UDK 528:519.2:517.9
Review / Pregledni znanstveni članak

New Formula for Finding the Correlation Coefficient in Geodetic Measurements for a Small Number of Observations

Valeriy GLADILIN – Bila Tserkva[1], Vadym BELENOK – Kyiv[2], Sergiy KRYACHOK – Chernihiv[3], Tetiana SIROSHTAN, Iryna HAMALII – Bila Tserkva[1]

ABSTRACT. When performing geodetic surveys, the number of measurements is usually small, which in the case of dependent series of measurements leads to an underestimation of the correlation coefficient calculated by the standard formula. A new formula for determining the correlation coefficient is proposed. The values of correlation coefficients calculated by the new formula and by the known ones were compared for the number of measurements n = 6. It was found that the correlation coefficient values calculated by the new formula will have higher values as compared to the values calculated by the known formulas, i.e. these values will be less biased with respect to the true values of the correlation coefficients. The correlation coefficients obtained by the new formula will be maximal at an appropriate level of significance and number of degrees of freedom. With a large number of measurements, the values of correlation coefficients obtained by the new formula and by the formulas that are widely used in geodesy and photogrammetry will not differ significantly.

Keywords: correlation coefficient, unbiased and biased estimator of correlation coefficient, calculating the correlation coefficient in geodesy, new method for calculating the correlation coefficient, regression equation.

[1] Assoc. Prof. Valeriy Gladilin, PhD, Department of Geodesy, Cartography and Land Management, Bila Tserkva National Agrarian University, Soborna pl. 8/1, UA-09117 Bila Tserkva, Kyivska oblast, Ukraine, e-mail: vgladilin.55@gmail.com
Assoc. Prof. Tetiana Siroshtan, PhD, Department of Geodesy, Cartography and Land Management, Bila Tserkva National Agrarian University, Soborna pl. 8/1, UA-09117 Bila Tserkva, Kyivska oblast, Ukraine, e-mail: tetiana.nedashkivska@btsau.edu.ua
Assoc. Prof. Iryna Hamalii, PhD, Department of Geodesy, Cartography and Land Management, Bila Tserkva National Agrarian University, Soborna pl. 8/1, UA-09117 Bila Tserkva, Kyivska oblast, Ukraine, e-mail: iryna.gamalii@btsau.edu.ua

[2] Assoc. Prof. Vadym Belenok, PhD, corresponding author, Department of Aerospace Geodesy and Land Management, National Aviation University, Liubomyra Huzara Ave 1, UA-03058 Kyiv, Ukraine, e-mail: belenok.vadim@gmail.com

[3] Assoc. Prof. Sergiy Kryachok, PhD, Department of Geodesy, Cartography and Land Management, Chernihiv Polytechnic National University, Shevchenko Street 95, UA-14035 Chernihiv, Ukraine, e-mail: geodesist2015@gmail.com

1. Introduction

The correlation coefficient is a standard measure of the possible linear relationship between two continuous random variables. Pearson defined the product-moment correlation coefficient (Pearson 1895), which is still used to quantify the relationship between two variables (Hu et al. 2020).

The method of calculating the correlation coefficient and constructing a regression equation has long been used in geodesy and surveying sciences, where it is usually limited to a small number of measurements.

The paper Pokonieczny et al. (2017) uses the correlation coefficient to analyze statistical and spatial relationships between land cover classes and relief (independent variables) and geodetic control points density (a dependent variable). Both correlated independent variables and variables weakly correlated with geodetic control point density were excluded from the studies in the paper.

In Thiagarajan et al. (2021), the classification process of satellite images using a gravity search algorithm based on the correlation coefficient was considered. The correlation coefficient considered in the feature selection process is used to avoid the irrelevant features from the feature set that are used in classification. As a result this procedure increases the classification accuracy of Satellite Images.

The authors Bloßfeld et al. (2018) found that the linear dependence between the estimated geodetic parameters decreased when combining satellites with different altitudes and orbit inclinations based on Satellite Laser Ranging measurements and calculation of the correlation coefficient. This allowed us to make a conclusion about the expediency of using a configuration, which differs from the standard 4-satellite constellation.

In Azuma et al. (2016) methods were developed for precisely analyzing GPS/A data, which makes it possible to determine the motion of the plates at a large distance from the coast. The proposed methods are based on the processing of the observed correlogram. The result is an improved estimation of the transponder location.

The authors Wang et al. (2012) analyzed the correlation between severity levels and spectral data obtained with ASD hand-held spectroradiometers. The results are the basis for studying the feasibility of monitoring the diseases and insect pests in tobacco using hyperspectral remote sensing.

