

Received 18.01.2018
Reviewed 13.07.2018
Accepted 13.08.2018A – study design
B – data collection
C – statistical analysis
D – data interpretation
E – manuscript preparation
F – literature search

Analysis of principal components used for modelling changes in glacitectonically disturbed areas

Maria MRÓWCZYŃSKA^{ABCDEF} ✉

orcid.org/0000-0002-4762-3999; University of Zielona Góra, Faculty of Civil Engineering, Architecture and Environmental Engineering, ul. Prof. Z. Szafrana 1, 65-516 Zielona Góra, Poland; e-mail: m.mrowczynska@ib.uz.zgora.pl

For citation: Mrówczyńska M. 2018. Analysis of principal components used for modelling changes in glacitectonically disturbed areas. Journal of Water and Land Development. No. 39 p. 119–123. DOI: 10.2478/jwld-2018-0066.

Abstract

Determination of vertical displacements of engineering objects is closely related to geodesic monitoring. Its purpose is to record the dynamics of changes in the deformation phenomenon. Geodesic monitoring requires the use of appropriate measurement equipment and appropriate methods for processing observation results, which make it possible to determine the correlation between the causes and effects of deformations in engineering objects. Progress in information technology resulted in the appearance of new methods for processing and compressing experimental data which are resistant to noise or interference and enable reduction of the amount of information.

The paper presents a method for statistical analysis of multidimensional data based on PCA (principal component analysis) transformation, implemented with the use of a neural network. PCA transformation, related to the Karhunen–Loeve transformation, is used for processing signals regarded as stochastic processes. This method makes enables reduction of the input data space on the basis of independent principal components with due attention to their significance. It also makes it possible to model changes occurring in both buildings and terrain in glacitectonically disturbed areas.

Key words: *geodesic monitoring, glacitectonically disturbed areas, PCA transformation, vertical displacements*

INTRODUCTION

Geodesic measurements are often carried out in tasks aimed at determining displacements and deformations of engineering objects and large areas of land vulnerable to changes caused by exogenous factors as well as human activity. In order to identify changes occurring in objects geodesic monitoring is used. It consists of geodesic measurements, their analysis and an engineering interpretation of the results. Geodesic measurements can be carried out using classical geodesic methods, methods based on satellite techniques and methods based on terrestrial or airborne laser scanning [KLAPA *et al.* 2017; KOSITSKY, AVOUAC 2010; ZACZEK-PEPLIŃSKA, KARSZNIA 2017]. In recent years, methods and instruments based on GIS techniques and remote sensing have gained in importance. They are used for determining displacements and terrain deformations, and also for assessing environmental risks and ones caused by human activity. These methods comple-

ment classical geodesic monitoring [SZEWRĄŃSKI *et al.* 2017].

Geodesic monitoring can be periodical (to assess the condition of an object at a particular moment) and permanent (to assess the dynamics of the phenomena occurring in an object and to identify processes caused by these changes). It is worth emphasizing that both construction work and the right operation of engineering objects require adequate geodesic monitoring, and its lack may lead to failures of even construction disasters [ZACZEK-PEPLIŃSKA *et al.* 2013].

Data obtained by monitoring have to be properly processed to produce reliable and adequately accurate results. Bearing in mind that geodesic monitoring may provide a large amount of information (especially if permanent), it is important that this information should be processed with methods that enable analysis and compression of multidimensional data [HEIDARI, MOATTAR 2017]. One of these methods is principal component analysis (PCA), which has

been widely used in recent years for modelling such non-linear phenomena as atmospheric studies, meteorology, biometrics, statistics, geophysics and civil engineering [NEJABAT *et al.* 2017; TIAMPO *et al.* 2017]. The PCA method has also been successfully used for solving geodesic problems. An example could be the use of the PCA method for optimizing the structure of geodesic measurement networks, which also enables detection of areas that are poorly represented in geodesic observations [NIEMEIER 1982] as well as improving the accuracy of the coordinates of points obtained using GPS technology [DONG *et al.* 2006].

