Preparation of Complex Landslide Simulation Results with Clustering Approaches for Decision Support and Early Warning

Eva Nuhn
Geoinformatics Research Group (AGIS), Institute of Applied Computer Science,
University of the Bundeswehr Munich, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
eva.nuhn@unibw.de

Erik Kropat
Institute for Theoretical Computer Science, Mathematics and Operations Research,
University of the Bundeswehr Munich, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
erik.kropat@unibw.de

Wolfgang Reinhardt
Geoinformatics Research Group (AGIS), Institute of Applied Computer Science,
University of the Bundeswehr Munich, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
wolfgang.reinhardt@unibw.de

Stefan Pickl
Institute for Theoretical Computer Science, Mathematics and Operations Research,
University of the Bundeswehr Munich, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
stefan.pickl@unibw.de

Abstract

Recent catastrophic landslide events caused large human and material damages. This shows again, that there is still not enough protection against such kinds of natural hazards. In an ongoing research project a new approach to improve early warning systems of landslides is pursued: complex simulations of landslides are coupled with geoinformation systems (GIS). This allows on the one hand for a detailed investigation of unstable slopes with the help of the simulation and on the other hand for a user friendly preparation of the complex simulation results in the GIS for decision support. In this paper the interconnection between a GIS and a simulation system is briefly introduced and the main focus is put on cluster methods for the processing of the complex simulation results in a suitable way to support the user in the decision support and in the learning system, respectively.

1. Introduction

Recent catastrophic landslide events caused a large human and material dam-age. The landslide in Angras dos Reis in Brazil in January 2010, which killed 74 people [15], demonstrated again, that there is still not enough protection against such kinds of natural hazards. In order to advance research in the field of early warning systems of land-slides the joint project "Development of suitable Information Systems for Early Warning Systems" was launched. The project aims at the development of components of an information system for the early recognition of landslides, their prototypical implementation and evaluation [6]. One subproject of the joint project addresses the coupling of complex finite element (FE) simulations with geoinformation systems [16,20]. In the project numerical simulations are setup to calculate the stability of slopes and to improve the understanding of the causes of slope instability and triggers of ground failure. This allows for the evaluation of unstable slopes and their imminent danger for humans and infrastructures. Because of the complex procedure of performing a simulation and the bulky simulation results such numerical simulations are mostly used by experts and scientists. For disaster prevention and management they are currently not available in practice, but would obviously be very helpful. The interconnection of the simulation with the geoinformation system (GIS) enables a broader use of such a complex system as the handling of it becomes more intuitive and user-friendly and GIS methods can

978-0-7695-4525-7/12 $26.00 © 2012 IEEE
DOI 10.1109/HICSS.2012.484

be used to prepare the input data as well as to make the bulky results usable.

In this paper the interconnection between a GIS and a simulation system is introduced and two operational modes of the coupled system are presented: the learning and the decision support system. The main focus will be put on cluster methods for processing of the complex simulation results in a suitable way to support the user in the decision support and in the learning system, respectively.

2. Application area

The methods developed in the framework of this project are evaluated on the basis of manifold data sets and specific application scenarios. Therefore as suitable application area a part of the slopes in the Isar valley in the south of Munich, Germany, has been selected in cooperation with the Bavarian Authority of Environment (see figure 1). In this area the Isar eroded a deep valley into layers of quaternary gravels and tertiary sediments of partially high plasticity [21]. Consequently, steep and unstable slopes were formed, where landslides occurred from time to time [2].

