Application of the rule extraction method to evaluate seismicity of Iran

Marziyeh Khalili[1]*, Ahmad Zamani[2]

[1] Department of earth sciences, college of sciences, Shiraz University, Shiraz, Iran
[2] Department of geology, Shiraz Branch, Islamic Azad University, Shiraz, Iran
**Corresponding author, e-mail: marzieh-khalili@shirazu.ac.ir*

(*received: 23/05/2016 ; accepted: 25/09/2016*)

Abstract

Assessing seismic hazards involves specifying the likelihood, magnitude and location of earthquakes in a region. Predicting the seismic hazards is the first step in reducing the impact of the damage caused by an earthquake. In this study, to fully utilize all the known parameters which may possibly affect the occurrence of earthquakes ($m_b \geq 4.5$); a data-driven rule-extraction method called the Classification and Regression Tree (CART) was used to find the rules governing the earthquakes that occur. The method produces Predictive Rule Based Seismicity Map (PRBSM) of Iran that shows regions with high earthquake hazards. The rules are based on a large number of geophysical and geological parameters. The PRBSM has been built based on earthquake data from the year 1900 up to the end of 2006 and has been validated using earthquakes from 2007 to the end of 2015. In addition, this method allows for the identification of the most important combination of parameters associated with earthquakes. For example, the isostatic anomaly has the highest correlation with earthquakes in Iran. A distinctive character of this paper is the predictive rule based method which can create online as well as offline maps which are flexible and readily automated.

Keywords: *Decision Tree; Predictive Model; Seismic Hazard Map; Iran.*

Introduction

Iran is one of the disaster-prone countries of the world, frequently suffering from destructive earthquakes that leave a large number of casualties and financial losses. Iran and its neighboring countries' seismic activities are closely related to their positions within the active Alpine-Himalayan Belt. During the past decades, many researchers have studied the seismicity of Iran. For example, (Nowroozi, 1976), (Tavakoli & Ghafory-Ashtiani, 1999), (Bonini *et al.,* 2003) and (Ashtari Jafari, 2010). In studies of this type, the factors contributing to the earthquake occurrence were studied independently. These studies did not consider the effects of other parameters on earthquake occurrence.

The application of machine learning and data mining methods is very common in various fields of science such as business, social sciences, biological and environmental sciences and engineering. During the last few years, some researchers used machine-learning methods to build classifiers or to predict earthquakes. For example, Zmazek *et al.* (2003) used a model tree to predict earthquakes based on soil radon data. A rough set and decision tree (C4.5 algorithm) have been used to characterize premonitory factors of low seismic activity (Iftikhar *et al.,* 2009). Reyes *et al.,* (2013) used artificial neural networks to predict earthquakes in Chile. None of these studies has shown a combination of parameters which leads to the identification of the regions with high active seismicity.

In this study, a multivariate analysis of parameters affecting the earthquake occurrence was investigated. A rule extraction method was used to determine the combination of parameters that correlate well with earthquakes and to explain the seismicity patterns in Iran. A major advantage of using rule base methods is that they are mostly data driven, nonparametric and without priori assumptions. Historical facts are the main players in model construction when models are based on data alone without any discrimination based on the researchers' opinions.

Seismic hazard analysis requires an assessment of the future earthquake potential (the likelihood, magnitude and location of earthquake), which might have damaging effects in a region. Identifying regions with high seismic hazards is important in planning risk mitigation strategies and can help city planners to identify where to enforce stricter construction standards as well. Generally, seismic hazard maps are produced based on limited factors such as ground velocity and ground acceleration for example (Tavakoli & Ghafory-Ashtiany, 1999).

Studies such as: Berg *et al.* 1964; Bouchon 1973; Davis and West 1973; Caputo *et al.* 1984, 1985; Geli *et al.* 1988; Johnston, 1997; Clift *et al.* 2000; Zamani and Hashemi, 2000; Chen *et al.* 2002;

Mishra et al., 2005; Li and Li 2009; Zamani *et al.* 2012; Wu *et al.,* 2015; Genti *et al.* 2016, indicate that some geological and geophysical parameters such as the isostatic anomaly, topography, gravity anomaly, and the electromagnetic field have effects on the seismicity of an area. Hence, in this study, to fully utilize all the known parameters which may possibly affect the occurrence of earthquakes (mb ≥ 4.5); a data-driven rule-extraction method called the Classification and Regression Tree (CART) (Breiman *et al.,* 1984) was used to find the rules governing the earthquakes that occur. CART analysis is a machine learning method based on statistical rules that proves fruitful in both prediction and rule extraction problems. The surface and sub- surface data were used to build a multivariate numerical database, and then based on a combination of major parameters rules governing high impact earthquakes were extracted.

The main aim of this paper is to introduce a modern data analytical and sorting technique to develop a useful new type of seismic hazard map, which can create online as well as offline and it is flexible and readily automated.

