Multidiszciplináris tudományok, 12. kötet. (2022) 3 sz. pp. 174-185 **https://doi.org/10.35925/j.multi.2022.3.16**

FACTOR ANALYSIS OF MULTI−ELECTRODE GEOELECTRIC FIELD DATA

Brigitta Turai-Vurom
PhD. student, University of Miskolc
Institute of Geophysics and Geoinformation Science, Department of Geophysics
3515 Miskolc, Miskolc-Egyetemváros, e-mail: brigitta.vurom@gmail.com

Norbert Péter Szabó
Full professor, University of Miskolc
Institute of Geophysics and Geoinformation Science, Department of Geophysics
3515 Miskolc, Miskolc-Egyetemváros, e-mail: gfnmail@uni-miskolc.hu

Abstract

The paper presents the results of the factor analysis of multi-electrode geoelectric data measured in the area of the Surány aquifer. The 9-variable dataset measured on the first profile was examined by factor analysis using 2, 3, 4 and 5 factors. The calculations show that the Factor profile 1 carries the effects of IP (Induced Polarization), while the Factor profile 2 shows the effect of apparent resistivity (Rho) with the greatest weight. In the Factor profile 3, the factor weight of the current (In) is the largest. Based on all this, it can be concluded that in multi-electrode geoelectric measurements, the first factor depends on the IP effect explaining the most of the variance of the observed data, the second factor on the apparent resistivity, and the third factor on the current. The obtained results confirm that the factor analysis is suitable for the correction and replenishment of noisy and erroneous data of multi-electrode measurements.

Keywords: *multi-electrode measurements, geoelectric data, factor analysis, factor profiles, data replenishment*

1. Introduction

Multi-electrode measurements have been widely used in both geological exploration and environmental studies (Loke, 2000). This measurement technique significantly increases the efficiency of geoelectric measurements (Kearey et al., 2002) and reduces the time of measurements. Measurement control programs and scripts can be entered in the memory of computer-controlled multi-electrode instruments. In the case of geoelectric measurements, the electric current is introduced into the ground between two current electrodes (A, B) and the potential difference between the two potential electrodes (M, N) and the parameters of the induced polarization (IP) are measured. The mutual spatial position of the four electrodes (A, B, M, N) determines the reference point (penetration depth) of the measurement. If many electrodes are sunk into the ground along a straight line (section), all electrodes can be used for current input and potential difference measurement as well. The measurement control scripts contain the addresses of the current electrodes (A, B) and the potential difference measuring electrodes (M, N) sequentially. Thus, up to hundreds of measurements can be programmed using different electrode

arrangements (Schlumberger, Wenner, pole-pole, pole-dipole, dipole-pole and dipole-dipole). The spatial distribution of the measured parameters (ρ_a - apparent resistivity, M - apparent chargeability, SP - spontaneous potential) can be determined in a vertical plane fitting the spreading profile (2D measurement) or in a subsurface volume if the electrodes are connected along several parallel profiles (3D measurement). From the spatial distribution of measured parameters, geophysical images (models) describing the geological structure of the earth can be generated by inversion methods (Dobróka et al., 2014). Thus, we can map the subsurface without trial pits or drilling. However, a significant difficulty with multi-electrode measurements is that we only see the measured file for analysis at the end of the entire measurement cycle (which can be up to several hours depending on the size of the control script). The quality of the measured data can be determined only at the end of the measurement. In several cases, our measurement results may contain incorrect data. These erroneous data are usually removed from the measured file by filtering, but in this way we cannot generate a model detail by inversion at the spatial reference point of the filtered data. Thus, the geological information (geological image) that can be obtained from the multi-electrode measurement will be truncated or distorted. This problem can be corrected by replenishment and correcting the filtered data afterwards. In this paper, we investigated the possibility of this replenishment using factor analysis on a real field multi-electrode dataset. Factor analysis is an effective mathematical method for exploring hidden relationships in multivariate datasets. In applied geophysics, this method has already been successfully applied in well log inversion for data replenishment and section repairment (Szabó, 2018; Szabó et al., 2021; Abordán and Szabó, 2020).

2. The measured dataset

The multi-electrode geoelectric dataset No. 1 (Turai and Nádasi, 2020) measured in the area of the Surány aquifer was chosen for the study. The columns of the dataset contain the values of the measurement control and the measured parameters, and the rows contain the values of the spatial measurement points determined by the current (AB) current and (MN) potential electrodes, in the chronological order of the measurement. During the field measurement, the parameters were determined and recorded at 744 reference points. The SYSCAL Pro 72-channel instrument from IRIS was used for the measurement. From the dataset, we selected the parameters measured for the ground, the apparent specific electrical resistivity (Rho, in ohmm), the average apparent chargeability (M, in mV / V), the apparent chargeability for 80 ms (M1, in mV / V), apparent chargeability for 160 ms (M2, in mV / V), apparent chargeability for 320 ms (M3, in mV / V), measurement standard deviation (Dev, %), spontaneous potential between MN electrodes before excitation (Sp, in mV), the potential between MN electrodes during excitation (Vp, mV) and the current between AB electrodes during excitation (In, mA). The data matrix of the factor analysis **D** was formed from the 9 parameters thus selected.