This paper presents an attempt to use the PCA transformation for compressing measurement data using neural networks [LEŚNIAK, ZIMA 2018]. They are vertical displacements observed in engineering objects located in glaciectonically disturbed areas. Glaciectonically disturbed soils, which have been affected several times in their history by the mass of the expanding glacier, have variable or lower soil strength parameters than those defined in the standard. As a result, engineering objects located on these soils are especially vulnerable to the impact of vertical forces, which results in uneven settlement. For this reason in a number of cases geodesic monitoring is necessary and measurements carried out in the process are aimed at examining the settlement of the foundations [GONTASZEWSKA-PIEKARZ, MRÓWCZYŃSKA 2008]. The results of geodesic measurements have been analysed using the PCA transformation, which generates new variables called principal components. These variables enable reduction of the database size and they are represented by the eigenvectors of the covariance matrix or the correlation matrix [MRÓWCZYŃSKA 2005].

METHODS

Principal component analysis (also called the Hotelling transformation, the Karhunen–Loeve transformation or the orthogonal decomposition method) is a method of statistical analysis of multidimensional data, which performs linear transformations of the input vector \mathbf{x} into the output vector \mathbf{y} according to the following relation:

$$\mathbf{y} = \mathbf{W}\mathbf{x} \quad (1)$$

Where \mathbf{W} is the PCA decomposition matrix (the eigenvector coordinate matrix).

The linear transformation performed by the linear PCA network (Fig. 1) makes enables reduction of the output space while preserving the most important information about the process analysed.

In a self-organizing PCA neural network the training process takes place without supervision, i.e. there is no need for a pre-set output pattern related to the input pattern. The training process consists in detecting significant characteristics of the correlation between input signals in order to use them in the reproduction phase for finding a solution for patterns that do not participate in the training process. During the training process the objective function is minimized. It is defined as [OSOWSKI 2006]:

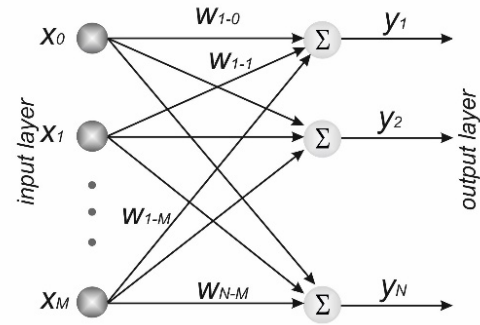


Fig. 1. The structure of a self-organizing PCA network; source: own elaboration

$$E = E_1 + E_2 \quad (2)$$

Where

$$E_1 = -\frac{1}{2} \sum_{i=1}^M \sum_{k=1}^M w_{ji} \mathbf{B}_{ik} w_{jk} \quad (3)$$

$$E_2 = -c_1 \sum_{i=1}^M w_{ji} - \frac{c_2}{2n} \left(\sum_{i=1}^M w_{ji} \right)^2 \quad (4)$$

The denotations in the formulas (3) and (4) are defined as: c_1 and c_2 are certain constants related to the input and the output vector, M is the number of training patterns, k is the number of neurons in the hidden layer, and \mathbf{B}_{ik} is the covariance matrix of the activity of the i^{th} and k^{th} neuron defined as

$$\mathbf{B}_{ik} = \frac{1}{p} \sum_{j=1}^p (x_i^j - \bar{x}_i)(x_k^j - \bar{x}_k) \quad (5)$$

Where: \bar{x}_i is the mean value of the components of the input vector \mathbf{x} , and p is the number the input vectors.

