear and nonlinear correlation and regression

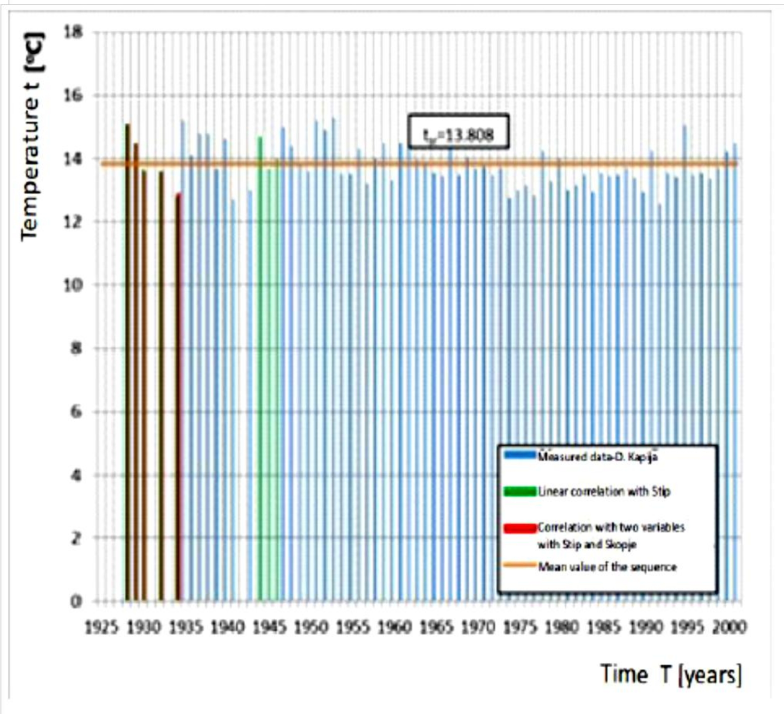


Figure 4 - Average yearly air temperatures in Stip with results obtained linear and nonlinear correlation and regression

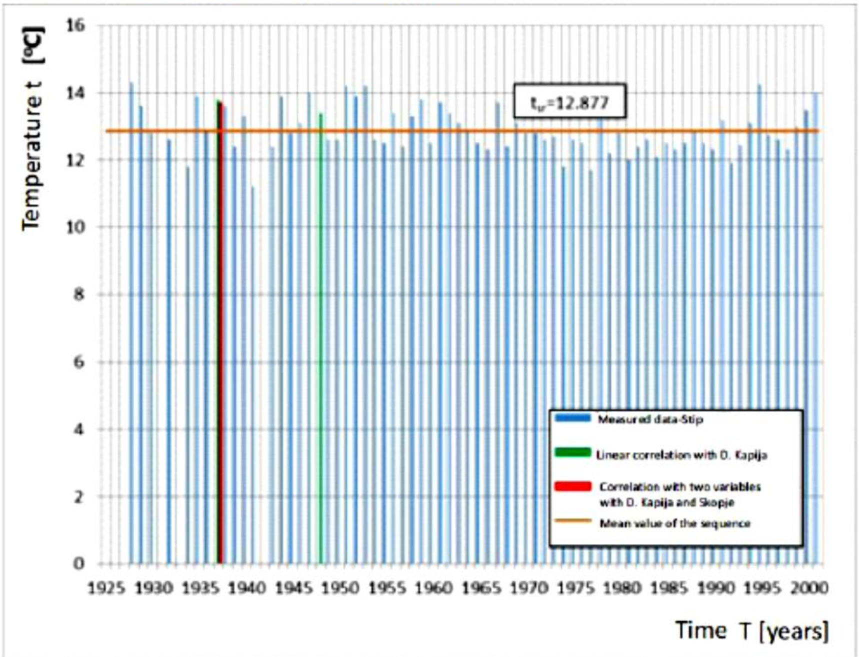
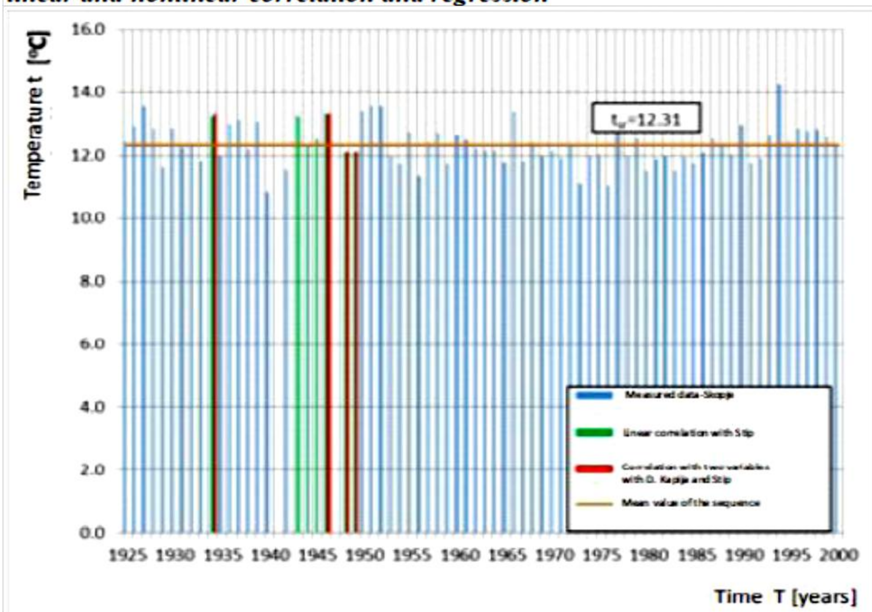