3. Mathematical basis of factor analysis

The birth of factor analysis is linked to the paper of Charles Spearman (1904), who used Karl Pearson's (1901) work on correlation calculations. Since then, numerous books and technical articles related to factor analysis have been published, among which the book written by Cudeck and MacCallum (2007) stands out, which summarizes the most important applications of factor analysis up to that point and its future perspectives. Let us generate the data matrix for the multi-electrode measurement

$$D = \{d_{n,k}\}, \qquad n = 1, 2, \dots, N, \qquad k = 1, 2, \dots, K, \tag{1}$$

where N - the number of measured data points,
K - number of variables (different measured parameters),
$d_{n,k}$ - the value of the k-th measured data (parameter) at the n-th data point, in the time order of the measurement.

Write the data matrix **D** with a factor model (Szabó and Kormos, 2012)

$$\mathbf{D} = \mathbf{F}\mathbf{L}^{\mathrm{T}} + \mathbf{E}, \tag{2}$$

where **F** - $N \times M$ size matrix of factor sections (factors),
M - number of factors,
L - a $K \times M$ matrix of factor weights,
E - the $N \times K$ error matrix (deviation matrix).

The aim of factor analysis is the above decomposition, and the estimation of factor loadings and factor scores to the same points where measurements were made. The reduction of the input data matrix allows to analyse the dataset in a simpler way as well as to explore hidden information in the dataset.

4. Factor analysis of the multi-electrode dataset

Prior to the calculation, the 9-variable dataset presented in Chapter 2 was standardized. For factor analysis, we used a computer code for Matlab system using a maximum likelihood method based function called factoran. For the rotation of factors the varimax algorithm was used (Kaiser, 1958). After reading in the dataset, the self-developed program displays the values of the covariance matrix on the screen, calculates the factor model and factor weights of the dataset. It also prints the matrix of the factor model and the matrix of factor weights to a Matlab file and a text file (txt). By setting the value of an input variable, the number of factors must be determined. Preliminary calculations were performed for all possible factor models. For the 9-variable dataset this was possible for 1-factor, 2-factor, 3-factor, 4-factor, and 5-factor models. The vectors of the factor weights are shown in the following five tables below for analysis. Table 1 shows the results of the 1-factor, Table 2 the 2-factor, Table 3 the 3-factor, Table 4 the 4-factor, and Table 5 the results of the 5-factor evaluation. It can be seen from the previous tables that the weight of factor 1 is highest for the M2 variable (IP apparent chargeability) in all evaluations. Among these, the maximum occurs in the 5-factor case (0.96). Based on this, it can be concluded that the effect of the variable M2 appears the most in factor 1. It can also be stated that in the factor weight vectors 1, the weights of the variables M, M1 and M3 are also high, above 0.8. Based on all this, it is clear that Factor 1 carries the effects of IP (Induced Polarization). In factor section 2, based on the 5-factor analysis, the effect of resistivity (Rho) appears with the highest weight (0.98). In addition, the factor weight of the variable Vp is relatively high, and in the 3-factor model the weight of this variable is over 0.95 (0.96). Thus, the factor section 2 seems to be sensitive to resistivity, since the voltage Vp also varies as a function of the resistivity of the medium. In factor section 3 the factor weight of the current (In) stands out, in the 4-factor model it is 0.98, and in case of the 5-factor model it is above 0.9. In factor sections 4 and 5 the weight of each variable is less than 0.4, suggesting that none of these variables have a prominent effect on these factors. In the following, the relationship between the standardized normalized variables and the individual factor profiles is shown in the 5-factor case.

Table 1. *Analysis with 1 factor.*

Variable	Factor weight
Rho	-0.2561
M	0.9034
M1	0.8578
M2	0.9788
M3	0.9637
Dev	0.1512
Sp	0.0350
Vp	-0.3683
In	-0.1397

Table 2. *Analysis with 2 factors.*

Variable	1st factor weight	2nd factor weight
Rho	-0.0720	0.6895
M	0.8380	-0.3455
M1	0.8673	-0.0920
M2	0.9631	-0.2143
M3	0.9074	-0.3220
Dev	0.0747	-0.2795
Sp	0.0208	-0.0524
Vp	-0.1532	0.8219
In	-0.0928	0.1825

Table 3. *Analysis with 3 factors.*

Variable	1st factor weight	2nd factor weight	3rd factor weight
Rho	-0.0903	0.5231	-0.3013
M	0.8486	-0.2901	0.1399
M1	0.8879	-0.1511	-0.1258
M2	0.9512	-0.1838	0.1222
M3	0.9106	-0.2107	0.3173
Dev	0.0632	-0.1025	0.3841
Sp	0.0259	-0.0590	-0.0204
Vp	-0.1235	0.9563	-0.2555
In	-0.0838	0.2123	-0.0708