In order to study the geodynamics in Croatia, a geological and geodesic model has been created in two separate ways. Based on the calculated correlation coefficient between these models, it was concluded that both models are reliable (Pribičević et al. 2011).

The theory of the correlation coefficient was applied in the article Tereshchuk et al. (2019) to compare the GPS-derived coordinate values of the rover locations obtained by processing measurements in the OCTAVA and GrafNav/GrafNet software packages. It was found that the correlation between the deviations in the plan and in heights of the differences in coordinate definitions is absent depending on the distance to the base station.

In Zatserkovny et al. (2019), the authors studied the Black Sea level changes using altimetry data. The correlation between sea level anomalies and sea surface temperature of the Black Sea was investigated.

The properties of the correlation coefficient can also be used in GIS. For example, the authors of Cen et al. (2005) applied the correlation coefficient to determine whether gross errors of the geospatial data can be detected in a given GIS model.

Tahouri et al. (2022) investigated a model to identify potential fragile areas prone to erosion in order to diagnose the state of soil degradation. The use of the correlation coefficient made it possible to identify climatic factors that improved the reliability of the model.

Mishra et al. (2022) used GIS and remote sensing data for drought risk assessment and drought analysis of the region. The correlation analysis performed between the vegetation condition index and the standardized precipitation index confirms the reliability of remote sensing data for drought analysis.

In Abdideh (2016), correlation coefficient and linear and nonlinear regression models were used to estimate fracture density in fault zones using petrophysical logging energy.

In the article Issaka and Kumi-Boateng (2020), the authors proved the possibility of using two artificial intelligence methods for an alternative method of quantifying solid earth tides, based on statistical indices and correlation coefficient.

In photogrammetry, the correlation coefficient is the basis for solving the problem of automatic identification of objects arising in the following cases (Dorozhynskyi 2019):

- automatic search in images for objects for which their standard representation is known in advance (e.g., coordinate marks or reference points);
- automatic search for clear contour points on neighboring images (stereo pairs or routes);
- automatic search for correspondent points on the left and right images of a stereo pair during relative orientation of images.

Correlation methods used in photogrammetry belong to the following main groups (Hong and Zhang 2007, Yi et al. 2013):

- area matching;
- feature based matching;
- relation matching.

Digital image correlation is described in Ackermann (1984) based on least squares window matching. The article also notes the high accuracy of the method with respect to the images of charge-coupled device cameras, which are actively used nowadays.

The article Sadeq (2020) notes that image matching is a main task in photogrammetry. It is noted that the area based matching method does not work on a discontinuous edge and in low light areas or with geometric distortion due to changes in the shooting location. The authors propose a new correlation method of detection, based on normalized cross-correlation, in each point, which is devoid of the above drawbacks.

The article Sharma et al. (2018) proposes a new method for solving correspondence problem in stereo-image matching. The method proposed by the authors is based on polynomial fitting and allows increasing of the percentage of accurately matched pixels in stereo pairs compared to the standard approach based on normalized cross-correlation by about 15%.

Area Based Matching has found wide application in the development of terrain survey methods using Unmanned Aerial Vehicles (Gao et al. 2022, Zhang et al. 2022).

At that point, if the camera illumination is different at the two ends of the baseline, then the values of pair correlation coefficients for the correspondent pixels of digital stereo pairs will be underestimated.

It is worth noting that traditional photogrammetry measures pixel values in six "standard zones" of a stereo pair located within a longitudinal overlap (sometimes two points can be defined in each zone, then the total number of measured pixels will be 12) (Dorozhynskyi 2019). With such a small number of measurements ($n = 6$), the value of correlation coefficient calculated by known formulas will be underestimated.

A common disadvantage of popular formulas for determining the correlation coefficient, which are used in geodesy and photogrammetry, is the underestimation of the correlation coefficient calculated with their help for a small number of observations. The resulting underestimation is due to the presence of bias with respect to the true values of the correlation coefficients.

Thus, the correlation coefficient is widely used in geodesy and photogrammetry, and there are a number of tasks where you need to improve the methodology for calculating the correlation coefficient.

2. Materials and Methods

2.1. Known methods of calculating the correlation coefficient

It is known that with a small number of measurements in the case of dependent series of measurements, the correlation coefficient calculated by the common formula (Wasserman 2004):

$$r_0 = \frac{\sum_{i=1}^{n} \delta(x_i)\delta(y_i)}{(n-1)\sigma(x)\sigma(y)} \tag{1}$$

it will be underestimated, and the standard and often used estimate of the Pearson correlation coefficient is slightly biased (Jurková et al. 2022).

