*Journal of Basic & Applied Sciences*, 2023, *19*, 151-162

Published by SET Publisher

Journal of Basic & Applied Sciences

ISSN (online): 1927-5129


# Methods of Cluster Analysis for Detection of Uniform Displacement Zones of Landslides and Anti-Landslide Structures

Khailak Andrii[1] and Shults Roman[2,*]

[1]*Research Institute of Building Production Named After V.S. Balitsky, Lobanovskyi Ave, 03037 Kyiv, Ukraine*

[2]*Interdisciplinary Research Center for Aviation and Space Exploration, King Fahd University of Petroleum and Minerals, 35712 Dhahran, Saudi Arabia*

**Article Info:**

*Keywords:*
Landslide,
cluster analysis,
clustering method,
similarity measure,
horizontal displacement,
evaluation.

*Timeline:*
Received: June 29, 2023
Accepted: July 26, 2023
Published: August 24, 2023

*Citation:* Andrii K and Roman S. Methods of cluster analysis for detection of uniform displacement zones of landslides and anti-landslide structures. J Basic Appl Sci 2023; 19: 151-162.

**Abstract:**

The primary goal of landslide monitoring is the development and implementation of appropriate prediction models. Such models will allow forecasting of the anticipated landslide movements and failures. The deployment of these models is only possible by the results of geospatial monitoring. However, the measured displacements of the monitoring targets mostly have different values that may deviate a couple of times for different parts of the observed landslide. Therefore, the correct prediction model can be developed for the points with similar displacements, or in other words, for the points with the same displacement velocities. The grouping of points with similar values is known as clustering or zoning task. Having the groups of similar displacements, it is possible to work out the proper prediction model for each group of displacements and detect the probable blunders in the measurements. The paper outlines the results of geospatial monitoring for landslide and anti-landslide structures carried out for small-scale landslide and a system of retaining walls in Kyiv, Ukraine. The efficiency of cluster analysis for uniform displacement zone identification has been studied by the results of geospatial monitoring. The basic principles and ideas of cluster analysis and clustering methods have been given. The different clustering methods have been examined. Each clustering method's efficiency has been estimated by distance determination methods and similarity measures. The quantitative analysis of the considered clustering methods was checked by evaluation analysis. The most reliable results in a line of the study have demonstrated centroid clustering and furthest neighbor clustering. The determined similarity measures for those two methods were almost the same.

*DOI:* https://doi.org/10.29169/1927-5129.2023.19.13

*Corresponding Author
E-mail: rshults@mtu.edu

## 1. INTRODUCTION

Landslide monitoring is a challenging problem faced by geologists, civil engineers, and surveyors in their practice. Apart from the threats that landslides create for people and urban facilities, their collapses have a significant socioeconomic impact [1, 2]. This is why landslide monitoring is essential for management in urban territories. Several new monitoring methods and technologies have been developed in the last decades. Integrating geospatial and non-geospatial technologies has allowed for achieving a prominent accuracy and quality of monitoring. Terrestrial and satellite technologies, such as satellite and ground-based radar interferometry, aerial and terrestrial laser scanning, close-range and UAV photogrammetry, GNSS, and robotic total stations [3-8] have been accompanied by various geotechnical technologies [9-11]. Today, we may admit that the data collection methods satisfy the requirements for detailed landslide analysis. However, the prediction of landslide activity is still challenging. State-of-the-art landslide susceptibility estimation and forecasting methods are based on the last achievements of mathematics and physics [12,13]. Among the various approaches, it is worth mentioning neural networks [14], machine learning methods [15], structural mechanics [16], and fuzzy logic [17]. The central premise of the correct model construction is the proper interpretation of the obtained values. The measurement results themselves may mislead researchers and make them accept the wrong decision. For such a complex problem as landslide monitoring, finding a single model describing the deformation process is often impossible. The reason is the complexity of an even small landslide to which different points may undergo displacements with different values and directions [18]. Moreover, urban landslides are the system that includes the landslide, structures emplaced on/nearby the landslide, and the system of anti-landslide structures, usually retaining walls [19-21]. That is why, before constructing any forecasting model, it is badly needed to analyze the results. Before the simulation, we must be sure that the data are free of gross and systematic errors and, what is more critical, correspond to the same deformation process, i.e., have similar displacements in a statistical sense. The first task is being solved by various statistical testing, while for the second task, we have to use more advanced methods. There are different ways to divide a set of deformation targets into zones or blocks where the displacements are uniform or changing evenly. Statistical methods were the first used to partition the results into groups with uniform displacements. However, statistical methods operate exclusively with displacements and provide a general picture of landslide behavior [22]. Meanwhile, such parameters as temperature, target coordinates, precipitation, etc., significantly impact the displacement value and direction. Many attempts have been made to highlight the regions of similar displacements, but most used simple time series analysis [23,24]. Among the methods that may incorporate additional parameters, cluster analysis is the most widespread and reliable [25-27]. The leverage of cluster analysis has grown significantly since disseminating powerful PC and machine learning algorithms. The review results have suggested to the authors the idea of applying the method of cluster analysis to detect the zones of uniform displacements.

The paper's primary goal is to estimate the opportunities of cluster analysis for detecting zones with uniform displacements for the case study of landslides and anti-landslide structures. The paper consists of five sections. The first section outlines the research problem introduction. The second section briefly describes the study objects and monitoring results. The third section deals with cluster analysis background, while the fourth presents the study results, their interpretation, and discussion. The fifth section contains conclusions and recommendations.

## 2. STUDY OBJECTS

As mentioned in the introductory section, two objects have been selected to study cluster analysis efficiency. For both cases, the monitoring workflow was similar. It was based on the conventional terrestrial observations by total station. The coordinate systems were not referenced to state or global coordinate systems to exclude possible errors. Generally, the monitoring workflow has the following steps given in Figure **1**. This scheme was applied in both cases.