If we adopt the diagonal matrix $\mathbf{\Lambda} = [\lambda_1, \lambda_2, \dots, \lambda_n]$, created of the eigenvalues λ_i , then the covariance matrix \mathbf{B} can be written as

$$\mathbf{B} = \mathbf{W}^T \mathbf{\Lambda} \mathbf{W} \quad (6)$$

Where \mathbf{W} is the matrix of the coordinates of eigenvectors. If in the problem in question we consider only the first m eigenvalues, ordered according to decreasing values, we will obtain a vector of principal PCA components in the form (1). The first of these components has the greatest impact on the accuracy of the reconstruction of the input vector \mathbf{x} , which is carried out according to the Karhunen–Loeve transformation:

$$\check{\mathbf{x}} = \mathbf{W}^T \mathbf{y} \quad (7)$$

Principal components were estimated based on the adaptation of the weight matrix in step $t + 1$ using Sanger sequencing [OSOWSKI 2006]

$$w_{ij}(t + 1) = w_{ij}(t) + \eta y_i(t) [x_j - \sum_{l=1}^i w_{lj}(t) y_l(t)] \quad (8)$$

At this point it should be noted that the results of the recreation of the input vector depend on the level of correlation of the input data and also to a high extent on the value of the training coefficient η , whose value should be within the range $0 < \eta < \frac{2}{\lambda_{\max}}$ (λ_{\max} is the highest eigenvalue of the matrix \mathbf{B}).

RESULTS AND DISCUSSION

Measurement data were compressed using principal component analysis for a data set of vertical displacements of points of a measurement and control network (Fig. 2a) stabilized on an object located on glaciectonically disturbed soils. In such soil conditions the object is at risk of uneven settlement, especially in variable water conditions. For this reason geodesic monitoring is necessary. The measurements were carried out in the object using the precision levelling method in the years 2016–2017, in 7 measurement cycles carried out at different time intervals, which depended on the situation in the object as well as the renovation work that was in progress (Fig. 2b). All measurements and calculations were compared to the zero measurement carried out in June 2016.

During the process aimed at determining vertical displacements of points of the measurement and control network mutually fixed points were identified, which defined the reference system according to an algorithm consisting of two stages. The first stage was carried out in the following steps:

- minimization of the objective function as the sum of absolute deviations,
- checking the condition for fixed points using the iterative method in order to eliminate points that are not mutually fixed,
- the condition for fixed points was checked in terms of the shortest path, and the points that satisfied it were adopted as a preliminarily identified reference system.

- The second stage consisted of the following four steps:
- the final definition of the reference system, formulated on the basis of the reaction of the observation system by an increment of the square of corrections,
 - additional observation adjustments with the assumption that the new points in the network are fixed,
 - analysis of the critical value of the increment of the square of corrections.

The displacements of the measurement points that were determined are relative in character and they provide sufficiently precise information about the stability of the building [PRÓSZYŃSKI, KWAŚNIAK 2006].

The principal components were estimated based on the adaptation of the weight matrix using rule (8) and self-organizing PCA neural networks. The results are the vectors of the principal components and the weight matrix **W**:

$$\mathbf{W} = \begin{pmatrix} 0.6901 & -0.4676 & -0.0213 & -0.3532 & 0.0671 & 0.0207 \\ 0.3808 & 0.3120 & -0.0341 & 0.7519 & -0.1781 & -0.0467 \\ -0.1846 & 0.3628 & -0.7601 & -0.3160 & 0.0483 & 0.0101 \\ -0.3549 & -0.2399 & 0.1413 & 0.2426 & 0.7628 & 0.0464 \\ -0.4663 & -0.4840 & 0.0364 & 0.0667 & -0.6168 & -0.0471 \\ -0.0651 & 0.5166 & 0.6378 & -0.3920 & -0.0834 & 0.0155 \end{pmatrix}$$

The vectors of the principal components are shown in Table 1. Table 2 presents the results of the measurements of vertical displacements of points of the measurement and control neural network and the reconstruction of the displacement values based on PCA transformation with the number of principal components adopted as 3 and the value of the training coefficient $\eta = 0.002$. The accuracy of the reconstruction, characterized by the mean square error, is $m = 0.26$ mm.