Table 4. *Analysis with 4 factors.*

Variable	1st factor weight	2nd factor weight	3rd factor weight	4th factor weight
Rho	-0.1053	0.6703	-0.1337	-0.2438
M	0.8597	-0.2807	-0.0618	0.1022
M1	0.8886	-0.0992	-0.0714	-0.1382
M2	0.9588	-0.1445	-0.0460	0.0971
M3	0.9260	-0.1774	-0.0218	0.3014
Dev	0.0664	-0.1241	-0.0985	0.3840
Sp	0.0266	-0.0617	-0.0326	-0.0251
Vp	-0.1910	0.7815	0.1267	-0.2521
In	-0.0568	0.0632	0.9752	-0.1916

Table 5. *Analysis with 5 factors.*

Variable	1st factor weight	2nd factor weight	3rd factor weight	4th factor weight	5th factor weight
Rho	-0.1037	0.9799	-0.0789	-0.1327	0.0151
M	0.8620	-0.2051	-0.0408	0.0881	0.3091
M1	0.8905	-0.0472	-0.0670	-0.1107	0.0274
M2	0.9656	-0.1123	-0.0492	0.1029	0.0328
M3	0.9232	-0.1557	-0.0248	0.3010	0.0763
Dev	0.0686	-0.1743	-0.1325	0.3767	-0.0776
Sp	0.0233	-0.0134	-0.0189	-0.0166	0.1399
Vp	-0.2291	0.5598	0.1534	-0.2421	-0.2543
In	-0.0616	0.0085	0.9039	-0.1745	-0.1272

Table 6. *Change in factor loadings between 2 and 5 factor models.*

Variable	Factor 1 loading in the 2-factor model	Factor 2 loading in the 2-factor model	Factor 1 loading in the 5-factor model	Factor 2 loading in the 5-factor model	Change in loading of factor 1	Change in the loading of factor 2	Displacement vector magnitude
Rho	-0.0720	0.6895	-0.1037	0.9799	0.0317	-0.2903	**0.2921**
M	0.8380	-0.3455	0.8620	-0.2050	-0.0240	-0.1404	0.1425
M1	0.8673	-0.0920	0.8905	-0.0472	-0.0232	-0.0448	0.0505
M2	0.9631	-0.2143	0.9656	-0.1123	-0.0025	-0.1020	0.1021
M3	0.9074	-0.3220	0.9232	-0.1557	-0.0158	-0.1663	0.1670
Dev	0.0747	-0.2795	0.0686	-0.1743	0.0060	-0.1052	0.1054
Sp	0.0208	-0.0524	0.0233	-0.0134	-0.0025	-0.0390	**0.0391**
Vp	-0.1532	0.8219	-0.2291	0.5598	0.0759	0.2620	0.2728
In	-0.0928	0.1825	-0.0616	0.0085	-0.0313	0.1740	0.1768

The factor sections and the profiles of the dataset variables were compared by calculating mutual statistical indicators (variance, covariance and correlation). The mutual evolution of the factor section 1 and the variable Rho is shown in Fig. 1, for variable M in Fig. 2, for variable M1 in Fig. 3, for variable M2 in Fig. 4 and for variable M3 in Fig. 5. (In each figure, the blue curve is on top so that where the two curves have the same value, the red curve is obscured.) Comparing the Pearson's correlation values shown in the figures and the factor weights obtained by factor analysis (Table 5), it can be concluded that the values, although not exactly identical in numerical value, follow each other well. The weak opposite correlation between the apparent resistivity (Rho) and the factor section 1 can be clearly seen in Fig. 1, because in some parts of the profile where the factor value is high, the Rho value is small and where the factor loading is small, the Rho value is high. The strongest correlation is found in Fig. 4 between factor section 1 and M2 chargeability. Based on this figure, it can be concluded that the chargeability M2 can be almost completely replaced by factor 1. For the other IP parameters (M, M1, and M3), Figures 2, 3, and 5 also show a close relationship with factor 1. However, the other parameters (Dev, Sp, Vp and In) have a very weak relationship (practically none) with factor 1.