In formula (1) the deviations $\delta(x_i)$ and $\delta(y_i)$ are found from expressions:

$$\begin{aligned} \delta(x_i) &= x_i - \bar{x}, \\ \delta(y_i) &= y_i - \bar{y}, \end{aligned} \tag{2}$$

where x_i, y_i are measurements in series X and Y;

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ are arithmetic mean values for these series of measurements;

n – number of measurements in series X and Y.

Empirical standards or root mean square errors of measurement results for series X and Y are determined by the known formulas (Rose 1991, Congalton and Green 2019):

$$\begin{aligned} \sigma(x) &= \sqrt{\frac{\sum_{i=1}^{n} \delta^2(x_i)}{n-1}}, \\ \sigma(y) &= \sqrt{\frac{\sum_{i=1}^{n} \delta^2(y_i)}{n-1}}. \end{aligned} \tag{3}$$

There is a formula by which you can find more accurate correlation coefficient values (Viduev and Kondra 1969):

$$r' = r_0 \left[1 + \frac{1 - r_0^2}{2(n-3)} \right], \tag{4}$$

but even this formula, as revealed when the number of measurements $n > 5$, gives a somewhat underestimated value of the correlation coefficient.

2.2. Theory of the proposed method for calculating the correlation coefficient

First, let's focus on the direct regression equations for the two dependent series of measurements X and Y, these are (Wasserman 2004):

$$\begin{cases} y = b_1 + k_1 \cdot x, \\ x = b_2 + k_2 \cdot y. \end{cases} \tag{5}$$

In these equations, the quantities b_1 and b_2 are represented by the expressions

$$\begin{cases} b_1 = \overline{y} - k_1 \cdot \overline{x}, \\ b_2 = \overline{x} - k_2 \cdot \overline{y}. \end{cases} \tag{6}$$

Regression coefficients or angular coefficients k_1 and k_2 we find by the formulas:

$$k_1 = \frac{\sum_{i=1}^{n} \delta(x_i)\delta(y_i)}{\sum_{i=1}^{n} \delta^2(x_i)}, \quad k_2 = \frac{\sum_{i=1}^{n} \delta(x_i)\delta(y_i)}{\sum_{i=1}^{n} \delta^2(y_i)}. \tag{7}$$

The formula (1) for determining the correlation coefficient can be presented taking into account (7) in the following form:

$$|r_0| = \sqrt{k_1 \cdot k_2}. \tag{8}$$

If the regression equations (5) coincide, i.e., the angle between these lines $\varphi = 0$, the correlation coefficient will be 1, $r_0 = 1$. As the angle increases, the correlation decreases, and when $\varphi = \pi/2$, the correlation equals zero, i.e. $r_0 = 0$. Thus, the correlation coefficient will depend on the angle between direct regression equations (5), such dependence is proposed to be represented as:

$$r = 1 - 2 \cdot \frac{\varphi}{\pi}, \tag{9}$$

where φ is the numerical value of the angle in radians.

To find the angle φ, we use the formula for the angle between two lines, for which we find the second equation (5) with respect to y, and then this system will take the form

$$\begin{cases} y = b_1 + k_1 \cdot x, \\ y = -b_2/k_2 + (1/k_2) \cdot x. \end{cases} \tag{10}$$

Then the angle between the two lines is defined as:

$$\tan\varphi = \frac{1/k_2 - k_1}{1 + k_1/k_2} = \frac{1 - k_1 \cdot k_2}{k_1 + k_2}. \tag{11}$$

Given (11) and (9), the formula for the correlation coefficient will look like:

$$r = 1 - \frac{2}{\pi} \tan^{-1} \frac{1 - k_1 \cdot k_2}{k_1 + k_2}. \tag{12}$$

The obtained formula is used to calculate the correlation coefficient with positive correlation. In expression (7), the coefficients k_1 and k_2 can only be with the same signs, since their expressions have the same numerator, and the denominators are always positive values. The product of the coefficients k_1 and k_2 according to (8) and the fact that $|r_0| \leq 1$ will be:

$$0 \leq k_1 \cdot k_2 \leq 1. \tag{13}$$

Therefore, the numerator in (11) will be positive, i.e.

$$1 - k_1 \cdot k_2 \geq 0. \tag{14}$$

Accordingly, the sign of the fraction in (11) will depend on the sign of the denominator, which in turn depends on the signs of the coefficients k_1 and k_2.