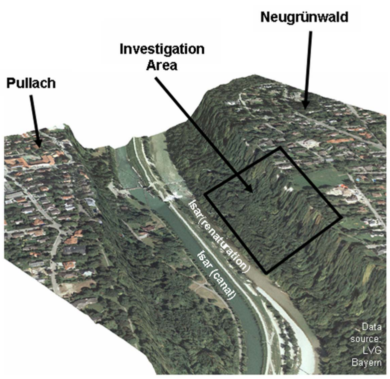

Figure1. Application Area

In the application area, the height difference of the slope reaches up to almost 40 meters and the potentially endangered infrastructure is located close to the edge of the slope. In the early and the mid-seventies there have been several landslides in this area. As a reaction to these events and because of the risk potential several sensors (extensometer, inclinometer and ground water level tubes) were installed and geodetic measurements were initiated by the Bavarian Authority of Environment. Further, the soil layers of the slope were investigated through numerous outcrops and boreholes. Today, after more than thirty years of investigations, extensive knowledge of the subsoil structure and the failure mechanism are available and can be used in the present project.

3. Operational modes and data flow of the coupled system

Complex simulations are computationally intensive and may be, in case of a rapidly required decision for early warning, too time consuming. Therefore two main operational modes of the coupled system have been identified: the use as decision support system for prevention or reaction to a hazardous event, which requires very fast results, and the use as learning system for better understanding and prognosis of landslide movements. The use as decision support system was designed within this project in principle, but for a usage by decision makers further extension would be necessary. The main focus was put on the learning system mode, which is provided for experts (e.g. geologist, geotechnical engineers), and which additionally supports the assessment and further GIS based analysis of the simulation results [16]. In this mode normally many simulations are performed and can be stored in the data base. In figure 2 the architecture of the coupled simulation and geoinformation system is shown. First, the required input data for the analysis is selected and prepared in the GIS. These parameters describe basically the geometry and the subsoil structure of the slope, the so called impacting loads and several boundary conditions. The impacting loads define the event, which influence to the stability of the slope shall be estimated by means of the simulation. For example a rainfall event may destabilize the slope and cause a potential danger.

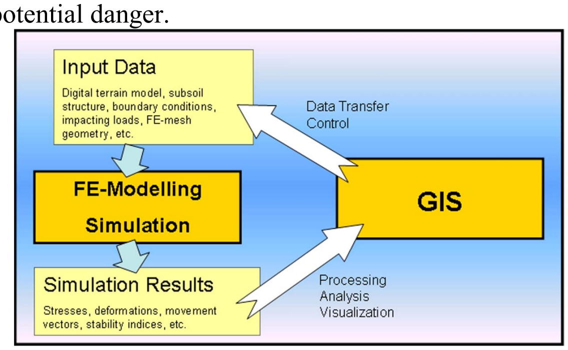

Figure 2. Architecture of the coupled system

The transfer of the input data to the simulation system is controlled by the GIS. The stability and deformation of the slope is investigated within the simulation system by application of the FE-method [20]. Therefore a FE-mesh is needed. An example of a FE-mesh for a 2D-simulation is shown in figure 3.

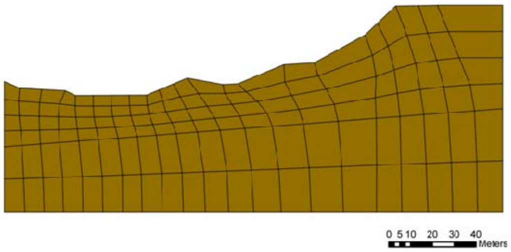

Figure 3. Example of a FE-mesh

After all information required for the FE-analysis has been collected, the data is put together in a single input file which then is used as the basis for computation (within the simulation). During the computation, the defined loads are applied incrementally on the slope and their influence on the slope stability is determined. The result of the simulation is a bulky field of vectors which indicate the instability of the slope. These vectors are transferred to the GIS for processing into a form which is understandable for decision makers and also more handsome for experts.

4. Preparation of complex simulation results with clustering approaches

Results of the simulations are vectors, which represent e.g. stresses, strains or deformation. In figure 4 deformation vectors are visualized. To identify the most important outcome of the simulation, namely the deformation direction and the length of the deformation vectors, the depiction has to be strongly enlarged. But therewith the overview of the slope situation as a whole will be lost. The problem is that the simulation results (vectors at mesh nodes) are too complex and confusing to be presented as a base for decision support (fig. 4).