The first study object is a small landslide emplaced in the center of Kyiv on the right bank of the Dnipro River. The landslide has approximate sizes of 100x130 m, with a height difference of up to 20 m (see Figure **1**). The borders of the landslide are portrayed in Figure **3**, where the designed monitoring network is also given. The suggested monitoring network ensured the spatial accuracy of approximately 5 mm and 2.9 mm along each coordinate axis, respectively. As shown in Figure **3**, the landslide is covered by trees and surrounded by buildings. After two observation epochs, the monitoring

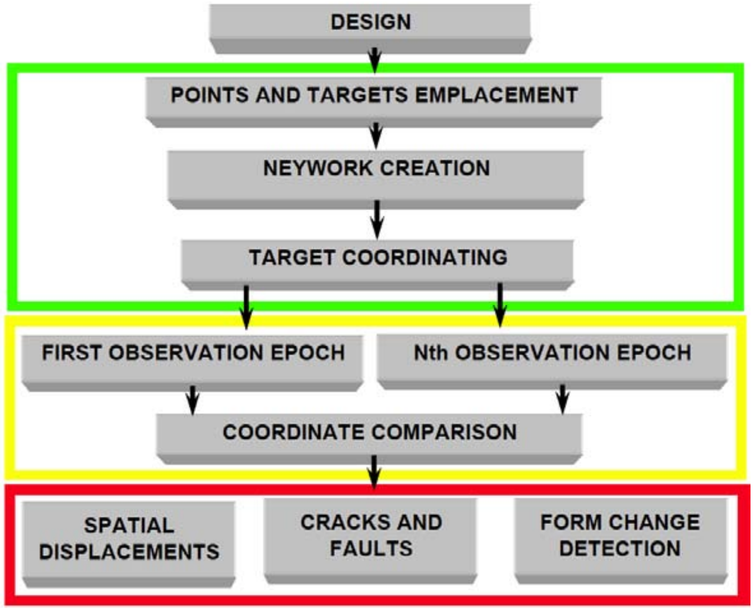


**Figure 1:** General workflow of landslide monitoring.

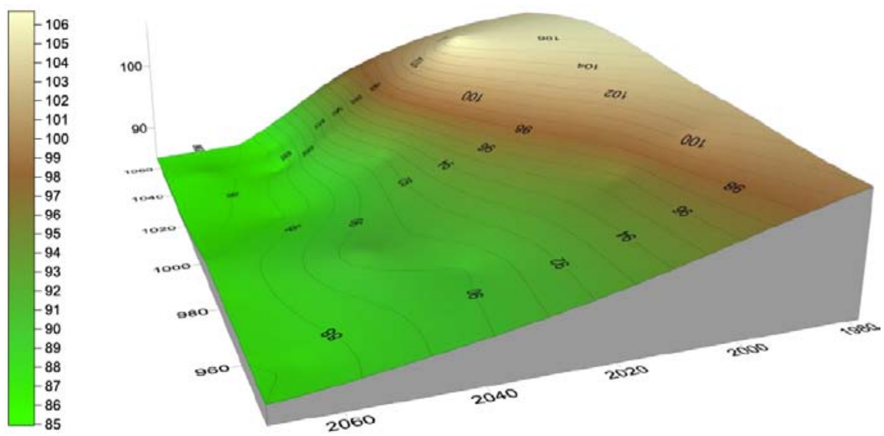

**Figure 2:** Landslide surface.

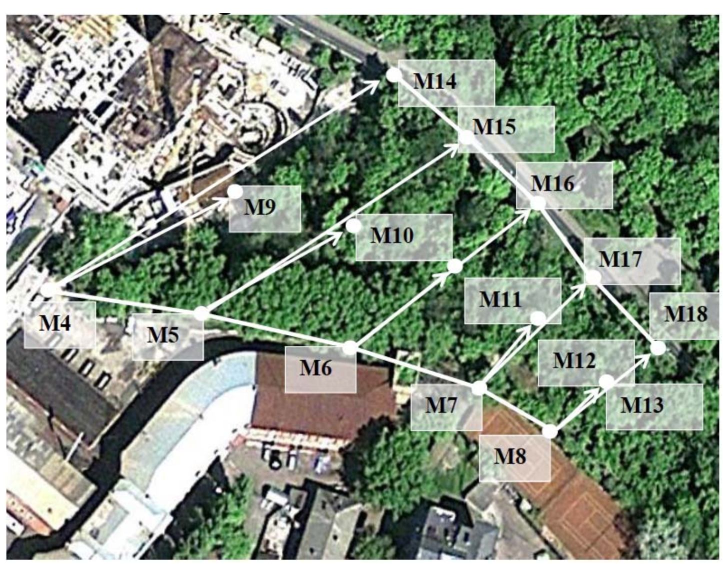


**Figure 3:** The landslide emplacement and designed monitoring network.

scheme was changed, and the final position of the monitoring points is presented in Figure **4**.

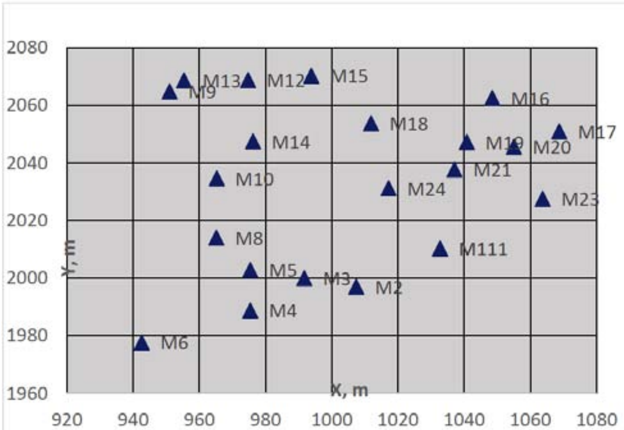


**Figure 4:** The landslide emplacement and designed monitoring network.

The monitoring lasted one year and contained four observation epochs. Finally, the target displacements were determined. The values of spatial displacements are given in the contour plot in Figure **5**, while their directions are presented in the vector plot in Figure **6**. Despite relatively small spatial displacements (-6.0 - -9.6 mm), for prediction, it is essential to know whether the landslide has moved as a whole body or as a set of separate blocks. This is why this landslide was selected for the following studies.

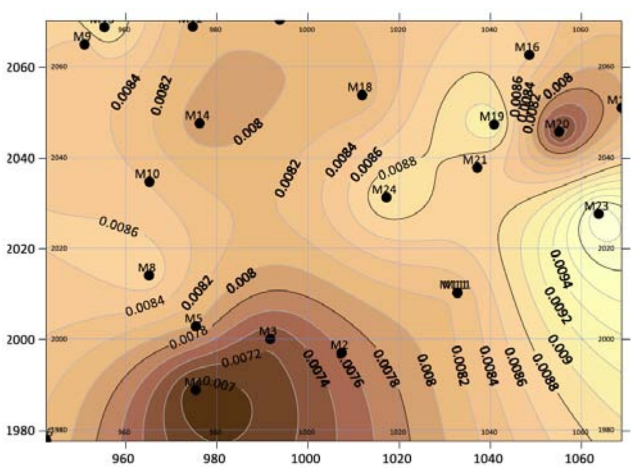


**Figure 5:** The landslide displacements field.

The second study object is a system of retaining walls that protects the buildings at the foot of the landslide. The landslide slope has a height of 40 meters and a width of over 300 meters. The landslide is held by four retaining walls (PS-1, PS-2, PS-3, PS-4) (Figure **7**). The retaining walls' height varies from 8 to 14 meters. All the retaining walls have a pile foundation with piles at a depth of 20 meters.

