UNCERTAINTY HANDLING IN DISASTER MANAGEMENT USING HIERARCHICAL ROUGH SET GRANULATION

H. Sheikhian[a], M.R. Delavar[b] and A. Stein[c]

[a] MSc. Student, GIS Dept., School of Surveying and Geospatial Eng., College of Eng., University of Tehran, Tehran, Iran, sheikhain@ut.ac.ir
[b] Center of Excellence in Geomatic Eng. in Disaster Management, School of Surveying and Geospatial Eng., College of Eng., University of Tehran, Tehran, Iran, mdelavar@ut.ac.ir
[c] Department of Earth Observation Science, University of Twente, The Netherlands, a.stein@utwente.nl

Commission II, WG II/4

KEY WORDS: Seismic vulnerability assessment, Granular computing, Rough set theory, Uncertainty management

ABSTRACT:

Uncertainty is one of the main concerns in geospatial data analysis. It affects different parts of decision making based on such data. In this paper, a new methodology to handle uncertainty for multi-criteria decision making problems is proposed. It integrates hierarchical rough granulation and rule extraction to build an accurate classifier. Rough granulation provides information granules with a detailed quality assessment. The granules are the basis for the rule extraction in granular computing, which applies quality measures on the rules to obtain the best set of classification rules. The proposed methodology is applied to assess seismic physical vulnerability in Tehran. Six effective criteria reflecting building age, height and material, topographic slope and earthquake intensity of the North Tehran fault have been tested. The criteria were discretized and the data set was granulated using a hierarchical rough method, where the best describing granules are determined according to the quality measures. The granules are fed into the granular computing algorithm resulting in classification rules that provide the highest prediction quality. This detailed uncertainty management resulted in 84% accuracy in prediction in a training data set. It was applied next to the whole study area to obtain the seismic vulnerability map of Tehran. A sensitivity analysis proved that earthquake intensity is the most effective criterion in the seismic vulnerability assessment of Tehran.

1. INTRODUCTION

Multi-criteria decision making problems are dealing with various criteria obtained from different sources as well as different data integration methods. Input data are one of the major sources of uncertainty in such a problem and thus data quality has a significant effect on the acquired results. The implemented model, however, may not be sufficiently accurate and its uncertainty has to be considered. In cases where expert judgment is involved in the decision making process, imprecise expert decisions may also affect the acquired results. Due to these uncertainties, multi-criteria decision making methods have to be assessed against the uncertainty sources.

A number of studies have focused on uncertainty management, such as Hansson (1996) and Zadeh (2005). Much research was also dedicated to manage uncertainty in multi-criteria problems (Paté-Cornell, 1996; Baker, 2008; Tesfamariam et al., 2010). In Tehran, researchers such as Jahanpeyma et al. (2007) and Khamespanah et al. (2013b) integrated uncertainty management with seismic vulnerability mapping.

In this paper, hierarchical rough set granulation is integrated with granular computing (GrC) for rule extraction to undertake quality assessment in extracting classification rules. Hierarchical rough set granulation provides quality measures for the granules describing the data set. Those measures are used to select a set of qualified granules to be used during the rule extraction by granular computing, which applies measurements between the granules and output classes to obtain classification rules.

Granular computing is adequate to handle uncertainty in classification procedures (Zadeh, 1997).

In this research, seismic physical vulnerability of Tehran, the capital of Iran, was examined. First, correlation of the selected criteria was investigated to ensure that the model is provided with non-redundant data. Next, granules were extracted by hierarchical rough set granulation and were provided to granular computing to extract the classification rules. To assess the model, statistical tests and a sensitivity analysis were carried out.

2. METHODOLOGY

This paper proposes hierarchical rough set granulation to extract granules to be used in the rule extraction using granular computing. Instead of classical granular computing algorithm which extracts the possible granules of the data and then attempts to extract the classification rules, we proposed a new model which uses hierarchical rough granulation for extracting best describing granules to be used in the process of granular computing rule extraction in order to overcome the problem of uncertainty propagation in the rule extraction procedure. An overview of the proposed methodology is provided in the succeeding sections.