Figure 4. Visualized 3D simulation results

A first step to reduce the complexity of the simulation results is the division of the slope into an area where significant deformations occurred (sliding body) and in an area with no or minor deformations by means of a given threshold value which is used to separate these areas. In figure 5 a section through the slope with the identified sliding body is shown. But on the basis of this depiction it is not possible to distinguish between regions with similar deformation characteristics indicated by the direction and length of the vectors.

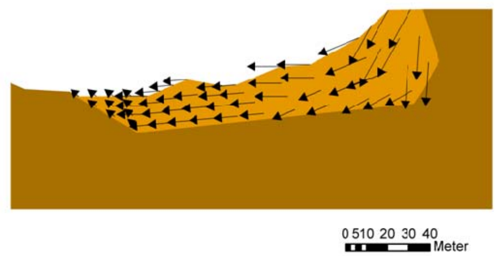

Figure 5. Deformation

5. Clustering approach

To prepare the complex simulation results a clustering algorithm is used. Cluster algorithms are statistical classification techniques for dividing a population into homogenous groups (“a cluster”). The similarities between members belonging to a cluster are high, while similarities between members belonging to different clusters are low [22]. In this study the single linkage algorithms was used [19]. This agglomerative hierarchical clustering approach was investigated in earlier studies and has performed best to group the simulation results [18]. Within this method at the beginning each object (in this case FE-nodes) is placed in a separate cluster, and at each step the closest pair of clusters is merged until a certain termination condition is satisfied. To decide which objects have to be merged a notion of cluster proximity has to be defined. In the single linkage, the proximity of two clusters is defined as the minimum of the distance between any two points within the two clusters. As distance function the widely used Euclidian distance is used. In the formula three features, which describe the to-be-clustered deformation vectors, are considered:

$$d = [(X_1-X_2)^2 + (Y_1-Y_2)^2 + (L_1-L_2)^2 + (R_1-R_2)^2]^{1/2}$$

- X, Y: Location of the FE-node,
- L: Deformation length and
- R: Deformation direction of the deformation vector.

The features have different ranges of values. To make them comparable a standardization function is used. An often used aid for this purpose is the z-transformation, which converts the values of a sample into *z-scores* [13]: The z-scores are calculated by subtracting the features mean (M) from the original value of the feature (x_i) and dividing the result by the standard deviation of the feature (s), according to the formula:

$$z_i = (x_i - M) / s.$$

A major issue in clustering is the determination of the appropriate number of clusters. Several methods have been proposed for this issue. In this study, the stopping rule from Mojena [14] is used. To cluster the deformation vectors a multivariate single-linkage cluster analysis is carried out. That means, all features (location, deformation length and direction) are considered in the distance function. The result of the multivariate cluster analysis in conjunction with the stopping rule from Mojena is shown in figure 6.

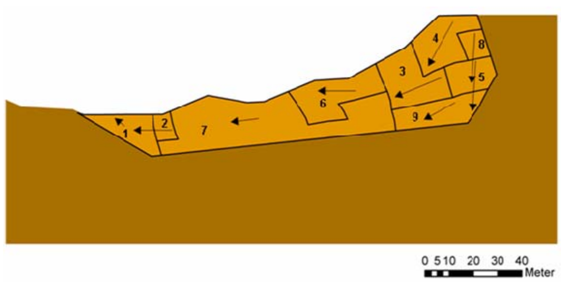

Figure 6. Result of the multivariate cluster analysis

Nine clusters have been determined. The deformation vectors in the clusters have been aggregated to one single deformation vector representing the vectors of theses cluster and the boundaries for the clusters have been determined according to [17]. But if we have a detailed look at some clusters, some malformations can be identified (see figure 7). First there are deformation vectors with different lengths and different directions, which belong to one cluster (e.g. cluster 7). Second there are deformation vectors with similar lengths and similar directions, which are splitted in two clusters (cluster 5 and 8).