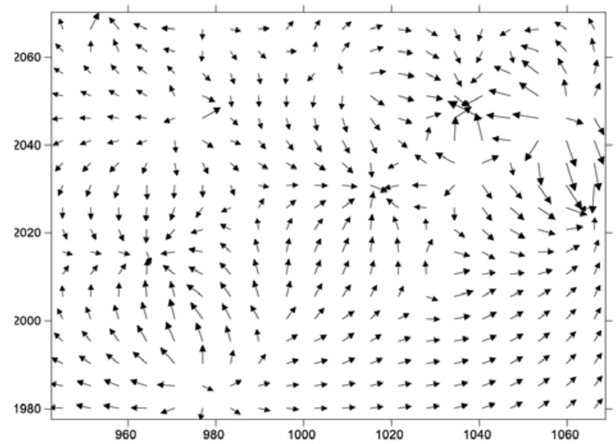


**Figure 6:** The vector plot of the landslide displacements.


**Figure 7:** A general view of the system of retaining walls.

To determine the spatial displacements, a network of five points was created. The main requirement was to determine the displacements in the direction perpendicular to the retaining walls' surface. Therefore, the highest observation accuracy was achieved along this direction (2 mm). The monitoring was carried out weekly for six months. The total number of observation epochs is 27. Figure **8** demonstrates the determined displacements for all targets.

The preliminary analysis shows that the displacements differ depending on the retaining wall, target position, observation epoch, etc. The general tendency demonstrates the increase of displacements over time. The largest displacements were determined for the wall PS-2 (up to 25 mm). Therefore, this wall was selected for further study.

## 3. CLUSTER ANALYSIS

Before we have an in-depth study of the objects considered above, we need to review the main

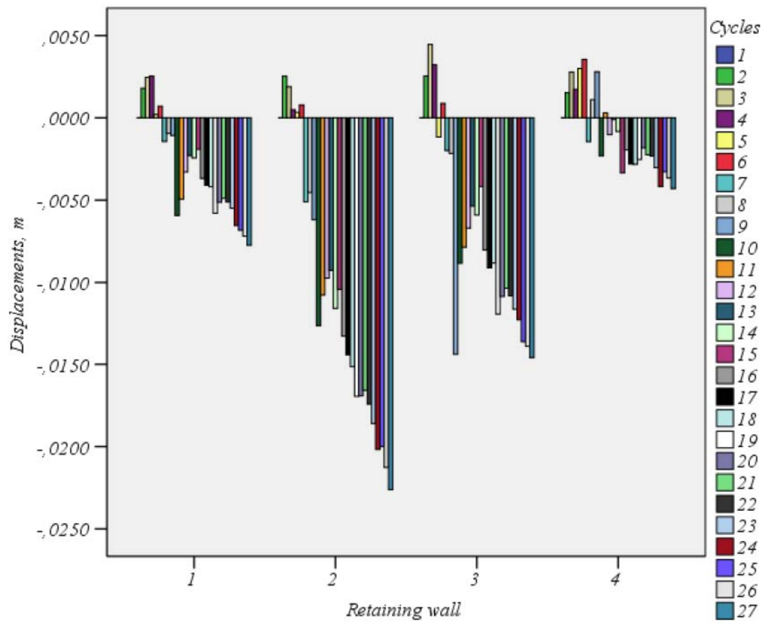


**Figure 8:** Displacements distribution for each retaining wall.

principles and characteristics of cluster analysis that we will apply in what follows. One may find a detailed description of the clustering algorithms, e.g., in [26,27]. The workflow of cluster analysis is given in Figure **9**.

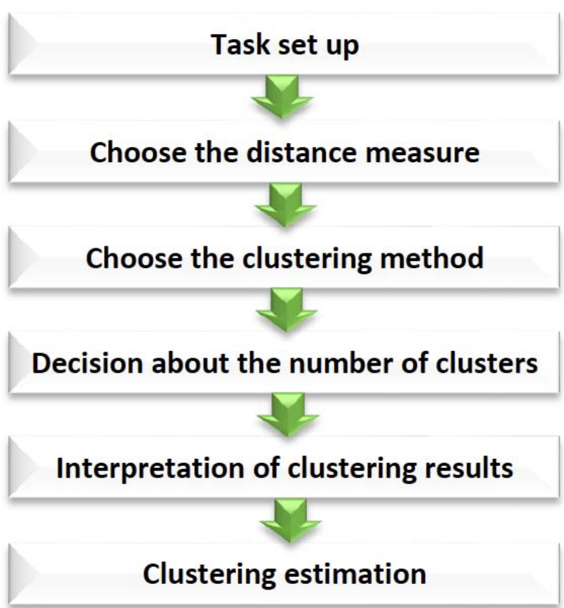


**Figure 9:** Clustering algorithm.

The first parameter that a researcher must define is a distance measure. There are different methods of distance measures. The most common methods of distance measure are outlined below.

*Euclidean distance* is the shortest distance between two points that, in a two-dimensional case, is defined as

$$D(X,Y)=\sqrt{\sum_{i=1}^{m}(X_i-Y_i)^2} \tag{1}$$

*Squared Euclidean distance* is the distance that better takes account of significant differences. This measure is suggested for centroid clustering, median clustering, and Ward's method. The measure is defined as

$$D^2(X,Y)=\sum_{i=1}^{m}(X_i-Y_i)^2 \tag{2}$$

*Correlation between two vectors or Pearson correlation.*

$$r=\frac{\sum_{i=1}^{m}(X_i-\bar{X})(Y_i-\bar{Y})}{(n-1)S_X S_Y} \tag{3}$$

*Cosine of distance vectors*. This measure ranges from -1 to +1, which is also valid for the previous measure.

$$r=\frac{\sum_{i=1}^{m}(X_iY_i)}{\sqrt{\sum_{i=1}^{m}(X_i^2)\sum_{i=1}^{m}(Y_i^2)}} \tag{4}$$

*Chebychev distance* is the longest of difference between two vectors.