In the case of a negative correlation, when an increase in one quantity corresponds to a decrease in the other, the signs of k_1 and k_2 will be negative, then the angle φ will be negative, and the formula for the correlation coefficient will look like:

$$r = -1 - \frac{2}{\pi} \tan^{-1} \frac{1 - k_1 \cdot k_2}{k_1 + k_2} . \tag{15}$$

3. Results

3.1. Calculation of correlation coefficients by known methods in geodetic constructions

Consider examples of calculating correlation coefficients first with a positive connection. At the triangulation point the directions were measured by twelve methods according to the method of circular methods, the deviations from the mean values of the directions are presented in Table 1.

Table 1. *Results of angle measurements by the method of rounds.*

Sets	$\delta(x_i)$	$\delta(y_i)$	**Sets**	$\delta(x_i)$	$\delta(y_i)$
1	–2.18	+0.57	**7**	–0.48	–1.13
2	–2.38	–2.03	**8**	–0.98	–1.03
3	+0.62	+1.17	**9**	+2.32	+0.77
4	–0.38	+0.07	**10**	+0.42	–0.23
5	+0.82	+2.07	**11**	+2.32	+1.97
6	+0.02	–1.63	**12**	–0.08	–0.53

According to Table 1: $\sum_{i=1}^{n} \delta(x_i)\delta(y_i) = 13.807$; $\sum_{i=1}^{n} \delta^2(x_i) = 23.757$; $\sum_{i=1}^{n} \delta^2(y_i) = 19.907$; $k_1 = 0.5812$; $k_2 = 0.6936$.

Correlation coefficient calculated by formula (1) or (8): $r_0 = 0.635$, by formula (4): $r' = 0.656$, by formula (12): $r = 0.721$.

This example shows that the correlation coefficient calculated by formula (4) is greater by the value of 0.021, and by formula (12) is greater by the value of 0.086 in comparison with its calculated value by formula (1) or (8).

Let us consider a negative correlation relationship. For two series of measurements at $n=5$, the following values were obtained: $\sum_{i=1}^{n}\delta(x_i)\delta(y_i)=-0.41$; $\sum_{i=1}^{n}\delta^2(x_i)=0.72$; $\sum_{i=1}^{n}\delta^2(y_i)=0.84$; $k_1=-0.5694$; $k_2=-0.4881$.

The correlation coefficient in this case we calculate by the formula:

$$r_0=\frac{\sum_{i=1}^{n}\delta(x_i)\delta(y_i)}{\sqrt{\sum_{i=1}^{n}\delta^2(x_i)\sum_{i=1}^{n}\delta^2(y_i)}}=-0.527\,. \tag{16}$$

By formula (4) correlation coefficient $r'=-0.622$ and by formula (15) $r=-0.619$, then at the number of measurements $n=5$ the correlation coefficients calculated by the formulas (4) and (12) are almost the same, but it should be noted, that in the correlation relation study the number of measurements is always more than five, therefore it is suggested to calculate the correlation coefficient by the formula (12), which value will be plausible in comparison with the values found by the formula (1), (8), (16) and by the formula (4) at the number of measurements more than five.

3.2. Study of the proposed formula for calculating the correlation coefficient

Let us investigate the proposed new formula for calculating the correlation coefficient. First, let us show that the number of measurements n in the series, at which the correlation coefficients calculated by formulas (4) and (12), are approximately the same and fluctuate around five. For this purpose, let us present formula (12) in a different form. The formula (8) shows that the correlation coefficient is equal to the geometric mean of the coefficients k_1 and k_2. If $(k_1+k_2)/2$ is the arithmetic mean, $k_1=k_2=r_0$, which follows from (8), then

$$k_1+k_2=2r_0\,. \tag{17}$$

Given (17) expression (12) will be:

$$r=1-\frac{2}{\pi}\tan^{-1}\frac{1-r_0^2}{2r_0}\,. \tag{18}$$

For the above example with positive correlation $r_0 = 0.635$, let's substitute this value in (4), we get $r = 0.720$. According to formula (12) $r = 0.721$, we can see, that these obtained values are practically equal, and the difference between coefficients $k_1 - k_2 = 0.69 - 0.58 = 0.11$ has a rather significant value.

To find the number of measurements n at which the correlation coefficients calculated by formulas (4) and (18) will be equal, for this purpose let's equate the right parts of equations (4) and (18), we obtain:

$$r_0 \cdot \left[1 + \frac{1-r_0^2}{2(n-3)}\right] = 1 - \frac{2}{\pi}\tan^{-1}\frac{1-r_0^2}{2r_0}. \tag{19}$$

From expression (19) we find:

$$n = \frac{r_0 \cdot \left(1-r_0^2\right)}{2 \cdot \left(1 - r_o - \frac{2}{\pi}\tan^{-1}\frac{1-r_0^2}{2r_0}\right)} + 3. \tag{20}$$

Setting different values of the correlation coefficient r_0 let's determine the value of n by the expression (20), they are given in Table 2.