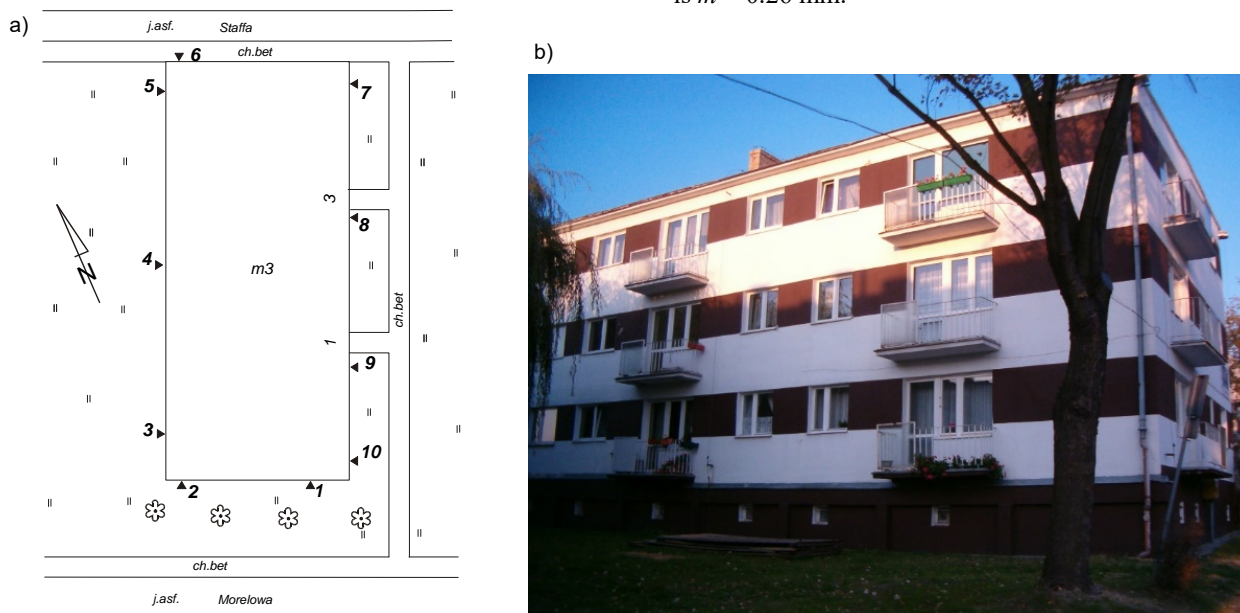


Fig. 2. Research object; source: own elaboration

Table 1. The vectors of the principal components

No.	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
1	-0.8806	-0.9014	0.4222	0.1548	0.7059	0.6313	0.9734	0.9105	0.5614	-0.8613
2	0.4397	0.4198	0.6558	0.7166	0.5767	0.4581	0.2154	0.0584	-0.2644	0.4526
3	-0.0218	-0.1126	-0.3377	0.6515	0.3035	-0.3539	-0.0570	-0.3511	0.4750	-0.0718

Source: own study.

Table 2. The vertical displacements of the control points obtained from the measurements and the ones reconstructed using the PCA method