Method of Analysis

Data Mining

Devices which automatically collect data, together with inexpensive storage devices, have accumulated a large database to analyze properly. Indeed, a large amount of data is not processed at all. Data mining aims to extract interpretable and actionable patterns of knowledge from large data sets. The patterns must be non-trivial, interesting, implicit, previously unknown and potentially useful. Stated differently, data mining is a technology that enables exploration, analysis, and visualization of information from very large databases at a high level of abstraction without a specific hypothesis in mind. Several tools and techniques, among which are classifiers such as decision trees, neural networks, SVM, clustering algorithms, like k-means or k-mode and association rules like a priori and FP-growth are deployed in data mining (Han & Camber, 2006).

Decision Tree

A decision tree is a classifier in the form of a tree structure with a top-down hierarchy used in statistics, machine learning and data mining (Fig1). This technique by deriving meaningful decision rules and maximizing differences on a dependent variable is often considered to improve knowledge representation structure (Daubie *et al.,* 2002).

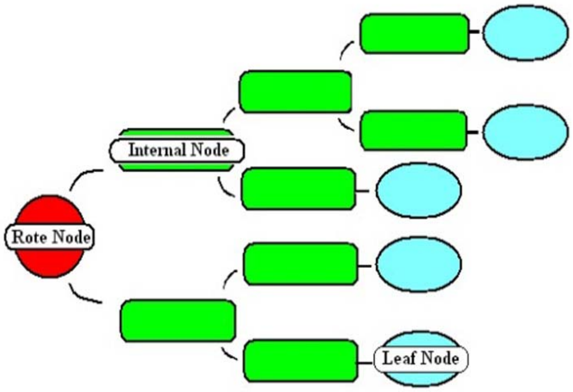

Figure 1. A typical binary decision tree

Root node-the topmost node in a tree- and the internal node denote a test on an attribute and the leaf node or terminal node denotes the predicted value of the target attribute given the values of the attributes represented by the path from the root.

The rules are easily interpretable allowing complex relationships to be represented in a comprehensible and an intuitive manner. The relationship between descriptions of objects and their assignment to a specific class is established by the rules. This technique eliminates redundant (unnecessary) attributes from the classification. Decision tree produces a directed tree as a predictive model. A database is classified by starting at the root node of the decision tree and testing the attribute by this node. The tree branch corresponding to the value of the attribute moves down to some internal nodes.

This process is then repeated until a leaf node is reached and provides the classification of the instance.

The Classification and Regression Tree is a tree-based classification and a prediction method which uses recursive partitioning to split the training dataset into segments with similar output field values. CART is an approach to generating a learning decision tree from a training dataset to predict a numeric value. This algorithm spawned a flurry of work on decision tree induction.

CART adopts a greedy approach in which a decision tree is constructed in a top-down recursive divide-and-conquer manner (Han and Camber, 2006). CART produces a binary decision tree which has a flowchart-like tree structure. The root node represents the source or full training dataset and is displayed at the top. The classification of a particular source set proceeds from top to bottom. The questions asked at each node concern a particular attribute or property of the data set, and the downward links or branches correspond to the possible answer (i.e. attribute values). Based on the answers, the appropriate branch or links to successive internal nodes are followed until a leaf or terminal node is reached where the value of the target attribute (class) is read. The classification in a decision tree proceeds from top to bottom. In the CART method, each leaf or terminal node stores a numeric prediction. In fact, it is the average value of the predicted attribute for the training dataset that reaches the terminal node.

CART creates a rule for each path from the root to a leaf node. Each splitting criterion along a given path is logically ANDed to form the rule antecedent (.IF” part). The class prediction is the resultant leaf node, forming consequent of the rule (.THEN” part).

R1: IF a AND b,… THEN c.

The rules are easily interpretable allowing complex relationships to be represented in a comprehensible and an intuitive manner. The relationship between descriptions of objects and their assignment to a specific class is established by the rules. Moreover, the rules can be used for the classification of new objects.

The advantage of using CART is that it is non-parametric and can evaluate data that are highly skewed or multimodal (Lewis, 2000). CART is well suited for data mining since it can reveal non-obvious and complex relationships between the splitting variables and the predicted variables. Datasets are partitioned into two subsets so that the records within each subset are more homogeneous than in the previous subset. Splitting is a recursive process, and the process is repeated until the stopping criterion or homogeneity criterion is reached. CART is quite flexible, allowing specifying the prior probability distribution in a classification problem. CART chooses a split at each node such that each child node created by the split is purer than its parent node. Here purity refers to the similarity of values in the target field. In a completely pure node, all the records have the same values as the target field. CART measures the impurity of a split at a node by defining an impurity measure. Depending on the type of the target field, three different impurity measures are used to find splits for CART models. For numeric target fields, the least squared deviation (LSD) impurity measure is used.

The process that controls how the algorithm decides when to stop splitting nodes in the tree is the stopping criteria. Tree growth continues until every leaf node in the tree triggers at least one stopping criterion.

Pruning is the process that examines a fully-grown tree and removes bottom-level splits, which do not play an important role on the accuracy of the tree. In pruning, the smallest tree is created by keeping the cost as low as possible. A tree branch is removed if the cost associated with having a more complex tree exceeds the gain associated with having another level of nodes (Han and Camber, 2006).