Table 2. *Determining the number of measurements in a row.*

***r*₀**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.99**
n	4.8	4.8	4.9	4.9	5.0	5.1	5.2	5.3	5.4	5.6	5.7

Table 2 shows that the equality of correlation coefficients found by formulas (4) and (12) or (18) corresponds when the number of measurements is about five.

In order to compare the values of the coefficients obtained by different formulas, it is first necessary to obtain formulas, by which we can find the coefficients k_1 and k_2 depending on the given values of ρ theoretical value of the correlation coefficient.

With equal exact measurements in two series *x* and *y* the maximum admissible ratio of empirical variances is determined by the inequality (Koch 2015, Suyunov et al. 2021):

$$\frac{\sigma^2(x)}{\sigma^2(y)} \le F_q, \ \left(\sigma^2(x) \ge \sigma^2(y)\right), \tag{21}$$

where F_q is the variance ratio subject to F-distribution (also called Fisher-Snedecor distribution) and chosen from the table, depending on the significance level q and the number of degrees of freedom $(n-1)$ for two series of measurements X and Y.

Given formulas (3), (7), and (21), the variance ratio will be:

$$\frac{\sigma^2(x)}{\sigma^2(y)} = \frac{\sum_{i=1}^{n} \delta^2(x_i)}{\sum_{i=1}^{n} \delta^2(y_i)} = \frac{k_2}{k_1} = F_q . \qquad (22)$$

Consider the case when the ratio of coefficients k_2 and k_1 will be maximum, then the values of correlation coefficients determined by formulas (18) and (12) will have the largest difference. Let us make a system of equations from (22) and (8), from the solution of which we find the coefficients of k_1 and k_2.

$$\frac{k_2}{k_1} = F_q, \ k_1 \cdot k_2 = r_0^2, \qquad (23)$$

where from

$$k_1 = \frac{r_0}{\sqrt{F_q}}, \qquad k_2 = r_0 \sqrt{F_q}. \qquad (24)$$

For the significance level $q = 0.10$ and the number of degrees of freedom $(n-1) = 6-1 = 5$, we find according to Table 3 of the Appendix in Mendenhall and Sincich (2013) dispersion relation $F_q = 3.5$, at such initial data the difference in the correlation coefficient values that are found by formulas (18) and (12) will be the largest, i.e. we consider the worst case.

Setting different values of the theoretical Pearson correlation coefficient ρ, we calculate the values k_1 and k_2 by formula (24), coefficient r_0 by formula (8), r' by formula (4), r by formula (18) and formula (12). All obtained values are given in Table 3.

Table 3. *Values of correlation coefficients calculated using different formulas*.

ρ	**Coefficients**		**Value of correlation coefficients**			
	k_1	k_2	**by (8)**	**by (4)**	**by (18)**	**by (12)**
0	0	0	0	0	0	0
0.1	0.05345	0.18708	0.100	0.116	0.127	0.152
0.2	0.10690	0.37417	0.200	0.232	0.251	0.296
0.3	0.16036	0.56125	0.300	0.346	0.371	0.427
0.4	0.21381	0.74833	0.400	0.456	0.484	0.543
0.5	0.26726	0.95541	0.500	0.562	0.590	0.645
0.6	0.32071	1.12250	0.600	0.664	0.688	0.734
0.7	0.37417	1.30958	0.700	0.760	0.778	0.813
0.8	0.42762	1.49666	0.800	0.848	0.859	0.882
0.9	0.48107	1.68375	0.900	0.928	0.933	0.944
1.0	0.53452	1.87083	1.000	1.000	1.000	1.000

The correlation coefficients calculated by the formula (18) were at $k_1 = k_2$, yet the values of r that are found by (18) are greater than by the formula (4). The highest values of correlation coefficients reach when found according to formula (12).

Let us evaluate the reliability of the correlation coefficient values calculated according to different formulas. For this purpose we will use Fisher's criterion (Fisher 1921, Mudholkar and Chaubey 1976):

$$z = f(r) = \frac{1}{2}\left[\ln\left(1+|r|\right) - \ln\left(1-|r|\right)\right]. \tag{25}$$

This function obeys the law of normal distribution.