*Block measure*. This measure is defined as sum of absolute differences between pairs of values.

$$D(X,Y)=\sum_{i=1}^{m}\left|X_i-Y_i\right| \tag{5}$$

*Minkowski distance.*

$$D(X,Y) = \left( \sum_{i=1}^{m} |X_i - Y_i|^r \right)^{\frac{1}{r}} \quad (6)$$

*Manhattan distance or city block distance* is calculated as the distance in the X direction plus the distance in the Y direction.

After the measure choice, it is necessary to choose the clustering method. Among the basic methods, the most popular are:

*k-Means clustering* is the method in which the centroid of the new cluster is obtained as a mean weighted of centroids of both initial clusters. The observation number determines the initial clusters' weight coefficient (Figure **10**).

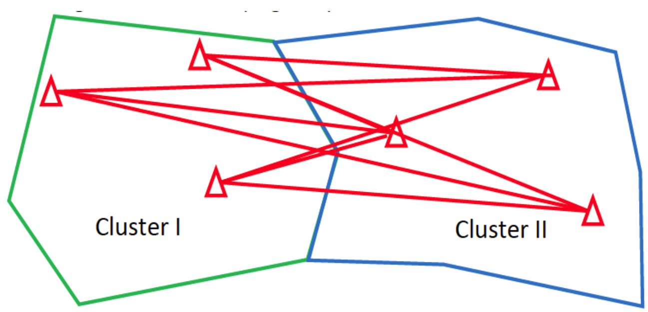


**Figure 10:** Centroid clustering.

*Median clustering* is similar to centroid clustering, but the weights of both clusters are the same.

*Mean linkage (between groups linkage) clustering*. For the between groups linkage, the distance between clusters equals to mean value of distances between all possible observation pairs. One observation is taken from one cluster, whereas the other is from another. Information for distance calculation is determined based on all possible observation pairs.

*Ward's method.* In the beginning, the mean values for all variables are calculated for both clusters. Then, squared Euclidean distances between each observation and the mean value of the particular variable are figured out. These distances are summed up. The clusters are joined for which the increment of the sum of distances has the least value.

*Mean linkage (within groups linkage) clustering.* For the within-groups linkage, the distance between clusters is figured out by all possible observation pairs, including pairs inside the clusters.

*Nearest neighbor clustering.* The distance between two clusters is determined as the shortest distance between observation pairs (Figure **11**).

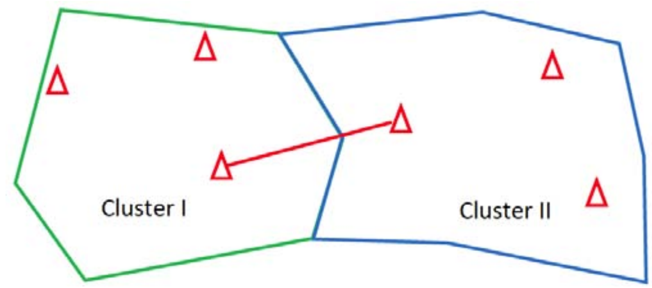


**Figure 11:** Nearest neighbor clustering.

*Furthest neighbor clustering.* The distance between two clusters is determined as the longest distance between observation pairs (Figure **12**).

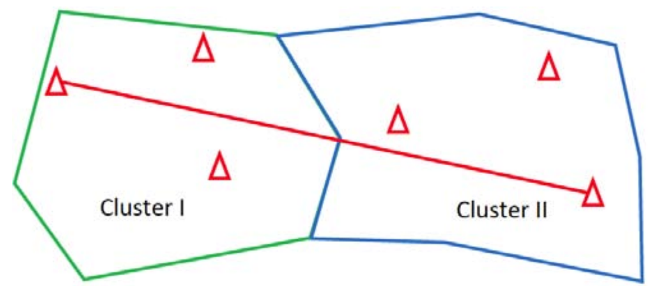


**Figure 12:** Furthest neighbor clustering.

All the considered methods can be grouped into more general clustering methods: Hierarchical clustering, k-Means and k-Medoids clustering, and k-Nearest Neighbor clustering. Some of them need a preliminary assignment of clusters' numbers. The decision about the number of clusters is a tricky question. However, for such small objects presented in this paper, the number of clusters does not exceed three. Moreover, four different criteria were used to evaluate the number of clusters in k-Means and k-Medoids clustering: Calinski-Harabazs, Davis-Bouldin, Silhouette, and Gap. The main questions are the existence confirmation of the different clusters and targets belonging to the different clusters.

## 4. CLUSTERING THE STUDY OBJECTS AND RESULTS DISCUSSION

Let's consider the results of cluster analysis efficiency. As it was mentioned, the first study object is a small landslide. The considered above clustering methods and measures have been tested on this landslide. Input parameters for cluster analysis were target spatial coordinates, target spatial displacements, and local hill