Cross Validation

Since data-driven methods may lead to the creation of models which are only "good" for the training data, cross validation methods are deployed to assess the performance of the treated models on "unseen" data. Cross validation is a technique for assessing how well the results of a statistical analysis can be generalized to a data set that is not used for training. It is used in a setting where the goal is classification or prediction, and one wants to estimate how accurately a predictive model will perform in practice. The data used is seldom seen at the time of model construction. In one round of cross-validation, a sample of data is partitioned into complementary subsets. It then performs the analysis on the training set, and validates the analysis on the validation set or testing set. A variety of cross validation techniques exist (Efron and Tibshirani, 1997), (Kohavi, 1995) and (McLachlan *et al.,* 2004). In this paper 10-fold cross validation, 10-CV was used. In K-fold cross-validation, the original data set is randomly partitioned into K- folds or subsample. The model is then generated excluding the data from each subsample in turn. The first tree is generated based on all of the cases except those in the first fold, and the second tree is based on all cases except those in the second sample fold. Therefore, K times the process of training and testing is performed, where in each round a single subsample is retained as the validation data for testing the model, and the remaining K−1 subsamples are used for training.

Gain Summary

The gain summary displays descriptive statistics for all terminal nodes in the tree.

The gain summary in target field with continuous parameter shows the weighted mean of the target value for each terminal node. Gains provide a measure of how far the mean or proportion at a given node differs from the overall mean. For the most part, the greater this difference, the more useful the tree is as a tool for making decisions.

Data Set

Databases are highly susceptible to missing, noisy, and inconsistent data. Low-quality mining results will be obtained from Low-quality data. If the data have been normalized, data mining methods provide better results (Han & Camber, 2006).

In order to construct an exact Predictive Rule Based Seismicity Map (PRBSM) of Iran, large numbers of updated, numeric and normalized geological, geophysical and seismological characteristics have been compiled for the 175 quadrangular sites of 1º area. The study area (Iran) with coordinates of 44°- 63° east longitude and 25°-39° north latitude is divided into 175 quadrangles each covering one degree of latitude and longitude (None of offshore Iran and islands is included in the data set). These quadrangles are used as observations (input samples) and a large number of possible measures of tectonic and seismotectonic characteristics are considered as variables (attributes). Each observation or case has been characterized by 46 variables which seem to characterize the intensity and degree of contrast between tectonic and seismotectonic structures in Iran.

In this research geological data had been obtained from digitized and regular geological maps of Iran (Geological Survey of Iran, 2004) including relative areas of surface rock (age and type) (%), fault length density and average Moho depth. Geophysical data were taken from Dehghani and Makris (1983), total magnetic intensity maps of Iran (Yousefi, 1989), and consist of magnetic intensity, free air anomaly, gravity anomaly, Bouger anomaly, residual Bouger anomaly and isostatic anomaly. Seismological data (historical and instrumental) were taken from earthquakes that occurred between the years 1900 to end of 2015 (Ambraseys, 2001); (Engdhal *et al.,* 2006); (ISC, 2015) and (NEIC, 2015). It is comprised of earthquake magnitude, energy and number of earthquakes greater than mb ≥ 4.5; (Table 1).

Generally, it is useful to forecast destructive earthquakes with magnitudes more than 6 but due to the fact that many towns, and villages in Iran have low resistance against earthquakes a threshold magnitude of MC = 4.5 has been selected (Zamani and Agh- Atabai, 2009; Zamani *et al.,* 2012). In addition, using a threshold magnitude of MC=4.5 makes model validation easier because of the poor statistics of the very few large earthquakes. Characteristically, the geophysical and geological variables gathered are not only correlated with each other, but each attribute is also influenced by the other attributes. Therefore, in many cases the attributes are interwoven in such a way that they yield little information about the region under investigation when analyzed individually (Zamani & Khalili, 2006; Zamani *et al.,* 2011, 2012).

Results and Discussion

An earthquake is a very complex process with several parameters. Knowing the most significant factors associated with earthquakes can help build predictive models. In this study, a numeric and updated catalogue of geological, geophysical and seismological data of Iran had been gathered. Then a rule extraction method was used to determine the combination of parameters that correlate well with earthquakes. In other words, almost all the possible factors and parameters that may have affected earthquake occurrences were collected and analyzed using the non-parametric data-driven method to forecast earthquakes.

To this end, the earthquake records from 1900 up to the end 2006 (Fig. 2) were collected to build a predictive model. The target for the model was the number of earthquakes which were greater than $m_b \geq 4.5$ (NEGMB). All the variables presented in Table 1 were used as potential predictor variables. Cross-validation was deployed to counter overfitting. Cross validation is a technique for assessing how the results of a statistical analysis will generalize an independent data set. It is used mainly in settings where the goal is prediction and where one wants to estimate how accurately a predictive model will perform in practice. As the input variables were mainly continuous (numeric), CART was run in the regression-tree mode. This method produced the decision tree diagram (Fig. 3) and the gain summery (Table 2).