The standard error of z is calculated by the formula (Wasserman 2004):

$$\sigma_z = \frac{1}{\sqrt{n-3}}. \tag{26}$$

With probability 0.90 ($t = 1.645$) the value of z can take values:

$$z - t \cdot \sigma_z \le z \le z + t \cdot \sigma_z. \tag{27}$$

The boundary values of z correspond to the boundary values of the correlation coefficient, which are calculated by the formula (25) inverse to the formula (25) (Wasserman 2004):

$$r = f^{-1}(z) = \frac{e^{2z}-1}{e^{2z}+1}. \tag{28}$$

The calculations performed by formulas (25)–(28) for the correlation coefficient values calculated by formulas (8), (4), (18), (12) are shown in Table 4 and Table 5, respectively.

Table 4. *Estimation of correlation coefficients (CC) calculated by formulas (8) and (4).*

CC value formula (8)	z value	Confidence Interval				CC value formula (4)	z value	Confidence Interval			
		z_{left}	z_{right}	r_{left}	r_{right}			z_{left}	z_{right}	r_{left}	r_{right}
0.000	0.000	–0.582	0.582	–0.524	0.524	**0.000**	0.000	–0.582	0.582	–0.524	0.524
0.100	0.100	–0.481	0.682	–0.447	0.593	**0.116**	0.117	–0.465	0.698	–0.434	0.603
0.200	0.203	–0.379	0.784	–0.362	0.655	**0.232**	0.236	–0.345	0.818	–0.332	0.674
0.300	0.310	–0.272	0.891	–0.266	0.712	**0.346**	0.361	–0.221	0.942	–0.217	0.736
0.400	0.424	–0.158	1.005	–0.157	0.764	**0.456**	0.492	–0.089	1.074	–0.089	0.791
0.500	0.549	–0.032	1.131	–0.032	0.811	**0.562**	0.636	0.054	1.217	0.054	0.839
0.600	0.693	0.112	1.275	0.111	0.855	**0.664**	0.800	0.218	1.382	0.215	0.881
0.700	0.867	0.286	1.449	0.278	0.895	**0.760**	0.996	0.415	1.578	0.392	0.918
0.800	1.099	0.517	1.680	0.475	0.933	**0.848**	1.249	0.667	1.831	0.583	0.950
0.900	1.472	0.891	2.054	0.712	0.968	**0.928**	1.644	1.062	2.225	0.787	0.977
0.999	3.800	3.219	4.382	0.997	1.000	**0.999**	3.800	3.219	4.382	0.997	1.000

Table 5. *Estimation of correlation coefficients (CC) calculated by formulas (18) and (12).*

CC value formula (18)	z value	Confidence Interval				CC value formula (12)	z value	Confidence Interval			
		z_{left}	z_{right}	r_{left}	r_{right}			z_{left}	z_{right}	r_{left}	r_{right}
0.000	0.000	–0.582	0.582	–0.524	0.524	**0.000**	0.000	–0.582	0.582	–0.524	0.524
0.127	0.128	–0.454	0.709	–0.425	0.610	**0.152**	0.153	–0.428	0.735	–0.404	0.626
0.251	0.256	–0.325	0.838	–0.314	0.685	**0.296**	0.305	–0.276	0.887	–0.270	0.710
0.371	0.390	–0.192	0.971	–0.190	0.749	**0.427**	0.456	–0.125	1.038	–0.125	0.777
0.484	0.528	–0.053	1.110	–0.053	0.804	**0.543**	0.608	0.027	1.190	0.027	0.831
0.590	0.678	0.096	1.259	0.096	0.851	**0.645**	0.767	0.185	1.348	0.183	0.874
0.688	0.844	0.263	1.426	0.257	0.891	**0.734**	0.937	0.356	1.519	0.341	0.909
0.778	1.040	0.459	1.622	0.429	0.925	**0.813**	1.136	0.554	1.717	0.504	0.938
0.859	1.290	0.708	1.871	0.609	0.954	**0.882**	1.385	0.803	1.966	0.666	0.962
0.933	1.681	1.099	2.263	0.800	0.979	**0.944**	1.774	1.192	2.355	0.831	0.982
0.999	3.800	3.219	4.382	0.997	1.000	**0.999**	3.800	3.219	4.382	0.997	1.000

Table 4 and Table 5 show that all calculated values of correlation coefficients lie within the confidence intervals. Thus, the obtained values of correlation coefficients are reliable.