Point no.	Specify method	Displacements of control points in different time (mm)					
		07.2016	08.2016	11.2016	12.2016	01.2017	02.2017
1	obtained	-2.79	-10.25	-13.47	-12.98	-12.86	-12.77
	reconstructed	-3.38	-9.04	-13.96	-12.73	-12.66	-13.34
2	obtained	-1.52	-3.35	-4.80	-4.37	-4.34	-4.26
	reconstructed	-1.58	-3.21	-4.86	-4.32	-4.34	-4.32
3	obtained	-0.71	-1.12	-0.90	-0.70	-0.57	-0.74
	reconstructed	-0.79	-0.95	-0.97	-0.64	-0.56	-0.83
4	obtained	+0.01	-0.14	0.02	0.04	+0.07	-0.43
	reconstructed	+0.03	-0.18	+0.4	-0.01	+0.10	-0.40
5	obtained	-0.06	-0.06	-0.02	-0.01	+0.05	-0.08
	reconstructed	-0.05	-0.08	-0.01	+0.01	+0.03	-0.08
6	obtained	+0.03	+0.12	0.00	+0.45	+0.32	+0.11
	reconstructed	+0.08	+0.02	+0.05	+0.32	+0.38	+0.18
7	obtained	0.00	+0.06	+0.26	+0.37	+0.49	+0.24
	reconstructed	0.00	+0.06	0.26	+0.40	+0.40	+0.47
8	obtained	+0.02	+0.02	+0.24	+0.42	+0.47	+0.42
	reconstructed	-0.01	+0.09	+0.21	+0.44	+0.48	+0.38
9	obtained	-0.31	-0.55	0.50	-0.03	-0.12	-0.02
	reconstructed	-0.48	-0.18	+0.35	-0.03	0.00	-0.19
10	obtained	-1.51	-5.59	-7.28	-6.80	-6.76	-6.59
	reconstructed	-1.88	-4.87	-7.57	-6.65	-6.62	-6.93

Source: own study.

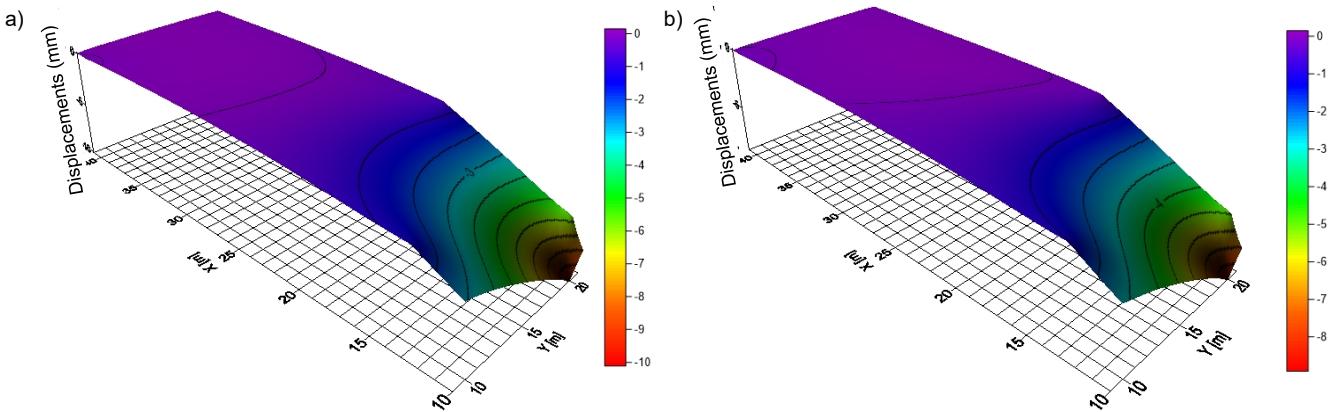


Fig. 3. The vertical displacements of the points: control (August 2016) obtained from the measurements, the ones reconstructed using the PCA method; source: own study

Figure 3a shows the results of the vertical displacement measurement carried out in August 2016, and Figure 3b shows the results of the reconstruction of these measurement data carried out with the PCA algorithm. For this measurement period the reconstruction of measurement data is the least favourable and its accuracy is $m_{08_16} = 0.47$ mm.

CONCLUSIONS

PCA transformation carried out using self-organizing neural networks can be used as a method of reducing the dataspace size, also in relation to data obtained by geodesic monitoring. The use of lossy transformation enables replacement of a large amount of input information with a set of principal components and eigenvectors of the covariance matrix.