Nodes with index values greater than 100% indicate that a better chance exists of accurate prediction by selecting records from these nodes instead of random selection from the entire sample.

Table 1. Attributes used for constructing the Predictive Rule Based Seismicity Map (PRBSM), measured within 1º Quadrangles

No.	Attributes	No.	Attributes
1	Maximum earthquake magnitude(mb), MXEMG	24	Minimum gravity anomaly (mgal), MIGRV
2	Number of earthquakes greater than mb≥4.5, NEGMB	25	Range of free air anomaly (mgal), RAFRA
3	Maximum seismic energy released (j), MXSER	26	Average free air anomaly (mgal), AVFRA
4	Fault length density (km^{-1}), FLTLD	27	Maximum free air anomaly (mgal), MXFRA
5	Range of isostatic anomaly (mgal), RA ISO	28	Minimum free air anomaly (mgal), MIFRA
6	Average isostatic anomaly (mgal), AVISO	29	Range of magnetic intensity (gamma), RAMGI
7	Maximum isostatic anomaly (mgal), MXISO	30	Average magnetic intensity (gamma), AVMGI
8	Minimum isostatic anomaly (mgal), MIISO	31	Maximum magnetic intensity (gamma), MXMGI
9	Range of regional Bouger anomaly (mgal), RAEGB	32	Minimum magnetic intensity (gamma), MIMGI
10	Average regional Bouger anomaly (mgal), AVREG	33	Average Moho depth (km), AVMOD
11	Maximum regional Bouger anomaly (mgal), MXREG	34	Range of elevation (m), RAELV
12	Minimum regional Bouger anomaly (mgal), MIREG	35	Average elevation (m), AVELV
13	Range of residual Bouger anomaly (mgal), RARES	36	Maximum elevation (m), MXELV
14	Average residual Bouger anomaly (mgal), AVRES	37	Minimum elevation (m), MIELV
15	Maximum residual Bouger anomaly (mgal), MXRES	38	Relative area of surface unconsolidatedcover(%),RAUNR
16	Minimum residual Bouger anomaly (mgal), MIRES	39	Relative area of surface sedimentary rocks (%),RASER
17	Range of Bouger anomaly (mgal), RABUG	40	Relative area of surface metamorphic rocks (%),RAMER
18	Average Bouger anomaly(mgal), AVBUG	41	Relative area of surface igneous rocks (%),RAIGR
19	Maximum Bouger anomaly(mgal), MXBUG	42	Relative area of surface ophiolitic rocks (%),RAOPR
20	Minimum Bouger anomaly(mgal), MIBUG	43	Relative area of surface Cenozoic rocks (%),RACER
21	Range of gravity anomaly (mgal), RAGRV	44	Relative area of surface Mesozoic rocks (%),RAMER
22	Average gravity anomaly (mgal), AVGRV	45	Relative area of surface Paleozoic rocks (%),RAPAR
23	Maximum gravity anomaly (mgal), MXGRV	46	Relative area of surface Proterozoic rocks (%),RAPTR

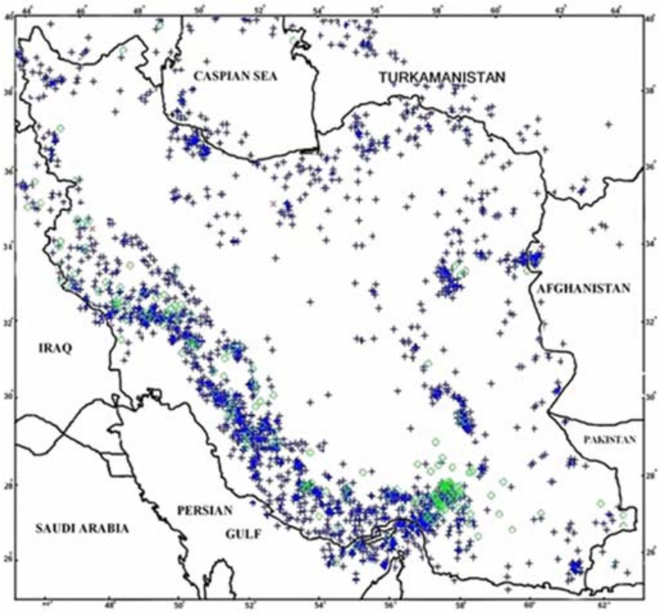

Figure 2 Seismicity map of IRAN (1900-2006)

Table 2. The Gains summery. Node Number: the number of node in the decision tree. Number of observations in node: The total number of records in each node. Node percentage: The percentage of all records in the dataset that fall into this node. Predicted values %: The percentage of predicted target for each node. Gains index value %: The Gains index measures how well a given node separates the attributes of the training examples according to their target classification.