2.1 Granulation by rough set approximation

Rough sets theory was introduced by Pawlak (1982). It is widely accepted as an excellent tool for handling vaguely described objects. It allows dealing with input data inconsistencies (Greco et al., 2001), while providing *a posteriori* information about the quality of the approximation.
Rough set approximations can be used for granulation of the universe. An information granule essentially comprises elements grouped together by similarity of their attributes (Yao, 1999).

We consider U as the finite and non-empty universe set of data. Then an equivalence relation $E \subseteq U \times U$ divides the U into disjoint partitions.

An arbitrary set $X \subseteq U$, however, may not be precisely described using equivalence granules and should be characterized by a pair of ordinary sets called lower and upper approximations (Pawlak, 1982):

$$lower(X) = \bigcup_{[x]_E \subseteq X} [x]_E \tag{1}$$

$$upper(X) = \bigcup_{[x]_E \cap X \neq 0} [x]_E \tag{2}$$

The lower approximation of *X*, *lower* (*X*), is the collection of all equivalence granules included in *X*, where granules are denoted as $[x]_E$. The upper approximation, *upper*(*X*), is the collection of non-elementary granules with non-empty intersection with X.
Equation (3) defines the accuracy of granule approximation (Baker, 2008):

$$ac(X) = \frac{|lower(X)|}{|upper(X)|} \tag{3}$$

where $|X|$ is the size of granule X, and $0 \le ac(X) \le 1$.
The quantity *q*(*X*) in (4) defines the quality of approximating *X* based on the available attributes (Greco et al., 2001):

$$q(X) = \frac{|lower(X)|}{|X|} \tag{4}$$

Obviously $0 \le ac \le q \le 1$.

2.1.1 Hierarchical rough granulation

The granules defined by rough set approximation can have a nested sequence, in which an equivalence granule $[X]_{E1}$ produced by equivalence relation E_1 considered to be included in an equivalence granule $[X]_{E2}$ produced by E_2, if $[X]_{E1} \subseteq [X]_{E2}$. This also means that E_1 produces a smaller granule than E_2 and E_2 is a union of some equivalence relations E_1. This notion can be extended to a sequence of granules or relations. The granules defined by rough set approximation can have a nested sequence:

$$[x]_{E1} \subseteq [x]_{E2} \subseteq \ ... \ \subseteq [x]_{En} \tag{5}$$

which means that:

$$ac([x]_{E1}) \subseteq ac([x]_{E2}) \subseteq \ ... \ \subseteq ac([x]_{En}) \tag{6}$$

thus forming a multi-layered granulation structure. The sizes of granules at different levels of granularity define the accuracy of classification. To achieve the best classification, a search within the layered granulations should be carried out to obtain the best granules to be used in GrC rule extraction.

2.2 Granular computing rule extraction

GrC is the science of information processing at different levels of granularity (Pawlak, 1982; Hobbs, 1985; Zadeh, 1997; Nguyen et al., 2001; Miao and Fan, 2002; Bargiela, 2003; Keet, 2008; Yao, 2008). In the GrC approach, information is divided into subsets or granules of information (Yao, 2001; Lin, 2003; Yao, 2008). By providing an information table about the problem, GrC constructs the granules considering the similarity of the attribute-values of the objects, and then focuses on induction of the classification rules.

In this paper, we replace GrC granulation with hierarchical rough set granulation because of its detailed quality measures that contribute to the uncertainty management. By obtaining the granules, GrC applies measures on them to obtain the best rules describing the relation of the granules with the assigned output classes.

GrC considers the attribute granules as the concepts ϕ and decision attributes as the concepts ψ. The rules are in the form of IF-THEN statements: "if an object satisfies ϕ, then the output class is ψ "(Yao, 2001) which is displayed by $\phi \to \psi$. In this paper, generality, absolute support and mutual support are used as rule describers.

Generality is the ratio of the objects that contribute in a specific rule (Yao, 2001):

$$G(\phi \to \psi) = \frac{|m(\phi \to \psi)|}{|U|} \tag{7}$$

where $|m(\phi \to \psi)|$ is the number of objects satisfying the concept ψ and $|U|$ is the size of universe. Rules having higher values of generality are more reliable (Yao, 2001).