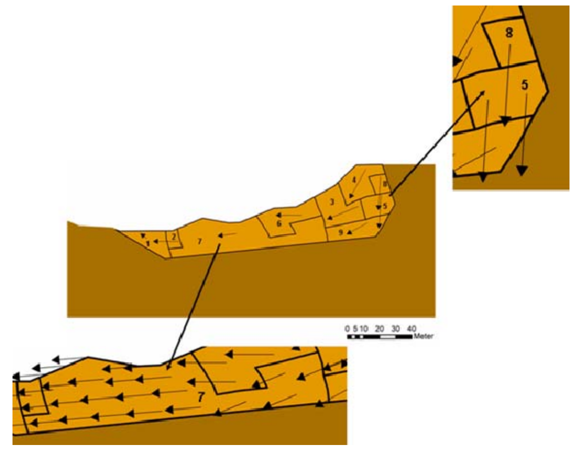

Figure 7. Malformations of the multivariate cluster analysis

Theoretical investigations have shown that these malfunctions are caused by the fact that the effect of the position within the cluster determination process is too large. Two of the four summands in the distance function are depending on the position. Because the deformation vectors in cluster 7 are positioned close to each other they are aggregated to one cluster although they have slightly different directions and lengths. In cluster 5 and 8 the deformation vectors are far apart, therefore they are splitted in two clusters. These examples have shown that the algorithm used is not suitable to cluster areas with vectors of similar lengths and directions as intended. To overcome this problem several other approaches have been selected. As one possible solution first a double bivariate cluster analysis was tested. This procedure starts with a first cluster analysis which considers only direction and length of the deformation vectors in the distance function. The following second bivariate cluster analysis considers only the position in the distance function. The result of the first bivariate cluster analysis is shown in figure 8.

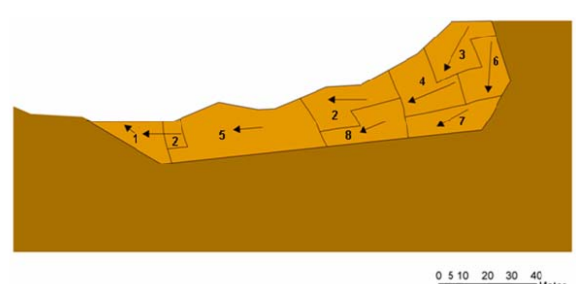

Figure 8. Result of the first bivariate cluster analysis

The cluster 7 from the result of the multivariate cluster analysis is now divided in two clusters (cluster 5 and cluster 8). And the former clusters 5 and 8 are now assembled to one cluster (cluster 6). But in this step another problem arises. With this bivariate cluster analysis, clusters are produced, which aggregate deformation vectors, which are not spatially adjacent to each other (cluster 2). This methodology now has to be extended to be able to overcome this problem and to separate these clusters. The result of the second bivariate cluster analysis, which now considers only the position is shown in figure 9. Like expected the cluster 2 is divided correctly in two subclusters (cluster 2_1 and 2_2). But with this second bivariate cluster analysis also other clusters (e.g. cluster 1) are divided in subclusters. Because this approach leads not to the desired results the approach of double bivariate cluster analysis was rejected.

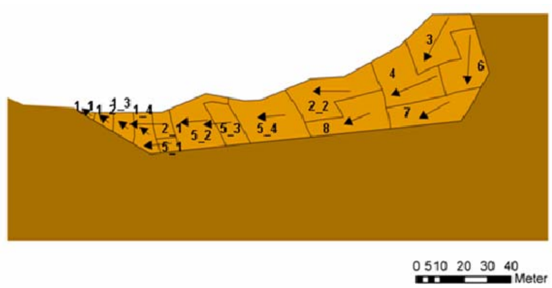

Figure 9. Result of the second bivariate cluster analysis

As another possible solution a bivariate cluster analysis with a subsequent investigation of the neighbourhood relations was tested. First the cluster analysis which considers only direction and length of the deformation vectors in the distance function is again carried out. The result is shown in figure 8. Subsequent to that it is investigated for each cluster, if the aggregated deformation vectors are spatially adjacent. Therefore the FE-mesh is used, which consists of a collection of nodes and edges, which defines the finite elements. The edges define the neighbourhoods between the nodes of the mesh and their corresponding deformation vectors. As a result of the neighbourhood analysis the cluster 2 from figure 8 is subdivided correctly into two subclusters (cluster 2_1 and 2_2). The other clusters remain the same, because all aggregated deformation vectors are spatially adjacent (see figure 10).