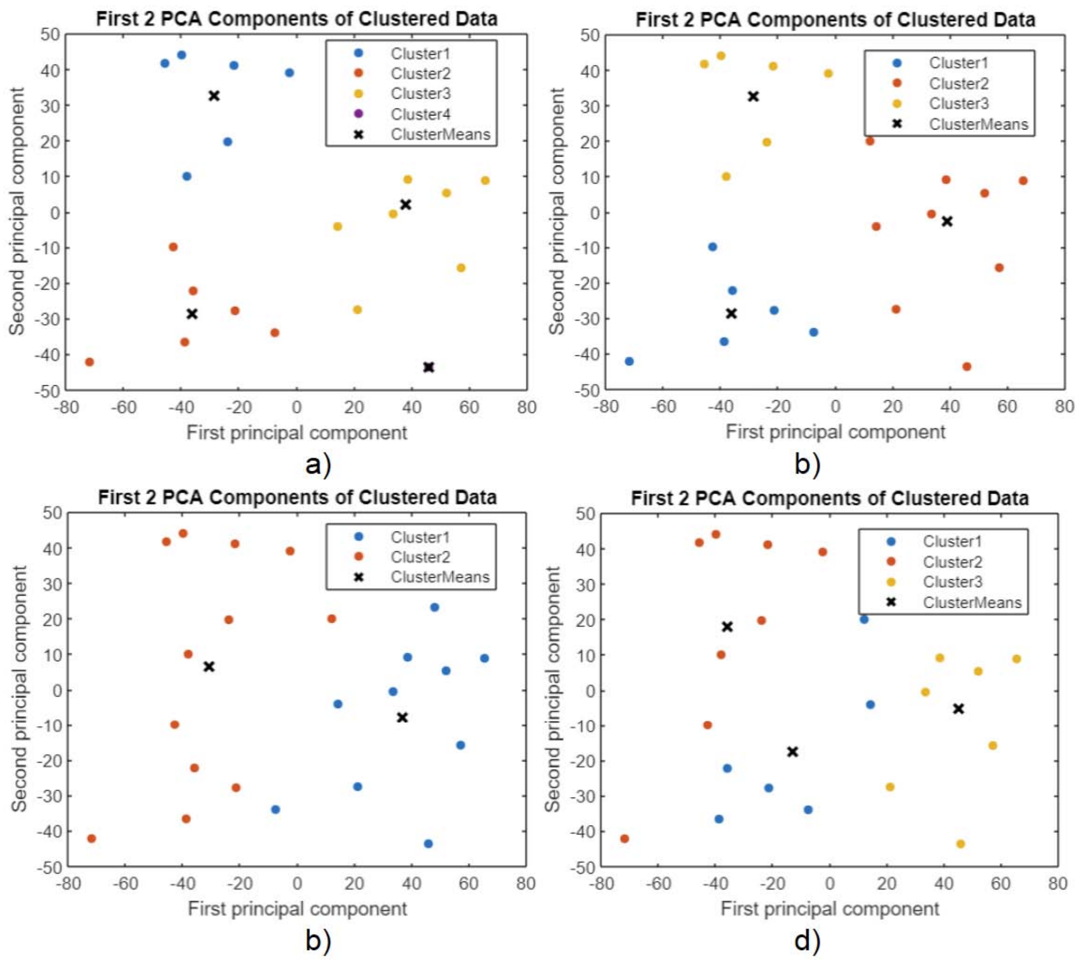


**Figure 13:** Centroid clustering for **a**) Euclidean distance; **b**) Cosine of distance vectors; **c**) Manhattan distance; **d**) Correlation between two vectors.

inclination. The data processing has been accomplished in IBM SPSS statistical software. Some visualizations have been done in MATLAB. In Figure **13**, one may see the clustering results for the centroid method for four different measures. Four criteria were used to estimate the clusters' optimal number: Calinski-Harabazs, Davis-Bouldin, Silhouette, and Gap. The sample results of these criteria applications are given in Figure **14**.

The obtained clustering results seem unreliable. The determined displacements are in a range of 6 – 10 mm. The differences between clusters' centers are less than 1 mm, which is undetectable with the measurement accuracy used. Due to the insignificant values of the displacements, different methods output different numbers of clusters. In this case, the right decision is to infer the absence of any separate clusters and determine the landslide as a solid body. This conclusion is supported by visual inspection that did not recognize any cracks or ruptures in the landslide body. With this inference, we may construct a single prediction model to simulate landslide movement at an arbitrarily chosen point. On the other hand, there are some caveats that we have to consider to rule out possible wrong decisions. Let's suppose that we have decided to apply the k-Means method. Therefore, we have to suppose that our displacements are grouped into some number of clusters. For our study, we have chosen two clusters. Under this premise, all the measures have delivered the same partitioning. The sample of such clustering for the k-Means method and Euclidean distance measure is given in Figure **15.**

This result looks very optimistic. However, we already know that, in real, this landslide has no separate blocks and has to be treated as a whole body. Therefore, before a final decision, as many clustering methods as possible have to be tested, and their evaluation must be accomplished.

The second case is the case of retaining wall monitoring. We have a data set that contains