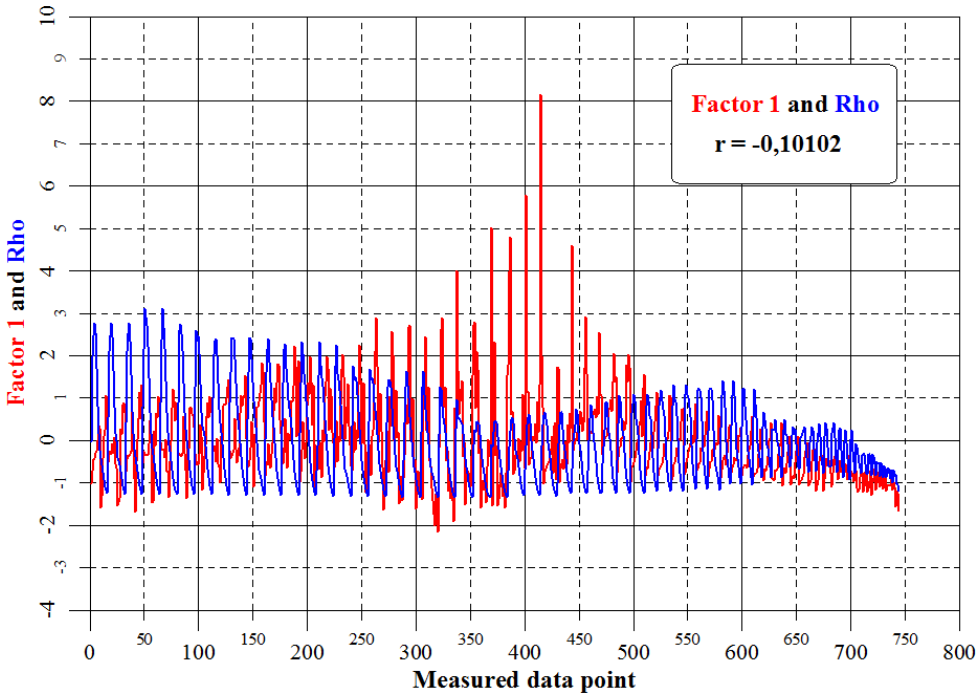

Figure 1. *Factor section 1 and Rho section.*

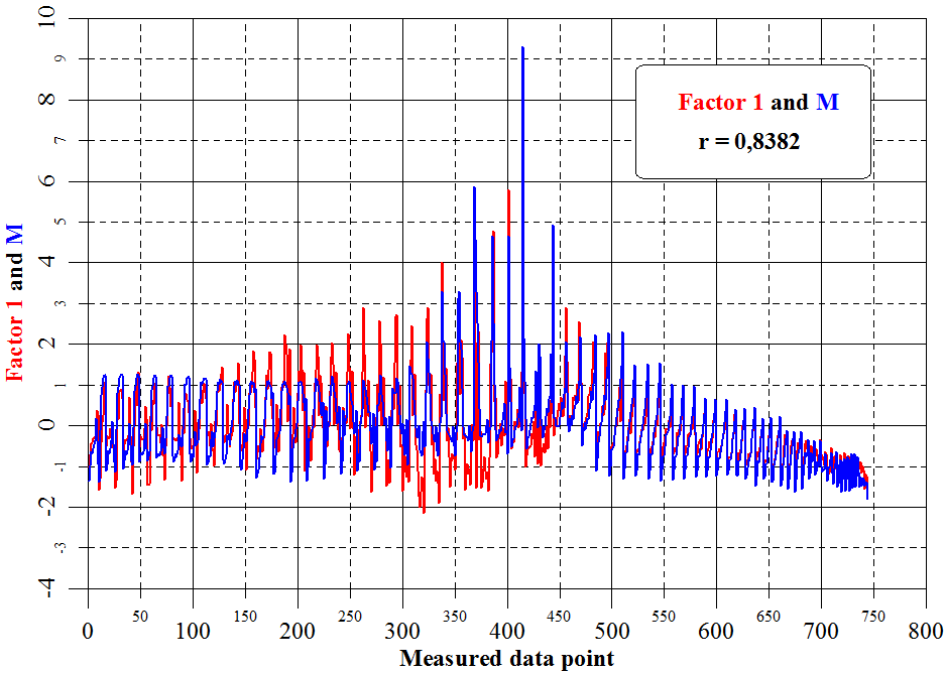

Figure 2. *Factor section 1 and M section.*

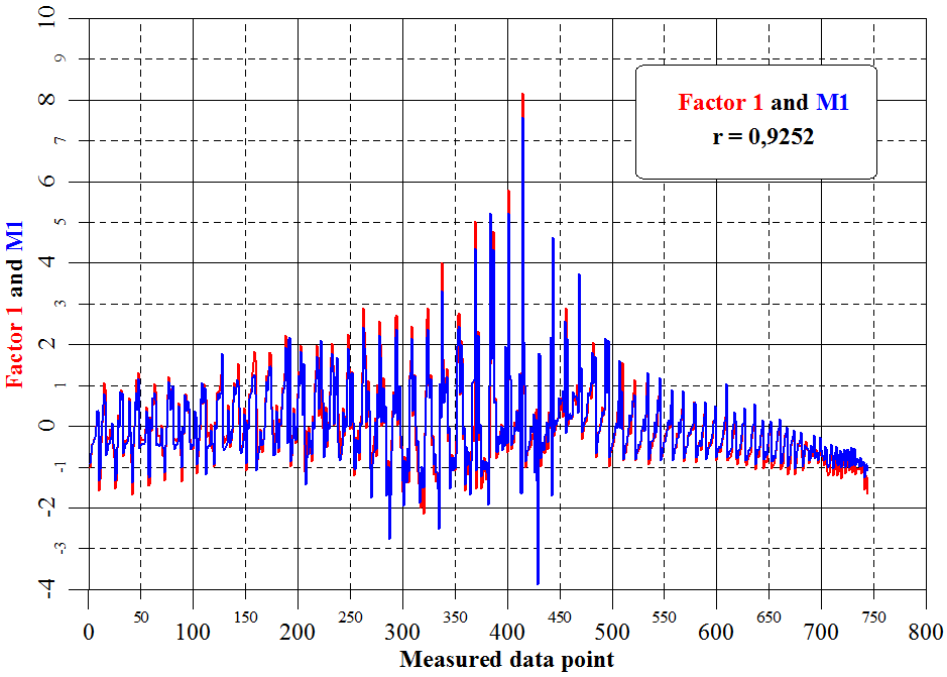

Figure 3. *Factor section 1 and M1 section.*

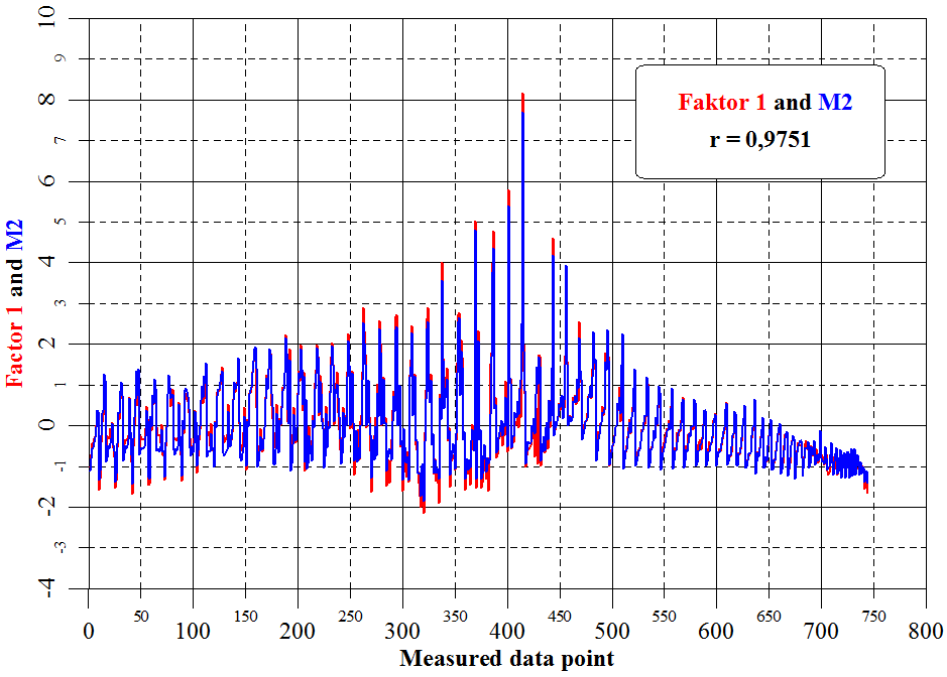

Figure 4. *Factor section 1 and M2 section.*

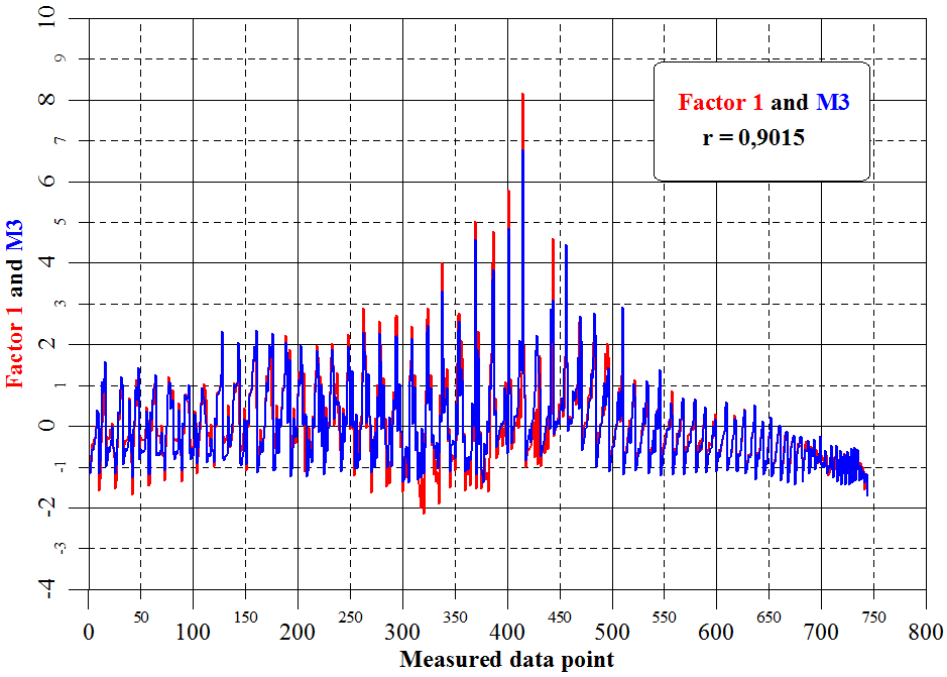

Figure 5. *Factor section 1 and M3 section.*

For the factor section 2 Fig. 6 shows the strongest correlation relationship with the apparent resistivity Rho. Examining the figure, we can see that the apparent resistivity (Rho) in all parts of the section can be excellently replaced by the value of factor 2. Factor 2 still has a moderately strong relationship with the Vp potential profile. This relationship is shown in Fig. 7.