The absolute support of the rule $\phi \to \psi$, *AS*($\phi \to \psi$), denotes the possibility of an object satisfying ϕ to have the class ψ (Yao, 2001) :

$$AS(\phi \to \psi) = \frac{|m(\phi \wedge \psi)|}{|m(\phi)|} \tag{8}$$

Mutual support ranges between 0 and 1, and indicates the degree to which ϕ confirms, and only confirms ψ. It is a measure of the strength of the two-way association $\varphi \Leftrightarrow \psi$ (Yao, 2002).

$$MS(\phi,\psi) = \frac{|m(\phi \wedge \psi)|}{|m(\phi \vee \psi)|} \tag{9}$$

3. EXPERIMENTAL RESULTS AND DISCUSSION

Census data of Tehran metropolitan area containing 3173 statistical units were accessed and used. In this data set, each statistical unit is characterized by six effective attributes, including average slope (slope), seismic intensity (MMI), percentage of weak buildings having less than or equal to four stories (WLE4), percentage of weak buildings having more than four stories (WH5), percentage of buildings constructed before 1966 (Bef-66), being the year of the commencement of executing building using standard construction rules in Iran, and the percentage of buildings constructed between 1966 and 1988 (Bet-66-88), where 1988 is the year that building was executed using earthquake resistance standards. 150 out of 3173 units were used as the samples to be ranked by experts against their degree of seismic vulnerability by numbers ranging from one to five, corresponding to very low vulnerability, low vulnerability, medium vulnerability, high vulnerability and very high vulnerability, respectively. 70% of the data were used for training and the rest used for testing the model. The six seismic parameters were discretized into four equal interval classes.

Position of the selected samples and their vulnerabilities are demonstrated in Figure 1. Experts' judgments on seismic vulnerability of the 5 selected samples are shown in Table 1. In addition, Figure 2 illustrates the proposed methodology for seismic vulnerability assessment of Tehran.

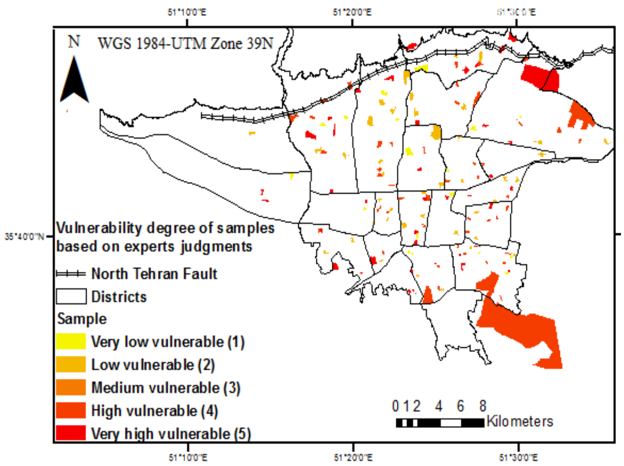

Figure 1. Position of the sample data in the study area

Sample id	Slope	MMI	WLE4	Bef-66	Bet-66-88	WH4	Expert remark
11	4	2	2	1	3	2	2
2799	1	2	1	1	4	1	2
1342	1	2	4	1	4	3	5
144	3	2	1	1	2	1	1
335	1	2	4	1	4	1	3

Table 1. Seismic physical vulnerability information classification for the five selected samples

To assess the quality of the underlying data, several tests were applied. As can be seen in the scatter plot diagrams (Figure 3), none of the criteria pairs showed a high correlation representing no redundancy between and among the employed data.