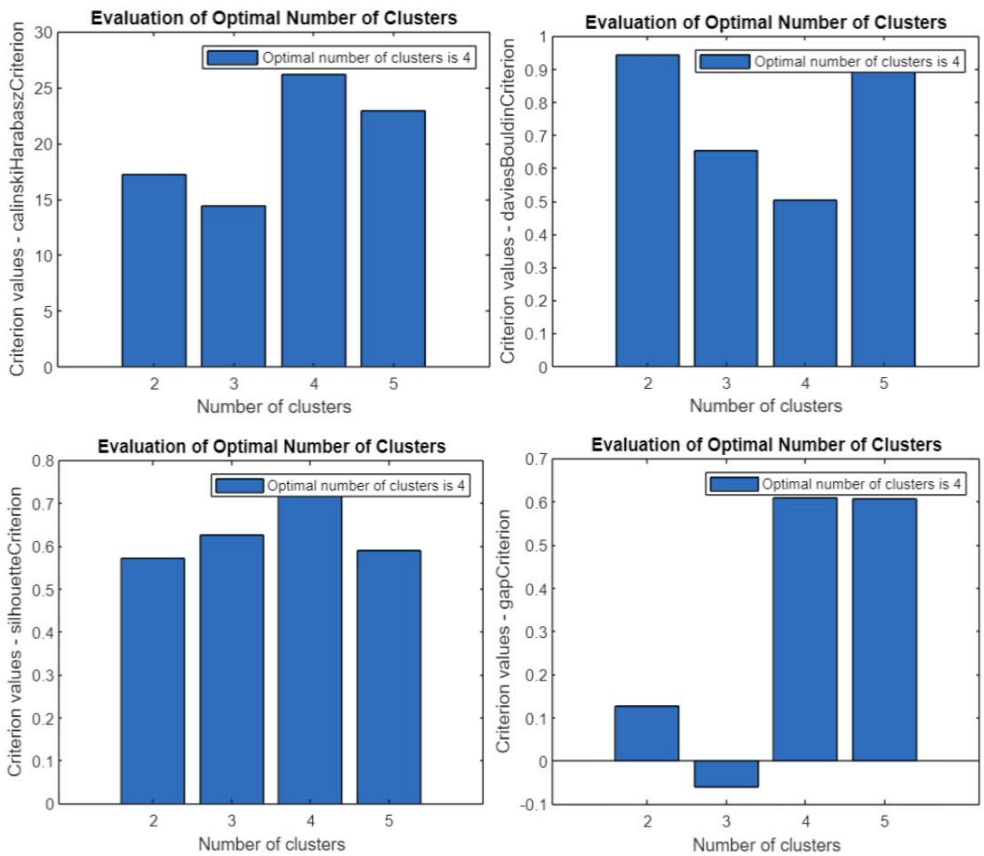


**Figure 14:** Evaluation of the optimal number of clusters.

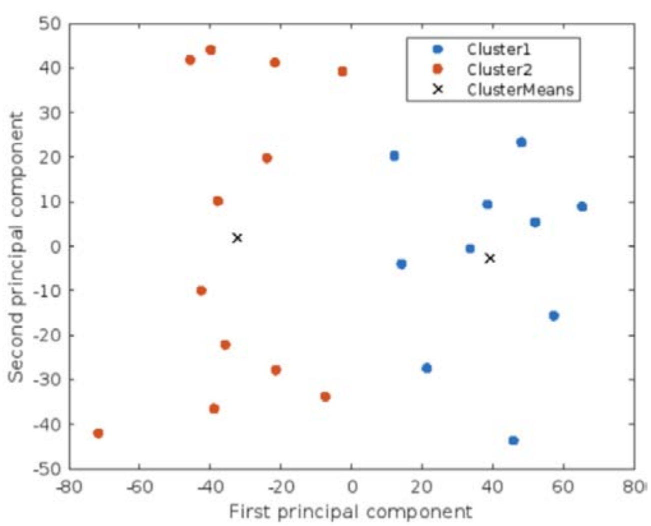


**Figure 15:** Clustering into two clusters for the k-Means method.

coordinates of deformation targets and displacements of these targets. The additional load from surrounding structures regarding the retaining wall was used as an extra parameter for improving cluster analysis (see Figure **16**).

The preliminary analysis of these displacements has shown that the displacements are grouped around two or three centers. So, as in a previous example, let's carry out the cluster analysis of this data set to determine which points belong to which cluster. In Figures **17** and **18**, there are sample results of the cluster analysis for different distance measures executed by the k-Means and nearest neighbor clustering.


**Figure 16:** The orientation of retaining walls regarding the landslide body.

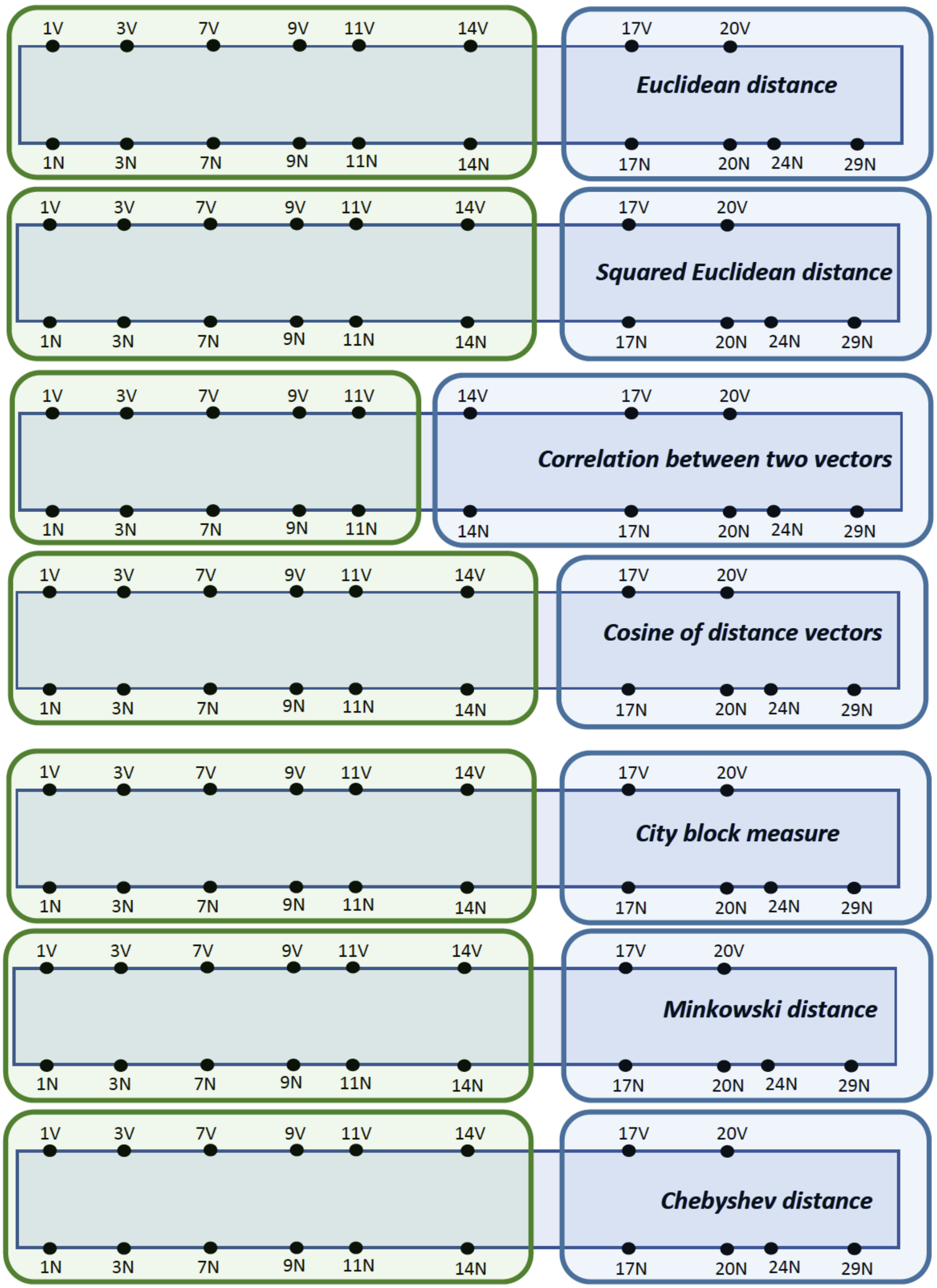


**Figure 17:** Furthest neighbor clustering for different measures.

Cluster analysis enabled identifying two blocks within which the displacements are uniform. Despite insignificant discrepancies in cosine and correlation measures, we may confidently believe that the wall is split into two separate blocks. For the first block, the average displacement is 2 mm; for the second block, 12 mm, the difference between the average values is 10 mm. For this case, the estimated accuracy of displacement determination was 2 mm. Using the well-known statistical measure for allowable accuracy of displacement determination.

$$\Delta = tm\sqrt{2}, \tag{7}$$

where $\Delta$ is an allowable measurement error, $t$ is a critical value that obeys *t*-distribution (for probability 99% and degrees of freedom 26, $t$ = 2.8), $m$ is a mean square error of measurements.