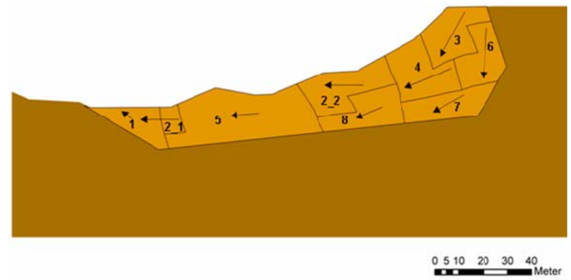

Figure 10. Result of the bivariate cluster analysis with subsequent examination of the neighbourhood

The results of this method are clusters, with aggregated deformation vectors with small, moderate and large deformation lengths. These clusters can be visualized according to the deformation length of the corresponding deformation vector, to support users in the learning and the decision support system, respectively (see figure 11).

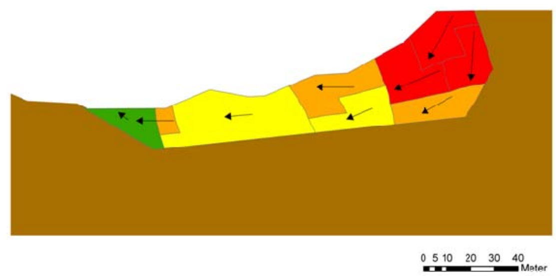

Figure 11. Classified deformation areas

6. Fuzzy Clustering Approach

For a user of the decision support system the division of all the FE-nodes to sharply separated clusters where each node is member of exactly one cluster provides a good basis for the determination of the deformation areas of the slope and is in most cases sufficient. For a user of the learning system this information is often not satisfactory. In particular, for the assessment and evaluation of the complex simulation results further information, like uncertainties, should be recognizable, in order to allow for a validation. Therefore a fuzzy clustering approach is used as another approach. In this approach the FE-node can belong to more than one cluster. The membership of a node to a cluster is defined on a scale from zero to one, where zero is no membership and one is full membership. The results are clusters with boundaries, which have vague or indeterminate locations, or which are gradual transitions between two zones [22]. In this study the prominent fuzzy c-means approach [5,11] is used. This method minimizes the intra-cluster variance, but like many other partitional clustering algorithms, fuzzy c-means requires the number c of clusters as an input variable. Once a fuzzy partition is determined by fuzzy c-means, the user has to decide whether or not it accurately presents the data structure. Various cluster validity indexes have been proposed to evaluate the quality and fitness of partitions produced by the clustering algorithm [24]. However, validity indexes are considered to be independent of the clustering algorithm and standard measures like the partition coefficient [3,4], modified partition coefficient [7] and partition entropy [3,4] are not directly related to the geometrical structure of the data. As different validity indexes can lead to variations in the number of clusters, the choice of an appropriate validity measure becomes an important issue. Fukuyama and Sugeno [10] proposed a validity function that measures compactness and separation of clusters in a noisy environment simultaneously. The evaluation of the fuzzy-c-means algorithm for different choices of c and a subsequent evaluation of the validity function enables the user to determine the optimal number of clusters with the best clustering performance.