Rule No.	Node Number	Number of Observations in Node	Node Percentage	Predicted values %	Gain Index value %
1	12	7	3.6	71	452.3
2	24	3	1.5	55.6	361
3	14	2	1.3	48.1	311
4	11	6	3.2	42	257.4
5	23	7	4.1	38	249.2
6	26	3	1.1	29	200.4
7	22	4	2.2	24	144.4
8	15	3	1.1	22	140.2
9	30	2	1.7	16	118.3
10	10	21	12.1	17	99.5
11	29	2	1	14.4	94.1
12	20	8	4.5	14.2	92.7
13	25	7	4.1	13.1	85.5
14	19	12	6.8	11.6	75.3
15	21	5	2.9	9.1	59.1
16	13	25	14.3	8.8	57.7
17	32	41	23.4	5.7	36.9
18	27	3	1.7	4.8	31.3
19	31	14	8	3	19.9

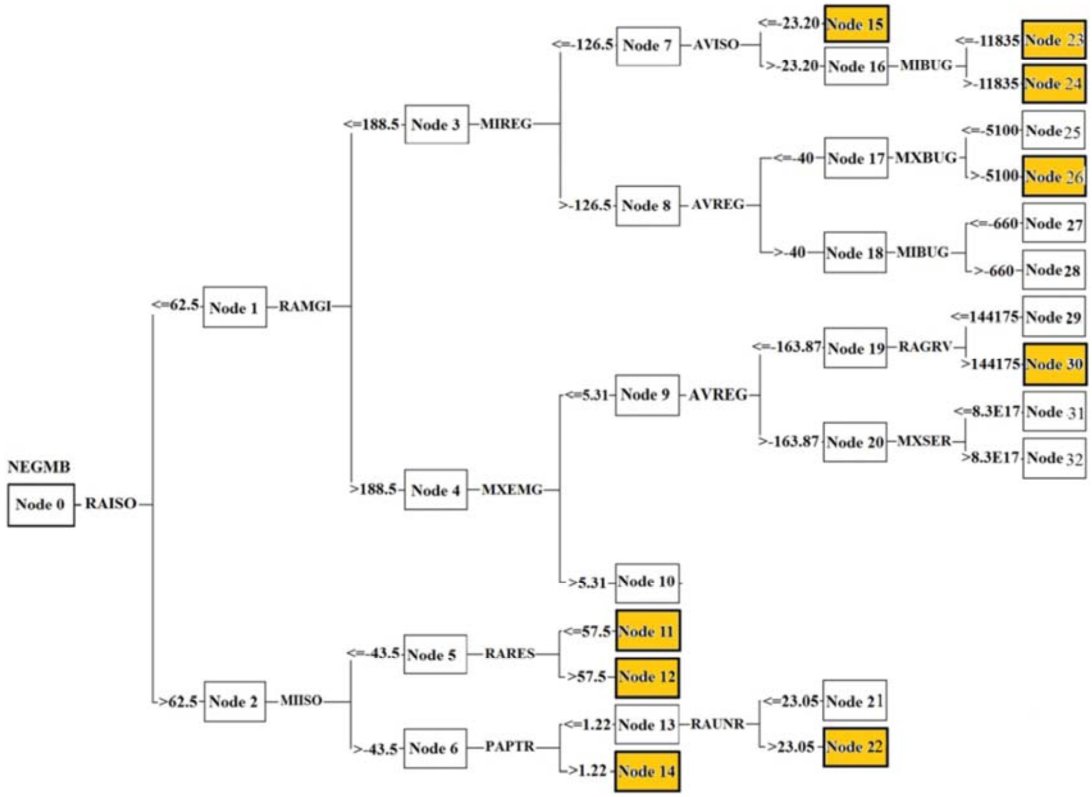

Figure 3. The CART binary decision tree. Nodes with gain index values greater than 100% are highlighted in the figure

The gain index percentage tells us how much greater the proportion of a given target at each node differs from the overall proportion.

The index values in this paper show that node 12 has the highest possible rate (a value of 452%) for the entire data. This node is thus almost 4.5 times more likely to get a hit with these records than using a random selection. The gain index values show that of the 19 nodes, 9 have index values greater than 100%. The rules of the top nine nodes are depicted in Table 3:

Table. 3The nine most reliable nodes, with their associated rules. The number listed after each set of initials in Table 3, is its attribute in Table 1.

Rule No.	**Node No.**	**IF**	**THEN NEGMB(4)**
1	12	RAISO(6) > 62.5 and MIISO(9) ≤ -43.5 and RARES(14) > 57.5	71%
2	24	RAISO(6) ≤ 62.5 and RAMGI(30) ≤ 188.5 and MIREG(13) ≤ -126.5 and AVISO(7) > -23.2 and MIBUG(21) > - 11835	55.6%
3	14	RAISO(6) > 62.5 and MIISO(9) > -43.5 and PAPTR(47) > 1.2	48.1%
4	11	RAISO(6) > 62.5 and MIISO(9) ≤ -43.5 and RARES(14) ≤ 57.5	42%
5	23	RAISO(6) ≤ 62.5 and RAMGI(30) ≤ 188.5 and MIREG(13) ≤ -126.5 and AVISO(7) > -23.2 and MIBUG(21) ≤ -11835	38%
6	26	RAISO(6) ≤ 62.5 and RAMGI(30) ≤ 188.5 and MIREG(13) > -126.5 and AVREG(11) ≤ -40 and MXBUG(20) > -5100	29%
7	22	RAISO(6) > 62.5 and MIISO(9) > -43.5 and PAPTR(47) ≤ 1.2 and RAUNR(39) >23.05	24%
8	15	RAISO(6) ≤ 62.5 and RAMGI(30) ≤ 188.5 and MIREG(13) ≤ -126.5 and AVISO(7) ≤ -23.2	22%
9	30	RAISO(6) ≤ 62.5 and RAMGI(30) >188.5 and MXEMG (3) ≤ 5.3 and AVREG(11) ≤ -163.9 and RAGRV(22) > 144175	16%