Figure 5 - Average yearly air temperatures in Skopje with results obtained linear and nonlinear correlation and regression



4. RESULT ANALYSIS

- ✓ Homogeneity analysis of the sequences of the annual temperature sums for the period $1925-2000$ for the measuring stations Prilep, Bitola, Demir Kapija and Stip, using different statistical tests (normalized z-test, Student's test and Fisher's test) is made. These parametric testes show that certain sequences, using certain test can be accepted, while using another test can not be accepted, with significance level of 5%. In hydrology, usually it is used the Fisher's test and his properties meets the needs and level of accuracy in this scientific field, so here only this test is presented. According to this test the sequences of the annual temperature sums for the period $1925-2000$ for the measuring stations Prilep, Bitola, Demir Kapija and Stip, can be accepted that these sequences are homogeneous with significance level of 5%.
- ✓ The choice of the measuring stations using to explore the correlation and regression connections is made according to smallest distance of the measuring stations.
- ✓ In establishing the correlative connection between two variables it is taken care of the values of the coefficient of correlation for dependent variable (Y) and independent variable $(X)-(r_{xy})$.
- ✓ Fulfilling the sequence of measured data for the measuring station Prilep is made through a correlation connection with one variable (using data from measuring station Demir Kapija) and correlative connection with two variables (using data from measuring stations Stip and Demir Kapija). The linear and polynomial regression between data sequences from Prilep and Demir Kapija is given. Obtained values by correlation with one variable and with two variables, as well as with regression, are very close (minimal difference in values). Also, we have very close values of the coefficients of regression in polynomial regression between sequences from Demir Kapija and Prilep ($r^2=0,848$) and linear regression between the same sequences ($r^2=0,847$).
- ✓ Fulfilling the sequence of measured data for the measuring station Bitola is made through a correlation connection with one variable (using data from measuring station Demir Kapija) and correlative connection with two variables (using data from measuring stations Demir Kapija and Stip). The obtained values with correlation with one or two variables, as well as the values with regression, are very close (there minimal differences in obtained values, except for 1940, where the temperature varies from $9,9°C$ (linear regression) to $10,5°C$ (correlation with one city). Also, we have very close values of the coefficients of regression in polynomial regression between sequences from Stip and Bitola ($r^2=0,878$) and linear regression between the same sequences ($r^2=0,877$).