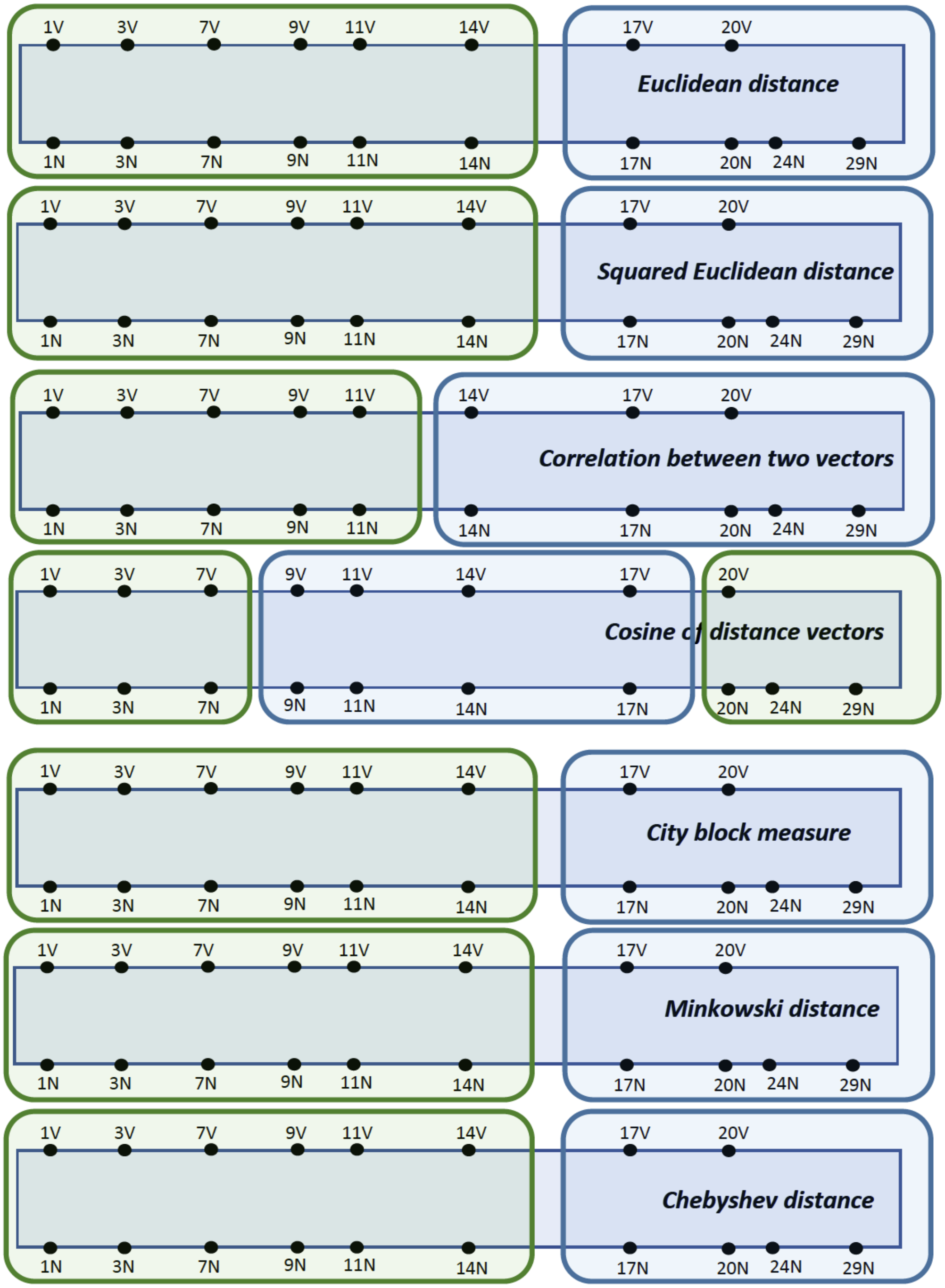


**Figure 18:** k-Means clustering for different measures.

We may conclude that the difference in average values of the two clusters exceeds the triple value of the accuracy of displacement determination of 8 mm, and therefore, it is likely that individual parts of the wall move differently.

## 5. CONCLUSIONS

The presented paper has considered one of the issues of landslide monitoring, namely the estimation of landslides' continuity. The goal was to apply cluster analysis and test its efficiency to detect the regions of uniform displacements. In general, it is known that with a large number of deformation targets and a large number of measurements, it is tough to determine the zones within which the law of displacements is uniform. The cluster analysis method can be successfully applied to solve the specified problem. The results yielded some interesting findings. The only approach proving that a landslide consists of different blocks is the application of different clustering methods. Methods that suppose a fixed number of clusters can mislead the study being that the number of clusters can be smaller or bigger than the suggested value. This is why testing various clustering approaches with different similarity measures is highly recommended. Each clustering result must be estimated by known evaluation criteria. If the estimation results differ, there is strong evidence that the clustering procedure failed, and we need an in-depth study of the particular landslide. On the other hand, when the picture is clear, all the clustering methods output similar results in general. Future studies will have to explore the efficiency of the evaluation criteria. The researchers will have to clarify which parameters are more suitable for different clustering methods of landslides and which will provide reliable results.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

[1] Schuster RL, Highland LM. Socioeconomic and environmental impacts of landslides in the western hemisphere. U.S. Geological Survey Open-File Report 01-0276, 2001. https://pubs.usgs.gov/of/2001/ofr-01-0276/

[2] Winter MG, Shearer B, Palmer D, Peeling D, Harmer C, Sharpe J. The Economic Impact of Landslides and Floods on the Road Network. Procedia Engineering 2016; 143: 1425-1434. https://doi.org/10.1016/j.proeng.2016.06.168

[3] Carlà T, Tofani V, Lombardi L, Raspini F, Bianchini S, Bertolo D, Thuegaz P, Casagli N. Combination of GNSS, satellite InSAR, and GBInSAR remote sensing monitoring to improve the understanding of a large landslide in high alpine environment. Geomorphology 2019; 335: 62-75. https://doi.org/10.1016/j.geomorph.2019.03.014

[4] Casagli N, Frodella W, Morelli S, Tofani V, Ciampalini A, Intrieri E, Raspini F, Rossi G, Tanteri L, Lu P. Spaceborne, UAV and ground-based remote sensing techniques for landslide mapping, monitoring and early warning. Geoenvironmental Disasters 2017; 4: 9. https://doi.org/10.1186/s40677-017-0073-1