Statistically, significant rules associated with the patterns of earthquakes were found. It is interesting that the results indicate that the isostatic anomaly is a very important parameter in seismic activity. Other important factors in decreasing the order of importance are: regional Bouger anomaly, Bouger anomaly and gravity anomaly respectively. In CART method, the decision tree was built to decrease the importance, that is, the most important of the parameters come first; the parameters for the next important come next, and so on.

Our results have a good correlation with previous studies. Zamani and Hashemi, 2000, have reported positive correlations between seismicity and isostatic anomalies from the Iranian plateau. Vertical differential stress in Kachchh may be further accentuated due to large-scale deposition of sediments in the adjoining north Arabian Sea (Clift *et al.,* 2000). Mishra *et al.,* 2005, have demonstrated which presence of thick crustal roots give rise to buoyancy and may lead to the accumulation of local stress, especially at the periphery of the crustal roots. Wu *et al.,* 2015, have explained that earthquakes in Mid-Yunnan and the surrounding area often occur at the maximum value of the gravity gradient zone and negative magnetic anomaly regions. Genti *et al.,* 2016 have concluded that flexural rebound induced by surface processes (gravitational potential energy associated with topography and dense crustal blocks; isostatic compensation in response to denudation and/or sedimentation) is able to explain the seismicity in Central–Western Pyrenees.

In this research, the PRBSM is defined as the map of regions with a high hazard of future earthquake occurrence (mb ≥4.5). Using the above-mentioned rules and attributes produced the PRBSM of Iran (Fig4). The map indicates that Bandar Abbas in southern Iran, parts of the Zagros simply folded belt, the Oman line (a limited area southern Iran) and the northern portion of the Lut block eastern Iran are regions of high hazards of future earthquakes m_b ≥4.5. To assess the accuracy of the rules applied, the seismicity map of Iran based on earthquakes (mb ≥4.5) that occurred from the year 2007 up to End of 2015 was produced (Fig5).

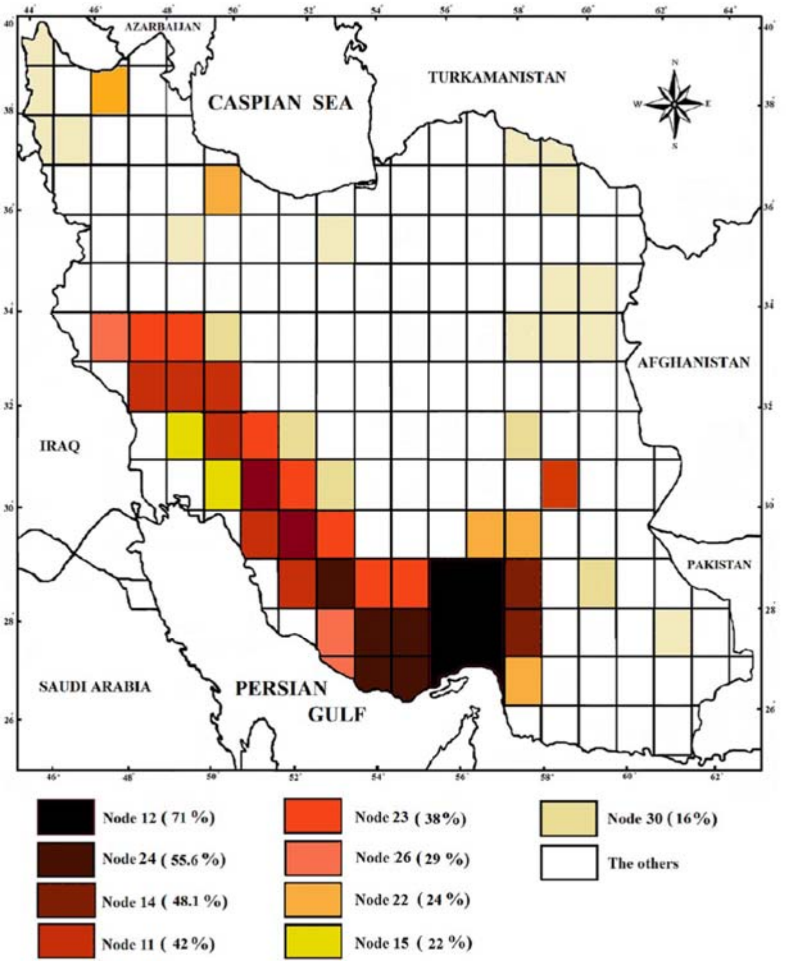

Figure 4. The Predictive Rule Based Seismicity Map of Iran (PRBSM)