- ✓ Fulfilling the sequence of measured data for the measuring station Demir Kapija is made through a correlation connection with one variable (using data from measuring station Stip) and correlative connection with two variables (using data from measuring stations Stip and Skopje). The linear and polynomial regression between the sequences data from the measuring station Demir Kapija and Stip is given. It can be noticed that obtained values with correlation with one or two variables, as well as with regression, are very close. We have very close values of the coefficients of regression in polynomial regression between sequences from Stip and Demir Kapija ($r^2 = 0,95$) and linear regression between the same sequences ($r^2 = 0,947$).
- ✓ Fulfilling the sequence of measured data for the measuring station Stip is made through a correlation connection with one variable (using data from measuring station Demir Kapija) and correlative connection with two variables (using data from measuring stations Demir Kapija and Skopje). The linear and polynomial regression between the sequences data from the measuring station Stip and Demir Kapija is given. It can be noticed that obtained values with correlation with one or two variables, as well as with regression, are very close. We have same values of the coefficients of regression in polynomial regression between sequences from Stip and Demir Kapija ($r^2 = 0,9471$) and linear regression between the same sequences ($r^2 = 0,9471$).
- ✓ Fulfilling the sequence of measured data for the measuring station Skopje is made through a correlation connection with one variable (using data from measuring station Stip) and correlative connection with two variables (using data from measuring stations Stip and Demir Kapija). The linear and logarithmic regression between the sequences data from the measuring station Skopje and Stip is given. It can be noticed that obtained values with correlation with one or two variables, as well as with regression, are very close. We have similar values of the coefficients of regression in linear regression between sequences from Stip and Skopje ($r^2 = 0,865$) and logarithmic regression between the same sequences ($r^2 = 0,864$).
- ✓ In establishing the correlative connection with one variable the biggest coefficient of correlation has the connection between the sequences of data from the measuring station Demir Kapija and measuring station Stip and vice versa ($r^2_{x_1 y} = 0,917$).
- ✓ In establishing the correlation connection with one variable the smallest coefficient of correlation has the connection between the sequences of data from the measuring station Bitola and measuring station Prilep and vice versa ($r^2_{x_1 y} = 0,691$).

- ✓ In establishing correlative connection with two variables, the biggest coefficient of correlation have the connections of Prilep with Stip and Demir Kapija ($r^2_{x_1x_2} = 0,917$), Bitola with Stip and Demir Kapija ($r^2_{x_1x_2} = 0,917$) and Skopje with Demir Kapija and Stip ($r^2_{x_1x_2} = 0,917$).

Correlative connections with two variables, the least coefficient of correlation have the connections of Stip with Prilep and Bitola ($r^2_{x_1x_2} = 0,691$), Skopje with Prilep and Bitola ($r^2_{x_1x_2} = 0,691$) and Demir Kapija with Prilep and Bitola ($r^2_{x_1x_2} = 0,691$).

5. CONCLUSIONS

In a hydrological point of view, the measured data as incoming parameters are very important for qualitative hydrological analysis. We must be careful with the type and quality of the sequence used into forming the relation. Also, it is very important to be careful about the type and strength of the established connection. Having coefficient of correlation less than $0,75$, we can state that connection is very weak or we state that there is no connection. When we have coefficient of correlation above $0,75$ of the correlation, we can state that the connection is good. Hence they can be used for fulfilling the sequences.

The application of the correlation and regression in hydrology is important for analysis of measured data examination of their applicability and continuation of certain sequences of data. The most used statistical techniques in hydrology are correlation and regression analysis. These techniques are complement each other, but there are significantly different. When using correlation, it does not matter whether certain phenomena is dependent or independent, the result is the same. Using correlation between three or more variables, one of the variables must be specified as dependent in advance and other must be specified as independent variables. The main goal of the correlation is to check and quantify whether exists consent between the variables (observed variables). In the regression analysis, we need to have information in advance about the (in)dependence of the variables. The main goal of the regression is to determine the type of the connection, i.e. dependence between observed phenomena.

The correlation and regression are methods for describing the mutual relation between two or more variables. Using these methods in any scientific field impose the following conditions: (1) Correlation and regression lines give linear and nonlinear connection, (2) Correlation and regression lines obtained by the least square method are susceptible on other influences. During the calculations always should be taken into consideration all possible parameters influencing the final result, (3) Correlation between two variables can be better understand if we take into consideration other variables. Lurking variables can make wrong correlation or regression.