4. Conclusion

Thus, as can be seen from Table 3 the estimate of the correlation coefficient starting from $n = 6$, which is performed by formula (4), will be shifted, so a more accurate estimate can be obtained by formula (12), formula (18) is not taken into account because it is obtained for the ideal case when $k_1 = k_2 = r_0$, this situation is rare. The values of r found by formula (18) will be less than those found by formula (12). In essence, the coefficients obtained by formula (12) will be maximum at the appropriate level of significance and the number of degrees of freedom, obtained on the basis that the level of significance $q > 0.25$ and a large number of measurements, $Fq \to 1$.

Calculation of more exact values of correlation coefficients (normalized correlation moments) is important in digital photogrammetry for mutual orientation and geodetic referencing of a stereo pair of images, as well as for investigation of systematic errors in measurement series. Underestimated values of correlation coefficients will influence the equation of correlation function and in the final case will give distorted representation about the values of systematic errors.

References

Abdideh, M. (2016): Estimation of the fracture density in reservoir rock using regression analysis of the petrophysical data, Geodesy and Cartography, 42 (3), 85–91.

Ackermann, F. (1984): Digital image correlation: performance and potential application in photogrammetry, The Photogrammetric Record, 11 (64), 429–439.

Azuma, R., Tomita, F., Iinuma, T., Kido, M., Hino, R. (2016): Development and examination of new algorithms of traveltime detection in GPS/acoustic geodetic data for precise and automated analysis, Earth, Planets and Space, 68 (1), 143.

Bloßfeld, M., Rudenko, S., Kehm, A., Panafidina, N., Müller, H., Angermann, D., Hugentobler, U., Seitz, M. (2018): Consistent estimation of geodetic parameters from SLR satellite constellation measurements, Journal of Geodesy, 92 (9), 1003–1021.

Cen, M., Gu, L., Yang, X., Zhou, G. (2005): The relationship between judgement matrix and correlation coefficient in the GIS data error separability analysis, American Society for Photogrammetry and Remote Sensing – Annual Conference 2005 – Geospatial Goes Global: From Your Neighborhood to the Whole Planet, 1, 69–73.

Congalton, R. G., Green, K. (2019): Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Third edition, CRC Press.

Dorozhynskyi, O. L. (2019): Photogrammetry and Remote Sensing, Book 1, Textbook, Lviv Polytechnic Publishing House, Lviv (in Ukrainian).

Fisher, R. A. (1921): On the probable error of a coefficient of correlation deduced from a small sample, Metron, 1, 3–32.

Gao, S., Yuan, X.-P., Gan, S., Hu, L., Bi, R., Li, R.-B., Luo, W.-D. (2022): UAV Image Matching Method Integrating SIFT Algorithm and Detection Model Optimization, Spectroscopy And Spectral Analysis, 42 (5), 1497–1503.

Hong, G., Zhang, Y. (2007): Combination of feature-based and area-based image registration technique for high resolution remote sensing image, International Geoscience and Remote Sensing Symposium (IGARSS), 4422809, 377–380.

Issaka, Y., Kumi-Boateng, B. (2020): Artificial intelligence techniques for predicting tidal effects based on geographic locations in Ghana, Geodesy and Cartography (Vilnius), 46 (1), 1–7.

Jurková, V., Žežula, I., Klein, D., Hutník, O. (2022): Unbiased estimator of correlation coefficient, Communications in Statistics – Theory and Methods, 51 (1), 95–115.

Koch, K.-R. (2015): Minimal detectable outliers as measures of reliability, Journal of Geodesy, 89 (5), 483–490.

Mendenhall, W., Sincich, T. (2013): A Second Course in Statistics: Regression Analysis, 7th Edition, Pearson New International Edition.

Mishra, D., Goswami, S., Matin, S., Sarup, J. (2022): Analyzing the extent of drought in the Rajasthan state of India using vegetation condition index and standardized precipitation index, Modeling Earth Systems and Environment, 8 (1), 601–610.

Mudholkar, G. S., Chaubey, Y. P. (1976): On the Distribution of Fisher's Transformation of the Correlation Coefficient, Communications in Statistics – Simulation and Computation, 5 (4), 163–172.

Pearson, K. (1895): Royal Society Proceedings, 58, 241.

Pokonieczny, K., Bielecka, E., Kamiński, P. (2017): Analysis of geodetic control points density depending on the land cover and relief – the Opoczno district case study, 10th International Conference on Environmental Engineering, ICEE.

Pribičević, B., Đapo, A., Medak, D. (2011): Geodetic-Geologic Research on Wider Zagreb Area based on Geodynamic Network of City of Zagreb (Geodetsko-geološka istraživanja na širem zagrebačkom području oslonjena na Geodinamičku mrežu Grada Zagreba), Geodetski List, 65 (1), 1–19.