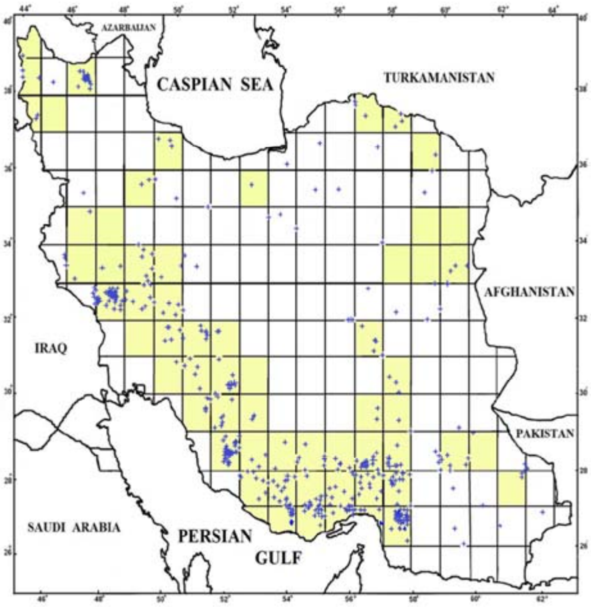

Figure 5. The Seismicity map of Iran based on earthquakes (mb ≥4.5) from the years 2007 up to end of 2015. Nodes of figure 4 are highlighted in this map

The model has been validated by comparing the predicted nodes (regions) in the model (Fig4) with the observed seismicity map of Iran (Fig5) for the period of 2007 up to end of 2015.

The evidence indicates that most of the earthquakes that occurred" between" 2007 to end of 2015 in fact happened in the areas predicted by the PRBSM model. The results indicate that the model has high accuracy and the introduced approach is a reliable method for knowledge extraction from the seismicity pattern of Iran. This paper applies a new class of data-driven rule–based model to create online as well as offline interactive seismic hazard map that is flexible and readily automated. This model might be improved and refined by the collection of new geological and geophysical data.

Conclusion

This research applies a modern data analytical and sorting method to develop a useful new type of seismic hazard map. A numeric and updated catalogue of geological, geophysical and seismological data from Iran have been used to build a multivariate numerical database, and then based on a combination of major parameters rules governing high impact earthquakes were extracted. The rules extracted from among the attributes were significant statistically to assure that the mapped patterns are not random and in fact relate to earthquake locations. The map shows Bandar Abbas in southern Iran, the Zagros simply folded belt, the Oman line and the northern portion of Lut block in eastern Iran have high seismic hazards for future number of earthquakes occurring with mb≥4.5. Furthermore, the results indicate that the model has high accuracy. Therefore, it is a reliable method for knowledge extraction from the seismicity pattern of Iran to forecast future earthquake hazards. The analysis also shows that the isostatic anomaly correlates best with these earthquakes. A distinctive character of this paper is the predictive rule based model that can create map online as well as offline that is flexible and readily automated. Our approach to seismic hazard analysis is a starting point and is expected to be improved and refined by collecting new data.

Acknowledgements

This study was supported by the center of Excellence for Environmental Geohazards and the Research Council of Shiraz University. Two anonymous referees are gratefully thanked for useful comments.

References

Ambraseys, N.N., 2001. Reassessment of earthquakes, 1900–1999, in the Eastern Mediterranean and the Middle East. Geophysical Journal International 45(2): 471–485.

Ashtari-Jafari, M., 2010. Statistical prediction of the next great earthquake around Tehran, Iran. Geodynamics 49, 14-18.

Berg, J.W., Gaskell, R., Rinehart, V., 1964. Earthquake energy release and isostasy. Bull Seismol Soc Am 54(2): 777–784.

Bonini, M., Corit, G., Sokoutis, D., Vannucci, G., Gasperini, P., Cloetingh, S., 2003. Insight from scaled analogue modeling into the seismotectonics of the Iranian region. Tectonophysics 376: 149-157.

Bouchon, M., 1973. Effect of topography on surface motion. Bull Seismol Soc Am 63, 615–632.

Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J., 1984. Classification and regression trees (Wadsworth, Inc. Monterey, U.S.A).

Caputo, M., Milana, G., Rayhorn, J., 1984. Topography and its isostatic compensation as a cause of seismicity of the Apennines. Tectonophys 102: 333–342.

Caputo, M., Manzetti, V., Nicelli, R., 1985. Topography and its isostatic compensation as a cause of seismicity; a revision. Tectonophys, 111: 25–39.

Chen, Y.T., Liu, KR., Zheng, J.H., Song, S.H., Liu, R.F., Lu, H.Y., Gu, F.Y., 2002. A review of the studies on the relationship between local gravity field changes and earthquakes. In: Sun S (ed) Advances in pure and applied geophysics. Meteorology Press, Beijing, 40–47 (in Chinese).

Clift, P., Shimizu, N., Layne, G., 2000. Fifty five million years of Tibetan evolution recorded in the Indus fan. EOS Trans., AGU 81: 277.