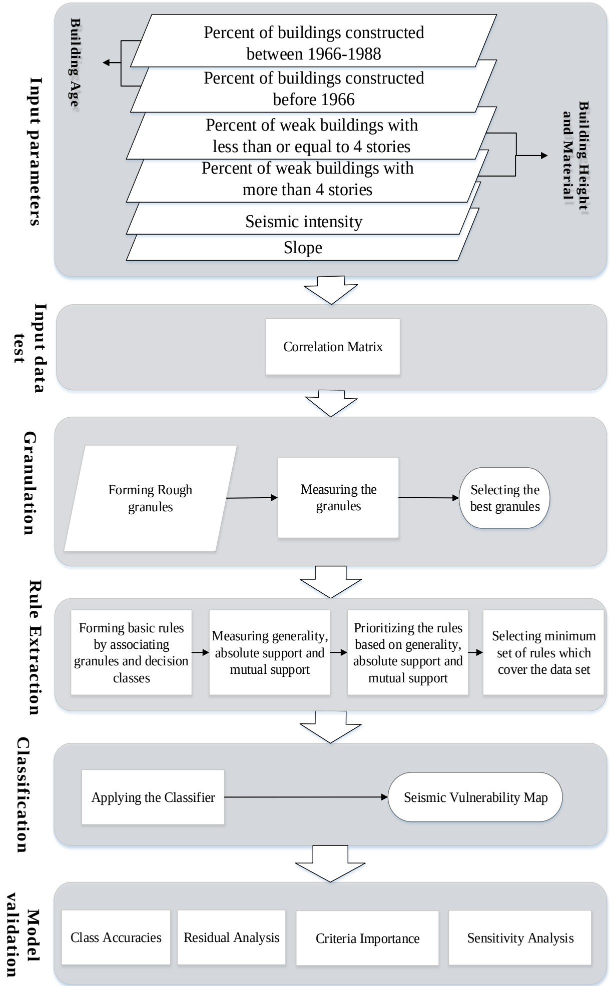

Figure 2: Proposed methodology

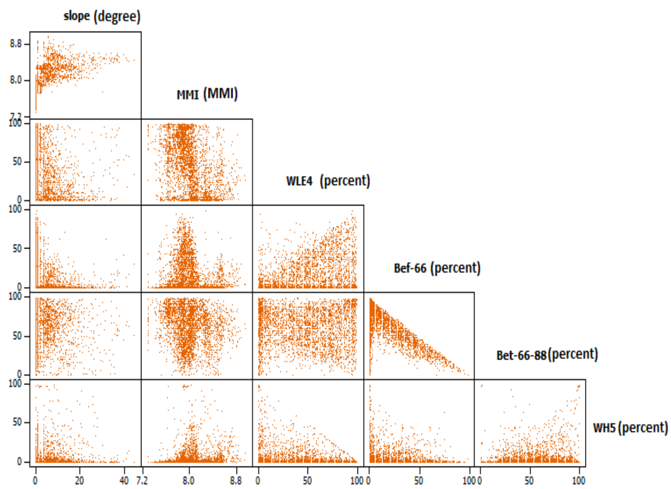

Figure 3. Scatter plot of the input criteria showing the correlation of each criteria pair

The data were used to extract high quality granules by means of hierarchical rough set granulation. In this regard, six levels of granulation were applied to the data with respect to the six attributes and 53 granules were selected among the full set of the extracted granules, according to the accuracy measures so that the selected granules cover the universe. The quality parameters for several extracted granules are presented in Table 2:

Granule id	Level of granularity	Accuracy (ac)	Quality (q)
1	6	1	1
5	6	0.89	0.9
32	5	0.94	0.96

Table 2. Quality measurements of granules showing level of the granule in the hierarchical structure, accuracy and quality values

GrC applied to the selected granules and 30 classification rules were extracted. Quality measures for some of these rules are shown in Table 3, whereas Figure 4 shows the obtained decision tree.

Rule id	Generality	Absolute Support	Mutual Support
9	0.06	0.95	0.95
18	0.09	0.92	0.96
22	0.06	0.82	0.9

Table 3. Quality measurements of the extracted rules

The acquired decision tree was applied to the study area. A seismic physical vulnerability map was produced (Figure 5). In this map, 9%, 11%, 44%, 16% and 20% of the units were assigned with the degree of vulnerability between 1 and 5, respectively.

The model resulted in 0.86 for R^2 (coefficient of determination) and an RMSE value of 0.43 for the training data, and 0.85 and 0.46 for the test data. The accuracies for classes one to five were equal to 94%, 83%, 86%, 77% and 92%, respectively.