PCA enables reconstruction of input data with different levels of accuracy. In the case of the problem presented in the paper, reduction of principal components led to

compression of experimental data using only the largest component y_i of the vector y which enabled triple reduction of the amount of processed information. At the same time the accuracy of the input data reconstruction was within the accuracy of the measurements. The use of PCA transformation also has a positive technical meaning, allowing for faster geodetic measurements and engineering calculations due to the smaller amount of information necessary to perform them, which leads to a reduction in their cost-effectiveness. This proves that the method discussed in the paper can be used for compressing data obtained by geodesic monitoring of engineering objects.

REFERENCES

DONG D., FANG P., BOCK Y., WEBB F., PRAWIRODIRDJO L., KEDAR S., JAMASON P. 2006. Spatiotemporal filtering using principal component analysis and Karhunen-Love expansion approaches for regional GPS network analysis. Journal of

- Geophysical Research Vol. 111. B03405. p. 1–16. DOI 10.1029/2005JB003806.
- GONTASZEWSKA-PIEKARZ A., MRÓWCZYŃSKA M. 2008. Charakterystyka nierównomiernych osiadań obiektu budowlanego posiadającego na gruntach ekspansywnych. W: Renowacja budynków i modernizacja obszarów zabudowanych [Characteristics of uneven settlements of a building object erected on expansive lands. In: Renovation of buildings and modernization of built-up areas]. Ed. T. Biliński. Zielona Góra. Ofic. Wydaw. UZ. p. 119–130.
- HEIDARI M., MOATTAR M.H. 2017. Discriminative geodesic Gaussian process latent variable model for structure preserving dimension reduction in clustering and classification problems. *Neural Computing and Applications* p. 1–14. DOI 10.1007/s00521-017-3273-4.
- KLAPA P., MITKA B., ZYGMUNT M. 2017. Application of integrated photogrammetric and Terrestrial Laser Scanning data to cultural heritage surveying. *WMESS 2017 IOP Conf. Ser. Earth and Environmental Science*. Vol. 95 p. 1–8. DOI 10.1088/1755-1315/95/3/032007.
- KOSITSKY A.P., AVOUAC J.P. 2010. Inverting geodetic time series with a principal component analysis-based inversion method. *Journal of Geophysical Research*. Vol. 115. B03401 p. 1–19. DOI 10.1029/2009JB006535.
- LEŚNIAK A., ZIMA K. 2018. Cost calculation of construction projects including sustainability factors using the Case Based Reasoning (CBR) method. Vol. 10(5) p. 1–14. DOI 10.3390/su10051608.
- MRÓWCZYŃSKA M. 2005. Kompresja wyników eksperymentu na podstawie transformacji PCA [Compression of experimental results based on PCA transformation]. *Prace Naukowe Instytutu Górnictwa Politechniki Wrocławskiej*. Vol. 114 p. 255–262.
- NEJABAT M., NEGAHDARSABER M., GHAHARI G. 2017. Range of soil and climate characteristics appropriate for *Pistacia atlantica* forest development and rehabilitation (case study: Fars province Iran). *Journal of Water and Land Development*. No. 32 p. 71–78. DOI 10.1515/jwld-2017-0008.
- NIEMEIER W. 1982. Principal Component Analysis and geodetic networks – Some basic considerations. In: Meeting of FIG Study Group 5B Survey Control Networks. Aalborg. 07–09.07.1982. Eds. K. Borre, W. Welsch. *Proceedings Survey Control Networks*. H. 7 p. 275–292.
- OSOWSKI S. 2006. Sieci neuronowe do przetwarzania informacji [Neural networks for information processing]. Warszawa. Ofic. Wydaw. PW. ISBN 978-83-7207-615-1 pp. 422.
- PRÓSZYŃSKI W., KWAŚNIAK B. 2006. Podstawy geodezyjnego wyznaczania przemieszczeń: Pojęcia i elementy metodyki [Basics of geodetic determination of displacements: Concepts and elements of methodology]. Warszawa. Ofic. Wydaw. PW. ISBN 8372076294 pp. 212.
- SAMAT A., GAMBA P., ABUDUWAILI J., LIU S., MIAO Z. 2016. Geodesic flow Kernel support vector machine for hyperspectral image classification by unsupervised subspace feature transfer. Vol. 8. Iss. 3 p. 1–23. DOI 10.3390/rs8030234.
- SKRZYPCZAK I., KOKOSZKA W., KOGUT J., OLENIACZ G. 2017. Methods of measuring and mapping of landslide areas. *World Multidisciplinary Earth Sciences Symposium 2017. Book Series: IOP Conference Series-Earth and Environmental Science*. Vol. 95 p. 1–9.
- SZEWRĄŃSKI S., KAZAK J., ŻMUDA R., WAWER R. 2017. Indicator-based assessment for soil resource management in the Wrocław larger urban zone of Poland. *Polish Journal of Environmental Studies*. Vol. 26. No. 5 p. 2239–2248. DOI 10.15244/pjoes/70178.
- TIAMPO K.F., GONZÁLEZ P.J., SAMSONOV S., FERNÁNDEZ J., CAMACHO A. 2017. Principal component analysis of MSBAS DInSAR time series from Campi Flegrei, Italy. *Journal of Volcanology and Geothermal Research*. Vol. 344 p. 139–153.
- ZACZEK-PEPLINSKA J., KARSNIA K. 2017. Wykorzystanie nowoczesnych technik pomiarów geodezyjnych w technicznej kontroli zapór. Budowle piętrzące – eksploatacja i monitoring [The use of modern surveying techniques in technical inspection of dams. High-rise buildings – operation and monitoring]. *Monografie Instytutu Meteorologii i Gospodarki Wodnej*. Warszawa. IMGW-PIB p. 19–35.
- ZACZEK-PEPLINSKA J., PASIK M., POPIELSKI P. 2013. Geodezyjny monitoring obiektów w rejonie oddziaływania budowy tuneli i głębokich wykopów – doświadczenia i wnioski [Geodetic monitoring of objects in the area of impact of tunnel construction and deep excavations – experience and conclusions]. *Acta Scientiarum Polonorum Architectura*. Vol. 12. No. 2 p. 17–31.