Davis, L.L., West, L.R., 1973. Observed effects of topography on ground motion. Bull Seismol Soc Am., 63: 283–298.

Daubie, M., Levecq, P., Meskens, N., 2002. A comparison of rough sets and recursive partitioning induction approaches: An application to commercial loans. International Transactions in Operational Research, 9: 681–694.

Dehghani, G.A., Makris, J., 1983. The gravity field and crustal structure of Iran. In: Geodynamic Project (Geotraverse) in Iran. Geol Suev Iran, 51–68.

Efron, B., & Tibshirani, R., 1997. Improvements on cross-validation: The 632 + Bootstrap Method. American Statistical

Association, 92 (438): 548–560.

Engdahl, E. R., Jackson, J. A., Myers, S. C., Bergman, E. A., Priestley, K.,. 2006. Relocation and assessment of seismicity in the Iran region. Geophysical Journal International, 167: 761–778.

Geli, L., Bard, P.Y., Jullien, B.A., 1988. The effect of topography ground motion: a review and new results. Bulletin of the Seismological Society of America, 78 (1): 42-63.

Genti, M., Chery, J., Vernant, Ph., Rigo, A., 2016. Impact of gravity forces and topography denudation on normal faulting in Central–Western Pyrenees: Insights from 2D numerical models. Comptes Rendus Geoscience, 348: 173-183.

Han, J., Camber, K., 2006). Data Mining: Concepts and Techniques. Multisciences Press, 743pp.

Iftikhar, U. S. Toshinori, M., 2009. Application of rough set and decision tree for characterization of premonitory factors of low seismic activity. Expert System Application, 36: 102–110.

ISC, (2015). International Seismological Centre. Newbury, Berkshire, UK.

Johnston, M.J.S., 1997. Review of electric and magnetic fields accompanying seismic and volcanic activity. Surveys in Geophysics 18: 441-475.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. 14th Int. Conf. Artificial Intelligence 2(12): 1137–1143.

Lewis, R.J., 2000. An Introduction to Classification and Regression Tree (CART) Analysis. Torrance, California, Harbor-UCLA Medical Centre.

Li, Z.X., Li, H., 2009. Earthquake-Related gravity field changes at Beijing-Tangshan gravimetric network during 1987-1998 Study of Geophysics. Geodynamic 53:185-197.

McLachlan, G.J., Do, K.A., Ambroise, C., 2004. Analyzing microarray gene expression data. Wiley.

Mishra, D.C., Chandrasekhar, D.V., Singh, B., 2005. Tectonics and crustal structures related to Bhuj earthquake of January 26, 2001: based on gravity and magnetic surveys constrained from seismic and seismological studies. Tectonophysics 396: 195– 207.

Mohajer-Ashjai, A., Nabavi, M.S., 1982. Seismicity and fault map of Iran. AEOI. Scale: 1/2,500,000.

NEIC, (2015). National Earthquake Information Center. Colorado, USA.

Nowroozi, A.A., 1976. Seismotectonics Provinces of Iran. Bulletin of Seismological Society of America, 66: 1249-1276.

Reyes, J., Morales-Estebanb, A., Martínez-Álvarez, F., 2013. Neural networks to predict earthquakes in Chile. Applied Soft Computing, doi:10.1016/j.asoc.2012.10.014

Tavakoli, B., Ghafory-Ashtiany, M., 1999. Seismic Hazard Assessment of Iran. Annali DI Geofisica 42: 1013-1021.

Wu, Guijua,b., Tan, Hongboa,b., Yang, Guanglianga,b., Shen, Chongyang., 2015. Research on the relationship between geophysical structural features and earthquakes in Mid-Yunnan and the surrounding area. Geodesy and Geodynamics 6 (5): 384 -391.

Yousefi, E., 1989. Total magnetic intensity maps of Iran. Geol Suev Iran. Scale: 1/250,000.

Zamani, A., Hashemi, N., 2000. A comparison between seismicity, topographic relief, and gravity anomalies of Iranian Plateau. Tectonophys 327: 25–36.

Zamani, A., Khalili, M., 2006. Application of multivariate statistical methods for integrated mapping in Geology. 8th Iranian Statistical Conference, Shiraz University, Shiraz, Iran (in Persian).

Zamani, A., Agh-Atabai, M., 2009. Temporal characteristics of seismicity in the Alborz and Zagros region of Iran, using a multifractal approach. J Geodyn 47: 271–279.

Zamani, A., Khalili, M., Gerami, A., 2011. Computer-based self-organized zoning revisited: scientific criterion for determining the optimum number of zones. Tectonophysics, 510: 207-216.

Zamani, A., Sami, A., Khalili, M., 2012. Multivariate rule-based seismicity map of Iran: a data-driven modeling. Bulletin of Earthquake Engineering, 10: 1667-1683.

Zmazek, B., Todorovski, L., Dzeroski, S., Vaupotic, J., Kobal, I., 2003. Application of decision trees to the analysis of soil radon data for earthquake prediction. Applied Radiation and Isotopes, 58: 697-706.