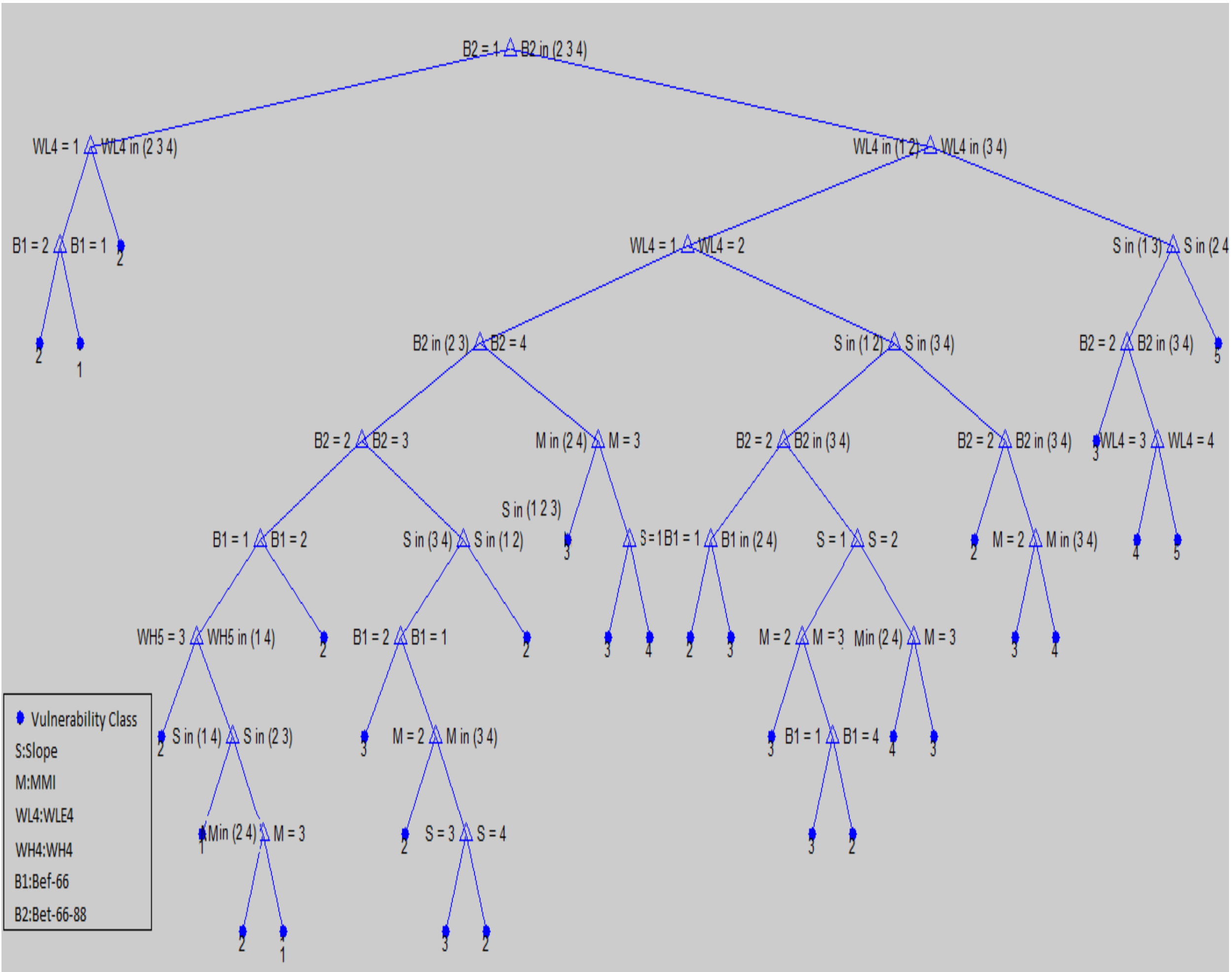

Figure 4. Decision tree of the extracted rules showing rule elements and resulting classes in a tree form

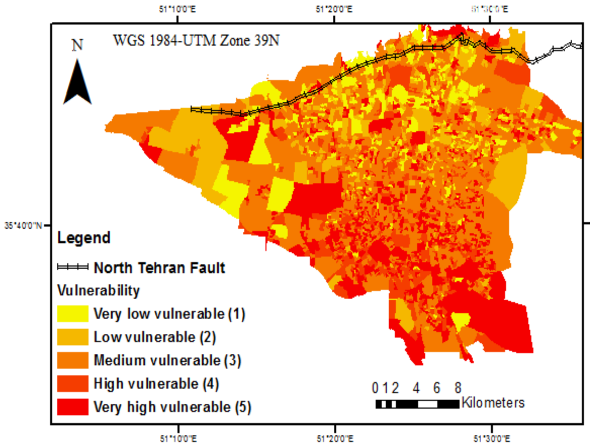

Figure 5. Seismic physical vulnerability map of Tehran considering North Tehran fault activation

Moreover, the relative importance of the criteria is calculated based on their frequency and influence in the rule set (Table 4).