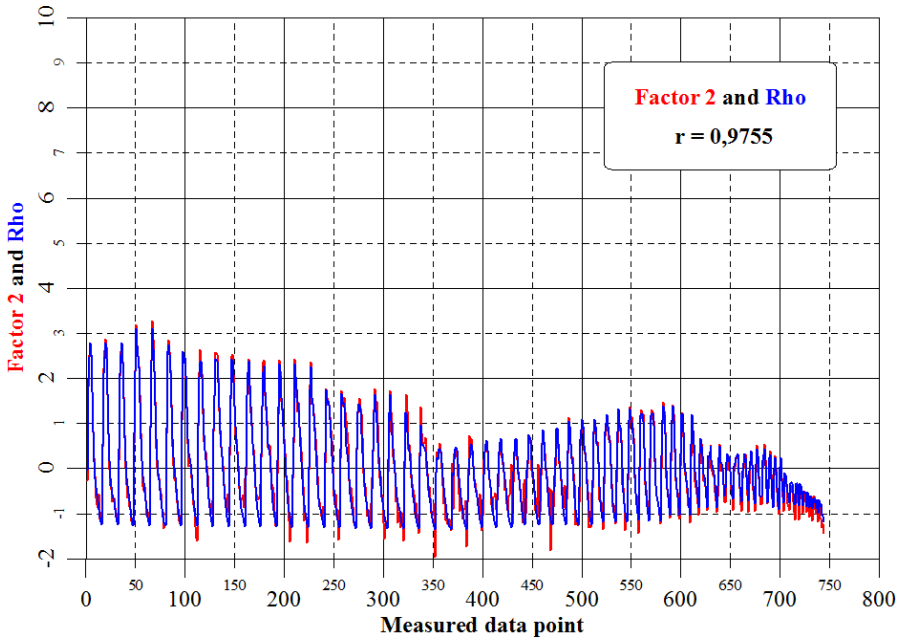

Figure 6. *Factor section 2 and Rho section.*

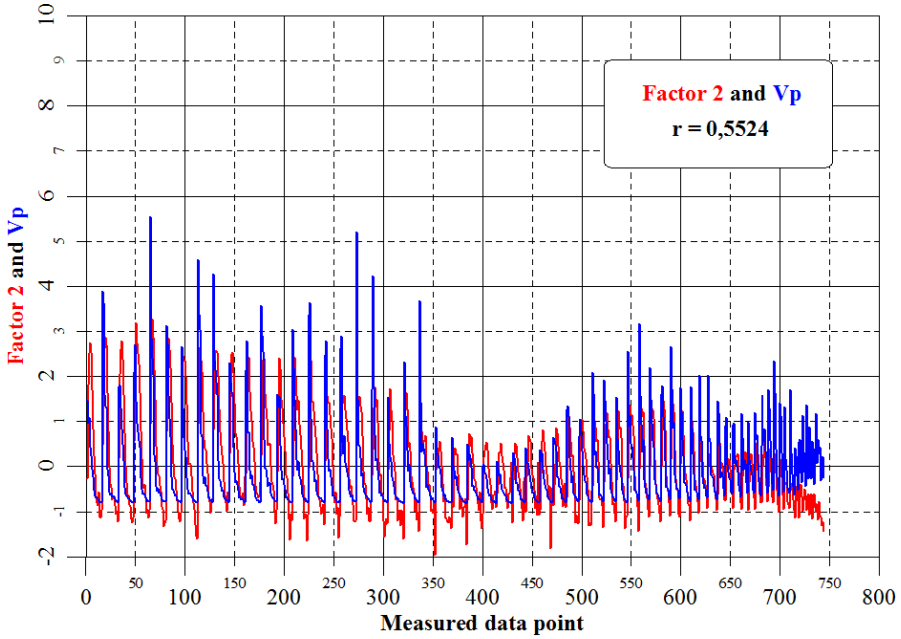

Figure 7. *Factor section 2 and Vp section.*

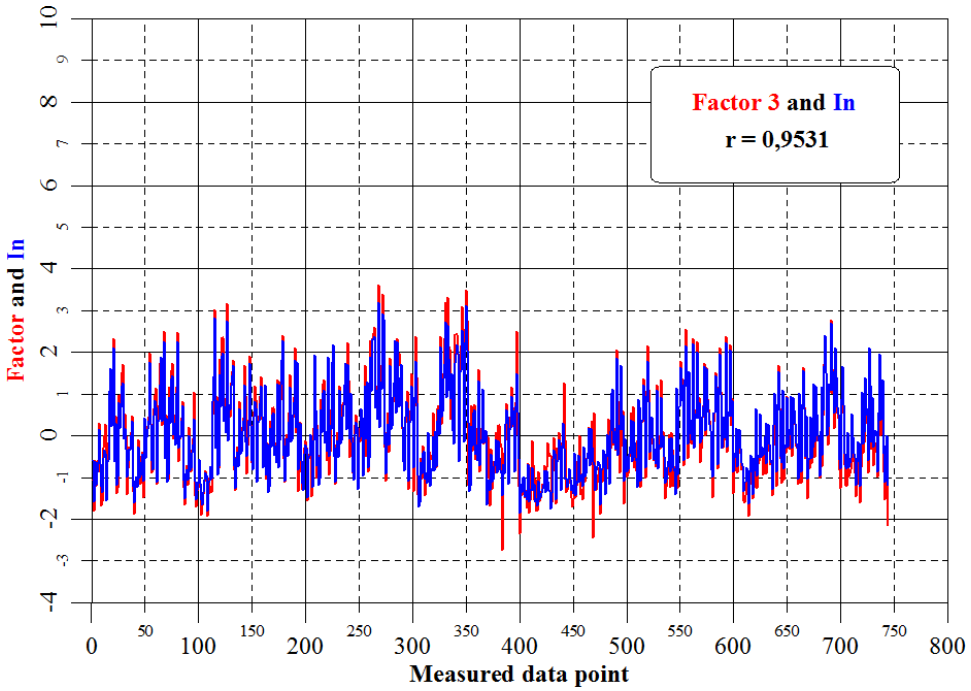

Figure 8. *Factor section 3 and In section.*

The other variables (M, M1, M2, M3, Dev, Sp, and In) have no significant relationship to factor 2. Factor 3 alone has a close relationship with the measurement current (In), which relationship is illustrated in Fig. 8. The other variables (Rho, M, M1, M2, M3, Dev, Sp, and Vp) have no significant relationship to factor 3. Factor 4 and 5 have a very weak relationship with all variables.

We also examined the difference in the factor loading of each variable if we used only a 2-factor model instead of the 5-factor model. The displacements of the factor loadings are shown in Fig. 9. Table 6 below summarizes