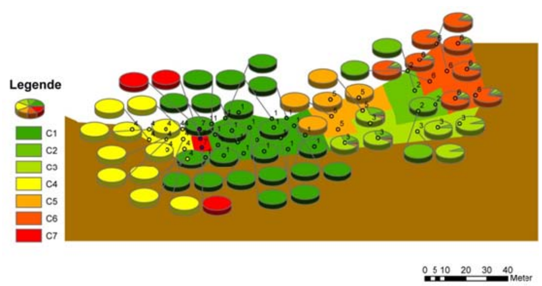

Figure 12. Fuzzy partition

For comparative studies, a fuzzy cluster analysis is performed with regard to the bivariate data from section 5 where the direction and the length of the deformation vectors are considered. Subsequent to that it is investigated for each cluster, if the aggregated deformation vectors are spatially adjacent. Figure 12 shows the results obtained with fuzzy c-means and subsequent neighbourhood examination. As each node can now belong to more than one cluster, the corresponding degree of membership is depicted by a pie chart. When each node is assigned to the cluster with maximal degree of membership, the underlying cluster partition can be revealed. In addition, cluster uncertainties can be visualized (cf. figure 13) offering more detailed information about the internal structure of clusters and the quality of cluster separation.

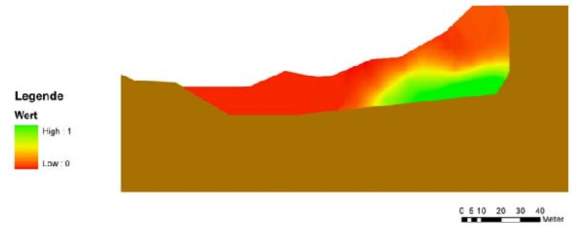

Figure 13. Fuzzy cluster

The cluster analysis performed in this study focussed on four features (location of the FE-node, deformation length and deformation direction of the deformation vector). In future studies, additional geophysical features will be available that allow for a deeper analysis and a more detailed description of clusters. For this reason, each region (cluster) is expected to be characterized by a specific group of features. Several methods have been proposed for *feature selection*, usually assigning binary relevance weights to the features to indicate whether or not a particular feature is considered important [1,12,23]. Such an approach is often applied to reduce the dimensionality of the problem by completely eliminating irrelevant features, but it is not suffcient here, because it does not reflect the relation between clusters and features. This relation is usually established by *feature weighting*, which assigns different feature subsets to distinct clusters. As some features might be useful but less important with regard to a specific cluster, *continuous feature weighting* can be applied as a further extension of feature selection in order to obtain a more balanced and cluster-dependent representation of feature relevance in an uncertain environment. Fuzzy c-means, like other clustering algorithms, does not provide cluster dependent feature weights. Any given feature must either be used (completely relevant) or ignored (irrelevant) for all clusters. Frigui and Nasraoui [9] pro-posed an extension of fuzzy c-means called *Simultaneous Clustering and At-tribute Discrimination* (SCAD) that performs clustering and feature weighting simultaneously. This algorithm learns the feature relevance for each cluster independently and it uses continuous feature weighting and so it provides much richer features relevance representation. SCAD can adapt to the variations that exist within the data set by categorizing it into distinct clusters, which are allowed to overlap because of the use of fuzzy membership degrees. This algorithm can be further extended and combined with *competitive agglomeration* [8] in order to determine the number of clusters automatically [9].

We tested SCAD to the simulation results with regard to the features position, deformation length and deformation direction. As number of c of clusters we used 8 and 9. This matches the number of clusters determined with the hard clustering approach. Nine clusters with subsequent examination of the neighbourhood and eight clusters without. The use of eight and nine clusters leads to the cluster partitions depicted in figure 14 and figure 15. Since SCAD obviously generates the same cluster structure as the bivariate cluster analysis with and without subsequent examination of the neighbourhood, the results obtained with hard clustering become validated.

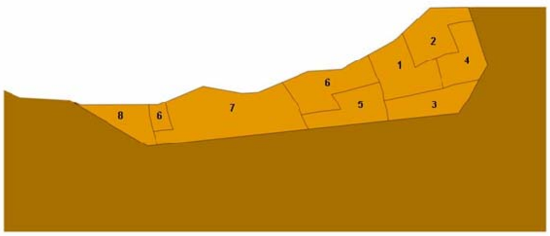

Figure 14. Fuzzy clusters based on SCAD

In case of the of nine clusters the position turns out to be an irrelevant feature. Deformation length and deformation direction mainly influence the assignment of objects and, thus, the shape of clusters. Table 1 shows the feature weights of the different clusters.