Criteria	Slope	MMI	less4	Bef-66	Bet-66-88	more4
Relative importance	0.04	1	0.96	0.56	0.32	0.62

Table 4. Relative importance of the criteria, normalized with respect to the most important criterion

To assess the acquired results, some quality measurements were carried out. Though there is no real earthquake damage measurements in Tehran, most of the tests are based on quality of predicting assigned values by experts. Accuracy indices of the implemented rough-GrC method in comparison with the GrC method are demonstrated in Table 5:

Algorithm	Rough-GrC			GrC		
Indices	R^2(coefficient of determination)	RMSE	p-value	R^2	RMSE	p-value
Train	0.89	0.37	0.0002	0.87	0.45	0.0002
Test	0.83	0.46	0.0002	0.83	0.49	0.0001

Table 5. Accuracy assessment of the implemented method in comparison with GrC algorithm

According to Table 5, both of the methods provided acceptable accuracies and residuals, while the proposed rough-GrC algorithm outperformed the basic GrC algorithm.

Table 6 shows the accuracies obtained for each vulnerability class. Best prediction performance resulted for the classes 1 and 5.

Assigned Class / True Class	Class 1	Class 2	Class 3	Class 4	Class 5	Accuracy (%)
Class 1	16	2	0	0	0	88
Class 2	2	35	4	0	0	85
Class 3	1	5	42	1	0	86
Class 4	0	0	2	13	1	83
Class 5	0	0	1	2	23	88

Table 6. Accuracy of classification per class

4. SENSITIVITY ANALYSIS

Finally, a sensitivity analysis revealed the uncertainty propagation in the model. This can be used as a test of the robustness of the results. In addition, it allows a user to identify the critical criteria, which cause significant change in the output if they are slightly changed. It is possible for a low-importance criterion to be more critical than higher importance criteria (Triantaphyllou, 1997). Here the sensitivity analysis is done with the assumption of change in each criterion, while the other criteria remained unaffected. The results of the analysis are represented in Figure 6 which shows that the MMI is the most critical criterion in seismic vulnerability assessment. Slope and WLE4 are the least critical criteria and other parameters have moderate effect. It has to be mentioned that the sensitivity analysis is based on the change in the acquired results, thus it could not be undertaken for the rule extraction section.

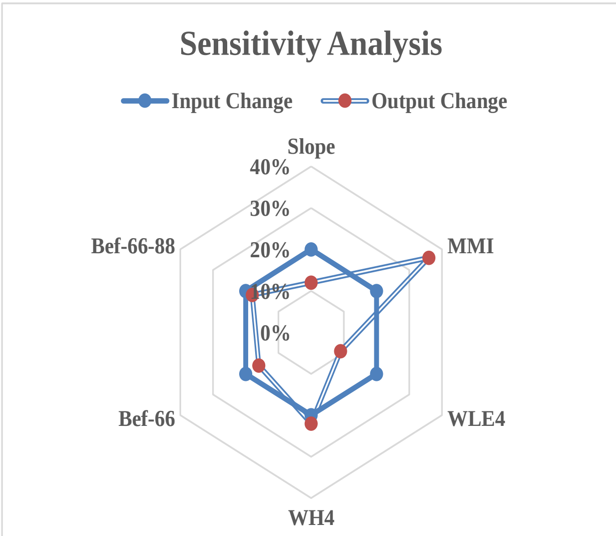

Figure 6. Sensitivity analysis of the model against 20% variation in the input values

5. CONCLUSION

This paper proposed a new methodology for seismic vulnerability assessment of Tehran using the integration of hierarchical rough granulation and granular computing rule extraction algorithm. Granules provided by hierarchical rough set granulation showed accuracies above 0.84 and qualities above 0.86. The rules extracted by granular computing provided acceptable results in quality measurements.

The model was tested and confirmed by various measurements including the prediction accuracy for the given training data, a residual analysis and a sensitivity analysis. In addition, the importance of the input criteria was determined.