the factor loadings, the extent of their change, and the magnitude of the displacement vectors. It can be seen that the change in the values of factor 2 is significantly larger than the change in the values of factor 1. The magnitude of the displacement vector is the largest for the variable Rho and the smallest for the variable Sp. Based on all this, the use of the 5-factor model is justified. This is also confirmed by the study of the independence of factor section 1 and factor section 2. The correlation coefficient between this two factor sections shown in Fig. 10 shows that there is practically no relationship between the two factor sections confirming the basic mathematical assumption that the factors must be uncorrelated.

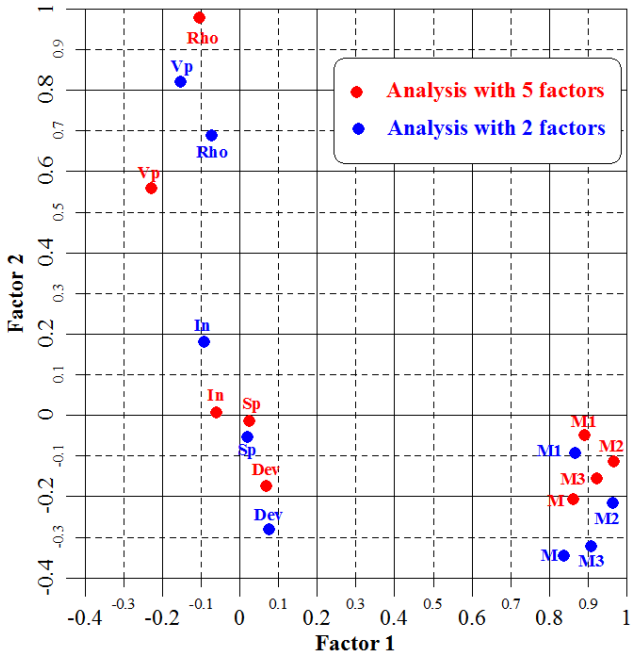

Figure 9. *The change in the first and second factor loadings of the 9-variable dataset.*