Cluster	Deformation length	Deformation direction
1	1.0000	0.0000
2	0.5813	0.4187
3	0.9961	0.0039
4	0.7763	0.2237
5	0.9998	0.0002
6	0.0002	0.9998
7	1.0000	0.0000
8	0.6765	0.3235
9	1.0000	0.0000

Table 1. Feature weights of the different clusters

Deformation length is the main feature for most clusters. Clusters 2,4,8 show a rather balanced influence of deformation length and direction. Only cluster 6 strongly depends on the deformation direction.

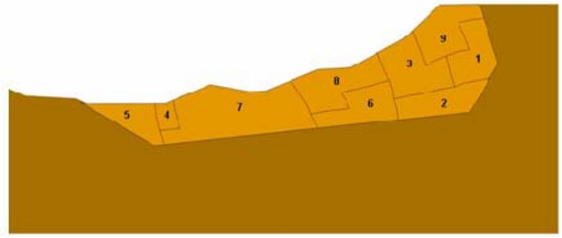

Figure 15. Fuzzy clusters based on SCAD

7. Conclusion

In this paper a coupled simulation and geoinformation system has been introduced. The main focus was put on cluster methods to prepare the complex simulation results in such a way that they are understandable and easy to use for decision makers in the learning and the decision support system respectively. Therefore a hard clustering approach was used, which divides all deformation vectors to exactly one cluster. Further a fuzzy clustering approach was used to allow for an evaluation of the complex simulation results (cf. table 1 for comparison of the applied cluster techniques). Both cluster algorithms deliver good results, which are comparable to each other. While the hard clustering approach offers information about the deformation areas the fuzzy clustering approach offers more detailed information about the quality of the cluster separation and the incorporated uncertainties. Hard and fuzzy algorithms complement and validate one another in identifying clusters which can be used for learning and decision support with respect to landslide analysis.

Table 1. Overall clustering approach

Cluster technique	Results
Single linkage with Mojena stopping rule	Not suitable to areas with vectors of similar lengths and directions as intended
Double bi-variate cluster analysis	Clusters not spatially adjacent
Fuzzy c-means clustering	Confirms hard clustering

8. Acknowledgement

The funding of the research project "Development of suitable information systems for early warning systems" by the German Ministry of Education and Research (BMBF) under grant no. 03G0645A et al. within the framework of the GEOTECHNOLOGIEN initiative (http://www.geotechnologien.de) is gratefully acknowledged.
This is publication no. GEOTECH-1316 of the R&D-Programme GEOTECHNOLOGIEN.

9. References

[1] H. Almuallim and T.G. Dietterich, "Learning with many irrelevant features". In: Ninth National Conference on Artificial Intelligence, 1991, pp. 547-552.

[2] H.J. Baumann, "Bruchvorgänge in Folge der Isareintiefung südlich Münchens und die kritischen Höhen der Talhänge", Schriftenreihe des Lehrstuhls und Prüfamt für Grundbau, Bodenmechanik und Felsmechanik. Technische Universität München. 1988.

[3] J.C. Bezdek, "Cluster validity with fuzzy sets", J. Cybernet. 3, 1974a, pp. 58-73.

[4] J.C. Bezdek, "Numerical taxonomy with fuzzy sets", J. Math. Biol. 1, 1974b, pp. 57-71.

[5] J.C. Bezdek, R. Ehrlicher, and W. Full, "FXM The fuzzy c-means clustering algorithm", Computers & Geosciences, 10, 2-3, 1984, pp. 191-203.

[6] M. Breunig, B. Schilberg, P.V. Kuper, M. Jahn, W. Reinhardt, E. Nuhn, S. Mäs, C. Boley, F.-X. Trauner, J. Wiesel, D. Richter, A. Abecker, D. Gallus, W. Kazakos, and A. Bartels, "EGIFF-Developing Advanced GI Methods for Early Warning in Mass Movement Scenarios",. In: GEOTECHNOLOGIEN Science Report - Early Warning Systems in Earth Management, No. 13, 2009, pp. 49-72.