[5] Gili JA, Corominas J, Rius J. Using Global Positioning System techniques in landslide monitoring. Eng Geol 2000; 55: 167-192. https://doi.org/10.1016/S0013-7952(99)00127-1

[6] Leva D, Nico G, Tarchi D, Fortuny-Guasch J, Sieber AJ. Temporal analysis of a landslide by means of a ground-based SAR Interferometer. IEEE Trans Geosci Remote Sens 2003; 41: 745-752. https://doi.org/10.1109/TGRS.2003.808902

[7] Peternel T, Kumelj Š, Oštir K, Komac M. Monitoring the Potoška planina landslide (NW Slovenia) using UAV photogrammetry and tachymetric measurements. Landslides 2017; 14: 395-406. https://doi.org/10.1007/s10346-016-0759-6

[8] Scaioni M, Longoni L, Melillo V, Papini M. Remote Sensing for Landslide Investigations: An Overview of Recent Achievements and Perspectives. Remote Sens 2014; 6: 9600-9652. https://doi.org/10.3390/rs6109600

[9] Auflič ME, Herrera G, Mateos RM, Poyiadji F, Quental L, Severine B, Peternel T, Podolszki L, Calcaterra S, Kociu A, Warmuz B, Jelének J, Hadjicharalambous C, Becher GP, Dashwood C, Ondrus P, Minkevičius V, Todorović S, Møller JJ, Marturia J Landslide monitoring techniques in the Geological Surveys of Europe. Landslides 2023; 20: 951-965. https://doi.org/10.1007/s10346-022-02007-1

[10] Brückl E, Brunner FK, Kraus K. Kinematics of a deep-seated landslide derived from photogrammetric, GPS and geophysical data. Eng Geol 2006; 88: 149-159. https://doi.org/10.1016/j.enggeo.2006.09.004

[11] Qiao G, Lu P, Scaioni M, Xu S, Tong S, Feng T, Wu H, Chen W, Tian Y, Wang W, Li R. Landslide investigation with remote sensing and sensor network: From susceptibility mapping and scaled-down simulation towards in situ sensor network design. Remote Sens 2015; 5: 4319-4346. https://doi.org/10.3390/rs5094319

[12] Chae B-G, Park H-J, Catani F, Simoni A, Berti M. Landslide prediction, monitoring and early warning: a concise review of state-of-the-art. Geosci J 2017; 21: 1033-1070. https://doi.org/10.1007/s12303-017-0034-4

[13] Zheng H, Qian G, Tordesillas A. Modelling High-Dimensional Time Series with Nonlinear and Nonstationary Phenomena for Landslide Early Warning and Forecasting. Eng Proc 2023; 39: 21. https://doi.org/10.3390/engproc2023039021

[14] Lin Z, Ji Y, Sun X. Landslide Displacement Prediction Based on CEEMDAN Method and CNN–BiLSTM Model. Sustainability 2023; 15: 10071. https://doi.org/10.3390/su151310071

[15] Song Y, Yang D, Wu W, Zhang X, Zhou J, Tian Z, Wang C, Song Y. Evaluating Landslide Susceptibility Using Sampling Methodology and Multiple Machine Learning Models. ISPRS Int J Geo-Inf 2023; 12: 197. https://doi.org/10.3390/ijgi12050197

[16] Jaboyedoff M, Couture R, Locat P. Structural analysis of Turtle Mountain (Alberta) using digital elevation model: toward a progressive failure. Geomorphology 2009; 103: 5-16. https://doi.org/10.1016/j.geomorph.2008.04.012

[17] Haberler-Weber M. Analysis and interpretation of geodetic landslide monitoring data based on fuzzy systems. Nat Hazards Earth Syst Sci 2005; 5: 755-760. https://doi.org/10.5194/nhess-5-755-2005

[18] Highland LM, Bobrowsky P. The landslide handbook—A guide to understanding landslides: Reston, Virginia, U.S. Geological Survey Circular 2008; 1325: 129. https://pubs.usgs.gov/circ/1325

[19] Chen T-HK, Prishchepov A V, Fensholt R, Sabel CE. Detecting and monitoring long-term landslides in urbanized areas with nighttime light data and multi-seasonal Landsat imagery across Taiwan from 1998 to 2017. Remote Sens Environ 2019; 225: 317-327. https://doi.org/10.1016/j.rse.2019.03.013

[20] Mateos RM, López-Vinielles J, Poyiadji E, Tsagkas D, Sheehy M, Hadjicharalambous K, Liscák P, Podolski L, Laskowicz I, Iadanza C, Gauert C, Todorović S, Auflič MJ, Maftei R, Hermanns RL, Kociu A, Sandić C, Mauter R, Sarro R, Béjar M, Herrera G. Integration of landslide hazard into urban planning across Europe. Landsc Urban Plan 2020; 196: 103740. https://doi.org/10.1016/j.landurbplan.2019.103740

[21] Frattini P, Crosta GB, Allievi J. Damage to buildings in large slope rock instabilities monitored with the PSInSARTM Technique. Remote Sens 2023; 5: 4753-4773. https://doi.org/10.3390/rs5104753

[22] Malamud BD, Turcotte DL, Guzzetti F, Reichenbach P. Landslide inventories and their statistical properties. Earth Surf Processes 2004; 29: 687-711.

[23] Gullà G, Calcaterra S, Gambino P, Borrelli L, Muto F. Long-term measurements using an integrated monitoring network to identify homogeneous landslide sectors in a complex geo-environmental context (Lago, Calabria, Italy). Landslides 2018; 15: 1503-1521. https://doi.org/10.1007/s10346-018-0974-4

[24] Soeters R, van Westen CJ. Slope instability recognition, analysis, and zonation. In Landslides: Investigation and Mitigation; Transportation Research Board Special No. 247; Turner AK, Schuster RL, Eds., National Academy Press: Washington, DC, USA, 1996; pp. 129-177.

[25] Lu P, Casagli N, Catani F, Tofani V. Persistent scatterers interferometry hotspot and cluster analysis (PSI-HCA) for detection of extremely slow-moving landslides. Int J Remote Sens 2012; 33: 466-489. https://doi.org/10.1080/01431161.2010.536185

[26] Everitt BS, Landau S, Leese M. Cluster Analysis, Fourth edition, Arnold 2001.

[27] Manly BFJ. Multivariate Statistical Methods: A primer, Third edition, Chapman and Hall 2005.