The proposed model resulted in higher accuracy in seismic vulnerability assessment as compared to previous research (Samadi Alinia and Delavar, 2011; Khamespanah et al., 2013a). Moreover, the resulting vulnerability map indicates a high amount of seismic vulnerability in Tehran, thus requiring comprehensive disaster risk reduction plans to be developed.

REFERENCES

Baker, W. 2008. Uncertainty propagation in probabilistic seismic loss estimation. *Structural Safety* 30(3), pp 236-252.

Bargiela, A. 2003. *Granular Computing: An Introduction*, The Springer International Series in Engineering and Computer Science, Vol. 717, New York, US, 452p.

Greco, S., B. Matarazzo and R. Slowinski. 2001. Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research* 129(1), pp 1-47.

Hansson, S. O. 1996. Decision making under great uncertainty. *Philosophy of the Social Sciences* 26(3), pp 369-386.

Hobbs, J. R. 1985. Granularity. Proc.Ninth International Joint Conference on Artificial Intelligence, Los Angeles, CA, Aug. 1985, pp 432-435.

Jahanpeyma, M., M. R. Delavar, M. Malek and N. Kamalian. 2007. Analytical evaluation of uncertainty propagation in seismic vulnerability assessment of Tehran using GIS. International Symposium on Spatial Data Quality (ISSDQ 2007), Enschede, The Netherlands, June 13-15 2007, pp 5-9.

Keet, C. M. 2008. A Formal Theory of Granularity, PhD Thesis, KRDB Research Centre, Faculty of Computer Science, Free University of Bozen-Bolzano, Italy, 268p.

Khamespanah, F., M. Delavar and M. Zare. 2013a. Uncertainty management in seismic vulnerability assessment using granular computing based on covering of universe. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1(1), pp 121-126.

Khamespanah, F., M. R. Delavar, H. Samadi Alinia and M. Zare. 2013b. Granular Computing and Dempster–Shafer Integration in Seismic Vulnerability Assessment. *Intelligent Systems for Crisis Management*. S. Zlatanova, Peters, R., Dilo, A., Scholten, H., Springer, pp 147-158.

Lin, T. Y. 2003. Granular computing. *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, Springer, Chongqing, China, pp 16-24.

Miao, D.Q. and S.D. Fan. 2002. The calculation of knowledge granulation and its application. *Systems Engineering-theory & Practice* 1, pp 78-82.

Nguyen, S. H., A. Skowron and J. Stepaniuk. 2001. Granular computing: A rough set approach. *Computational Intelligence* 17(3), pp 514-544.

Paté-Cornell, M. E. 1996. Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering & System Safety* 54(2), pp 95-111.

Pawlak, Z. 1982. Rough sets. *International Journal of Computer & Information Sciences* 11(5), pp 341-356.

Samadi Alinia, H. and M. Delavar. 2011. Tehran's seismic vulnerability classification using granular computing approach. *Applied Geomatics* 3(4), pp 229-240.

Tesfamariam, S., R. Sadiq and H. Najjaran. 2010. Decision making under uncertainty - An example for seismic risk management. *Risk Analysis* 30(1), pp 78-94.

Triantaphyllou, E. 1997. A Sensitivity analysis approach for some deterministic multi-criteria decision-making methods. *Decision Sciences* 28(1), pp 151-194.

Yao, Y. 1999. Stratified rough sets and granular computing. Proc. of the 18th International Conference of the North American Fuzzy Information Processing Society, Jun. 10-12, 1999, New York, US, pp 800-804.

Yao, Y. 2002. Granular computing using information tables. *Data Mining, Rough Sets and Granular Computing*, Springer, Heidelberg, Berlin, pp 102-124.

Yao, Y. 2008. A Unified Framework of Granular Computing. *Handbook of Granular Computing*. W. Pedrycz, A. Skowron and V. Kreinovich (Editors), John Wiley & Sons, Chichester, UK, pp 401-410.

Yao, Y. Y. 2001. On Modeling data mining with granular computing. Proc. Of the 25th Annual IEEE International Conference on Computer Software and Applications, Chicago, US, Oct. 8-12 2001, pp 638-643.

Zadeh, L. A. 1997. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems* 90(2), pp 111-127.

Zadeh, L. A. 2005. Toward a generalized theory of uncertainty (GTU)—an outline. *Information Sciences* 172(1), pp 1-40.