[7] R.N. Dave, "Validating fuzzy partition obtained through c-shells clustering", Pattern Recognition Lett. 17, 1996, pp. 613-623.

[8] H. Frigui and R. Krishnapuram, "Clustering by competitive agglomeration", Pattern Recognition, 30, 7, 1997, pp. 1109-1119.

[9] H. Frigui and O. Nasraoui, "Unsupervised learning of prototypes and attribute weights", Pattern Recognition, 37, 2004, pp. 567-581.

[10] Y. Fukuyama and M. Sugeno, "A new method of choosing the number of clusters for the fuzzy c-means method", Proceedings of fifth Fuzzy Syst. Symp., 1989, pp. 247-250.

[11] F. Höppner, F. Klawonn, R. Kruse, and T. Runkler, "Fuzzy Cluster Analysis: Methods for Classification Data Analysis and Image Recognition", Wiley, New York 1999.

[12] K. Kira, and L.A. Rendell, "The feature selection problem: trditional methods and a new algorithm", in: Tenth National Conference on Artificial Intelligence, 1192, pp. 129-134.

[13] L.S. Meyers, G. Gamst, and A.J. Guarino, "Applied multivariate research: design and interpretation", Sage Publications, London, 2006.

[14] R. Mojena, "Hierarchical grouping methods and stopping rules: an evaluation", The Computer Journal 20(4), 1977, pp. 359-363.

[15] NTDTV, New Tang Dynasty Television - Mudslide Rescue Efforts Continue in Brazil (2009) http://english.ntdtv.com/ntdtv_en/ns_sa/2010-01-06/792541539031.html, (accessed 05.01.2010).

[16] E. Ortlieb, W. Reinhardt, and F.-X. Trauner, "Developement of a coupled geo information and simulation system for early warning systems", Proceedings from the Cartographie and Geoinformatics for Early Warning and Emergency Management: Towards better solutions Conference. Prague, Czech Republic, 2009a, pp. 483-492.

[17] E. Ortlieb, W. Reinhardt, and F.-X. Trauner, "Kopplung von Geoinformation- und Simulationssystemen zur Entscheidungsunterstützung", Mitteilungen des Bundesamtes für Kartographie und Geodäsie Band 43, Verlag des Bundesamtes für Kartographie und Geodäsie, Frankfurt am Main, Germany, 2009b, pp. 97-106.

[18] M. Reus, "Aufbereitung von komplexen Simulations-ergebnissen mit Methoden der Clusteranalyse", Institute for Applied Computer Science, University of the Bundeswehr Munich. 2009.

[19] P.H.A. Sneath, "The Application of Computers to Taxonomy", Journal of General Microbiology 17, 1957, pp. 201-226.

[20] F.-X., Trauner, and C. Boley, "Application of geotechnical models for early warning systems to mitigate threats by landslides", Proceedings from the 17th International Conference on Soil Mechanics & Geotechnical Engineering. Alexandria, Egypt (2009).

[21] F.-X., Trauner, C. Boley, and E. Ortlieb, "Neue innovative Methoden zur Frühwarnung auf Basis von gekoppelten Informations- und Simulationssystemen am Beispiel gravitativer Massenbewegungen", Proceedings des 6. Kolloquium Bauen in Boden und Fels. Esslingen, Germany, 2008, pp. 433-439.

[22] T. Wade, and S. Sommer, "A to Z GIS", ESRI Press. Redlands, California. ISBN 9781589481404, 288, 2006.

[23] D. Wetterschreck, D.W. Aha, and T. Mohri, "A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms", Artif. Intell. Rev., 11, 1997, pp. 273-314.

[24] K.-L. Wu, and M.-S. Yang, "A cluster validity index for fuzzy clustering", Pattern Recognition Letters 12, 2005, pp. 1275-1291.