Maria MRÓWCZYŃSKA

Zastosowanie analizy składników głównych do modelowania zmian na terenach zaburzonych glaciectonicznie

STRESZCZENIE

Wyznaczenie przemieszczeń pionowych obiektów inżynierskich jest ściśle związane z monitoringiem geodezyjnym, mającym za zadanie zarejestrowanie dynamiki zmian zjawiska deformacji. Monitoring geodezyjny wymaga zastosowania odpowiedniego sprzętu pomiarowego oraz odpowiednich metod przetwarzania wyników obserwacji, umożliwiających określenie związku między skutkami a przyczynami deformacji badanego obiektu. Wraz z rozwojem technologii informacyjnej zaczęto przetwarzać dane eksperymentalne w sposób odporny na szумы i zakłócenia oraz stosować kompresję danych, pozwalającą na zmniejszenie ilości informacji.

W pracy przedstawiono metodę analizy statystycznej wielowymiarowych danych za pomocą transformacji metodą analizy głównych składowych (PCA – principal component analysis), realizowanej z wykorzystaniem sieci neuronowej. Transformacja PCA, związana z transformacją Karhunen–Loevego, znajduje zastosowanie w przetwarzaniu sygnałów traktowanych jako procesy stochastyczne. Omawiana w pracy metoda umożliwia zmniejszenie przestrzeni danych wejściowych na podstawie wyznaczonych niezależnych składników głównych z uwzględnieniem ich znaczenia oraz modelowanie zmian na terenach zaburzonych glaciectonicznie, zarówno w odniesieniu do obiektów budowlanych, jak i samego terenu.

Słowa kluczowe: grunty zaburzone glaciectonicznie, monitoring geodezyjny, przemieszczenia pionowe, transformacja PCA