Rose, Ch. M. (1991): Error Theory as Applied to Mapping, Charting, and Geodesy, DMA Technical Report, Defense Mapping Agency, Springfield.

Sadeq, H. (2020): Using total probability in image template matching, International Arab Journal of Information Technology, 17 (3), 347–357.

Sharma, L., Sharma, J. K., Anand, D., Sharma, S. (2018): An Adaptive Window Based Polynomial Fitting Approach for Pixel Matching in Stereo Images, Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT, 8473146, 657–661.

Suyunov, A., Suyunov, S., Aminjanova, M., Rakhmatullaeva, K. (2021): Improvement of the method for comparing subsidence of structures using the Fischer's F-test and the Foster-Stuart test, E3S Web of Conferences, 227, 04005.

Tahouri, J., Sadiki, A., Karrat, L., Johnson, V. C., Chan, N. w., Fei, Z., Kung, H. T. (2022): Using a modified PAP/RAC model and GIS-for mapping water erosion and causal risk factors: Case study of the Asfalou watershed, Morocco, International Soil and Water Conservation Research, 10 (2), 254–272.

Tereshchuk, O. I., Korniyenko, I. V., Kryachok, S. D., Malik, T. M., Belenok, V. Y., Skorintseva, I. B. (2019): Research of systematic errors according to the results of processing satellite observations by software complexes, News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences, 4 (436), 199–212.

Thiagarajan, K., Anandan, M. M., Stateczny, A., Divakarachari, P. B., Lingappa, H. K. (2021): Satellite image classification using a hierarchical ensemble learning and correlation coefficient-based gravitational search algorithm, Remote Sensing, 13 (21), 4351.

Viduev, N. G., Kondra, S. G. (1969): Probabilistic-statistical analysis of measurement errors, Textbook, Nedra, Moscow (in Russian).

Wang, M., Li, X.-J., Yao, Q.-Q., Liu, Y. (2012): Extraction of Diseases and Insect Pests for Tobacco Based on Hyperspectral Remote Sensing, Geodetski List, 66 (3), 209–216.

Wasserman, L. (2004): All of Statistics, A Concise Course in Statistical Inference, Textbook, Springer Texts in Statistics.

Yi, B., Yu, P., He, G.-Z., Chen, J. (2013): A Fast Matching Algorithm with Feature Points Based on NCC, Proceedings of the 2013 International Academic Workshop on Social Science, 955–958.

Zatserkovny, V. I., Babych, V., Belenok, V., Frolov, H. O., Hebryn-Baidy, L. (2019): Black sea level change monitoring using altimetry data and geo-information technologies, 18th International Conference on Goinformatics – Theoretical and Applied Aspects, May 2019, Kiev, Ukraine, European Association of Geoscientists & Engineers, 13–16.

Zhang, Y., Ma, G., Wu, J. (2022): Air-Ground Multi-Source Image Matching Based on High-Precision Reference Image, Remote Sensing, 14 (3), 588.

Nova formula za izračun koeficijenta korelacije u geodetskim mjerenjima za mali broj opažanja

SAŽETAK. Prilikom izvođenja geodetskih izmjera, broj mjerenja je obično mali što u slučaju serije ovisnih mjerenja dovodi do niske procjene koeficijenta korelacije izračunatog pomoću standardne formule. Predlaže se nova formula za određivanje koeficijenta korelacije. Uspoređene su vrijednosti koeficijenta korelacije izračunate pomoću nove formule i pomoću poznatih formula za broj mjerenja n = 6. Utvrđeno je da će vrijednosti koeficijenta korelacije izračunate novom formulom biti veće u usporedbi s vrijednostima izračunatima pomoću poznatih formula, odnosno te vrijednosti će biti manje pristrane u odnosu na prave vrijednosti koeficijenata korelacije. Koeficijenti korelacije dobiveni pomoću nove formule bit će maksimalni na odgovarajućoj razini značaja i broja stupnjeva slobode. Za veliki broj mjerenja, vrijednosti koeficijenata korelacije dobivene novom formulom i formulama čija je upotreba raširena u geodeziji i fotogrametriji neće se značajno razlikovati.

Ključne riječi: *koeficijent korelacije, nepristrani i pristrani procjenitelj koeficijenta korelacije, izračunavanje koeficijenta korelacije u geodeziji, nova metoda izračunavanja koeficijenta korelacije, regresijska jednadžba.*

Received / Primljeno: 2022-06-01

Accepted / Prihvaćeno: 2022-06-23