References

[1] М. Георгиева, *Случајни грешки,* Математичка школа, Скопје, 1982.
[2] S. E. Edgell, S.M. Noon, *Effect of violation of normality on the* t*–test of the correlation coefficient,* Psychological Bulletin 95 (1984), 576-583.
[3] D. C. Gayou, *Effects of temperature on the mating call of Hyla versicolor,* Copeia (1984), 733-738
[4] Z. A. Ivkovic, *Matematicka statistika,* Naucna knjiga, Beograd, 1982.
[5] С. Јадровски, *Примена на статистички методи во хидрологијата,* Магистерска работа, Градежен факултет, Скопје, 2018.
[6] Р. Малчески, В. *Малческа, Теорија на веројатност,* Армаганка, Скопје, 2019.
[7] J. H. McDonald, K. W. Dunn, *Statistical tests for measures of colocalization in biological microscopy,* Journal of Microscopy 252 (2013), 295-302.
[8] B. H. McCardle, *Lines, models, and errors: Regression in the field,* Limnology and Oceanography 48 (2003), 1363-1366.
[9] D. S. Moore, *The Basic Practice of statistics,* W.H. Freedman and Company, New York, 2010.
[10] I. Pavlic, *Statisticka teorija i primjena,* Tehnicka knjiga, Zagreb, 1977
[11] A. Papoulis, Probability, *Random variables and Stochastic Processes,* International Student Edition, New York, 1965.
[12] Ц. Поповска, В. Ѓешовска, *Хидрологија-теорија со решени задачи,* Градежен факултет-Скопје, 2012.
[13] S. Prohaska, V. Ristić, *Hidrologija kroz teoriju i praksu,* Univerzitet u Beogradu, Rudarsko-geološki fakultet, Beograd, 1996.
[14] S. Prohaska, *Hidrologija I,* Univerzitet u Beogradu, Rudarsko-geološki fakultet, Beograd, 2006.
[15] S. Prohaska, *Hidrologija II,* Univerzitet u Beogradu, Rudarsko-geološki fakultet, Beograd, 2006.
[16] D. Srebrenović, *Primijenjena hidrologija,* Tehnička knjiga, Zagreb, 1986.
[17] *Стратегија за развој и унапредување на хидрометеоролошката служба во Македонија (2009-2011),* Министерство за земјоделство, шумарство и водостопанство на РМ.
[18] N. Kannan, J. P. Keating, R. L. Mason, *A comparison of classical and inverse estimators in the calibration problem,* Communications in Statistics: Theory and Methods 36 (2007), 83-95.
[19] R. G. Krutchkoff, *Classical and inverse regression methods of calibration,* Technometrics 9 (1967), 425-439.

[20] R. G. Krutchkoff, *Classical and inverse regression methods of calibration in extrapolation*, Technometrics 11 (1969), 605-608.
[21] Р. Ѓ. Кукутанов, *Метеорологија,* Универзитет „Гоце Делчев “, Штип, 2014.
[22] I. Ličina, D. Lisjak, *Primjena statističkih metoda s naglaskom na trend, korelaciju i regresiju*, Ekonomski faklutet u Osjeku, 2014.
[23] T. Lwin, J. S. Maritz, *An analysis of the linear-calibration controversy from the perspective of compound estimation*, Technometrics 24 (1982), 235-242.

Faculty of Civil Engineering, Ss. Cyril and Methodius University, Skopje, Macedonia
E-mail address: violetag@gf.ukim.edu.mk
Faculty of Civil Engineering, Ss. Cyril and Methodius University, Skopje, Macedonia
E-mail address: velinov.daniel@gmail.com
BALKAN ENERGY GROUP, Skopje, Macedonia
E-mail address: sjadrovski@hotmail.com

Publisher
Union od Mathematicians of Macedonia – ARMAGANKA

Editor in chief
Prof. d-r Aleksa Malcheski

CIP - Каталогизација во публикација
Национална и универзитетска библиотека "Св. Климент Охридски",
Скопје

51(082)

PROCEEDINGS of CODEMA 2020 = Зборник на трудови од CODEMA
2020. -
Skopje : Armaganka, 2021. - 375 стр. : илустр. ; 25 см

Текст на мак. и англ. јазик. - Фусноти кон текстот. - Библиографија кон
трудовите

ISBN 978-608-4904-09-0

а) Математика -- Зборници

COBISS.MK-ID 53570309