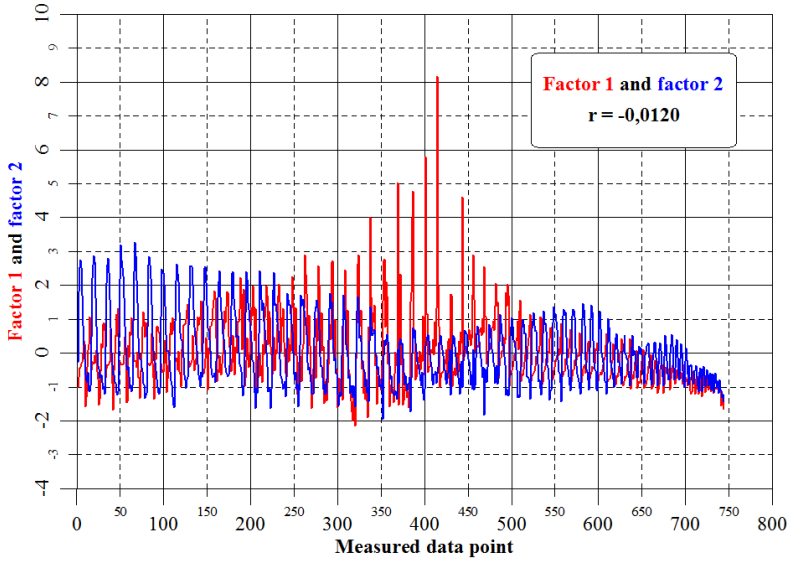

Figure 10. *Relationship between factor section 1 and factor section 2.*

5. Conclusions

Summarizing the results of the factor analysis, it can be concluded that factor 1 is closely related to the IP parameters (M, M1, M2 and M3) and factor 2 is mainly related to the apparent resistivity (Rho). The

section of the variable Rho can be replaced by the factor section 2 and the section of the variable M2 can be replaced almost entirely by the section 1. Factor section 3 shows a very close relationship with the measuring current (In) section. Factors 4 and 5 are not closely related to any of the variables. Based on the examination of the change of the factor values and the independence of the factor 1 and 2 sections, it seems justified to use the 5-factor model in the evaluation. Regarding the practical applicability of the results, the use of factor analysis to correct and replenishment erroneous data from multi-electrode measurements is very promising. Factor 2 can be used to replenish and correct apparent resistivity data, while factor 1 can be used to replenish and correct IP data. Multi-electrode geophysical measurements are widely used in international practice in the fields of mineral exploration and environmental investigations. The correction of erroneous field multi-electrode data by factor analysis can therefore represent a very important and effective new way in both domestic and international applications.

References

[1] Loke, M. H. (2000). Electrical imaging surveys for environmental and engineering studies. *A Practical Guide to 2-D and 3-D Surveys*, 61, Springer-Verlag.

[2] Kearey, P., Brooks, M. and Hill, I. (2002). *An introduction to geophysical exploration*. 3rd ed, Oxford: Blackwell Science Ltd.

[3] Dobróka, M., Szabó N. P. and Szegedi H. (2014). *Geophysical Information Processing – Methods and Applications (in Hungarian).* CriticEl Monográfia sorozat 4, Miskolc, Milagrosa Kft., ISBN:978-963-08-9325-1.

[4] Szabó, N. P. (2018). *Interpretation of well log-geophysical data by factor analysis and inversion procedures (in Hungarian)*. DSc. thesis, Budapest-Miskolc, extent: 102 p.

[5] Szabó, N. P., Valadez-Vergara, R., Tapdigli, S., Ugochukwu, A., Szabó, I. and Dobróka, M. (2021). Factor analysis of well logs for total organic carbon estimation in unconventional reservoirs. *ENERGIES, 14*(18*)*, Paper: 5978. **https://doi.org/10.3390/en14185978**

[6] Abordán, A. and Szabó, N. P. (2020). Uncertainty reduction of interval inversion estimation results using a factor analysis approach. *GEM - INTERNATIONAL JOURNAL ON GEOMATHEMATICS, 11*(1), Paper: 11. **https://doi.org/10.1007/s13137-020-0144-4**

[7] Turai, E. and Nádasi, E. (2020). *Multielectrode geophysical measurements in the area of Surány water base in the framework of the "Clean drinking water" project (in Hungarian)*. Exploration report, University of Miskolc, extent: 41 p.

[8] Spearman, C. (1904). The proof and measurement of association between two things. *Am. J. Psychol, 15*, 72-101. **https://doi.org/10.2307/1412159**

[9] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 2*, 559-572. **https://doi.org/10.1080/14786440109462720**

[10] Cudeck, R. and MacCallum, R. C. (2007). *Factor analysis at 100: Historical developments and future directions.* New Jersey: Lawrence Erlbaum Assoc. Inc. **https://doi.org/10.4324/9780203936764**

[11] Szabó, N. P. and Kormos, K. (2012). Shale content of freshwater formations estimated by factor analysis of borehole logs (in Hungarian). *Magyar Geofizika, 53(2),* 1-11.

[12] Kaiser, H. F. (1958). The varimax criterion for analytical rotation in factor analysis. *Psychometrika, 23*, 187-200. **https://doi.org/10.1007/BF02289233**