Alternative models for the calculation of the RMR and Q indexes for granite rock masses

T. Miranda, A. Gomes Correia & I. Nogueira
University of Minho, Department of Civil Engineering, Guimarães, Portugal

M. F. Santos & P. Cortez
University of Minho, Department of Information Systems, Guimarães, Portugal

L. Ribeiro e Sousa
University of Porto, Department of Civil Engineering, Porto, Portugal

ABSTRACT: Empirical classification systems like the RMR and Q are often used in current practice of geotechnical structures design built in rock masses. They allow obtaining an overall description of the rock mass and the calculation, through analytical solutions, of strength and deformability parameters which are determinant in design. To be applied these systems need a set of geomechanical information that may not be available or can be difficult to obtain. In this work it is intended to develop new alternative regression models for the calculation of the RMR and Q indexes using less data than the original formulations and keeping a high accuracy level. It is also intended to have an insight of which parameters are the most important for the prediction of the indexes and in the rock masses behaviour. This study was carried out applying Data Mining techniques to a database of the empirical classification systems applications in a granite rock mass. Data Mining is a relatively new area of computer science which concerns with automatically find, simplify and summarize patterns and relationships within large databases. The used Data Mining techniques were the multiple regression and artificial neural networks. The developed models are able to predict the two geomechanical indexes using less information that in the original formulations with a good predictive capacity.

1 INTRODUCTION

Rock mass characterization is normally carried out through the application of empirical classification systems which use a set of geomechanical data and provide an overall description of the rock. Moreover, they allow obtaining other important information like support needs, stand-up time, geomechanical parameters among others.

The different classification systems have some well known drawbacks and limitations due mainly to their empirical base (Miranda, 2003). However, they are still very useful in practice therefore there is a need to improve their efficiency.

Two of the most used classification systems are the RMR-Rock Mass Rating (Bieniawski, 1989) and the Q-system (Barton et al, 1974). The RMR system is based on the consideration of six geological/geotechnical parameters. To each parameter is assigned a relative weight related to the rock mass characteristics and the final RMR value is the sum of these weights and can vary from 0 to 100. The parameters considered by this system are the following: P_1 – uniaxial compressive strength; P_2 – Rock Quality Designation (RQD); P_3 – Discontinuities spacing; P_4 – Discontinuities conditions; P_5 – Underground water conditions; and P_6 - Discontinuities orientation.

The Q-system also uses six parameters to which values have to be assigned depending on the rock mass characteristics. The final Q value is then obtained through the following expression:

$$Q = \frac{RQD}{J_n} \cdot \frac{J_r}{J_a} \cdot \frac{SRF}{J_w} \qquad (1)$$

where the different parameters are related with: J_n - number of discontinuities sets, J_r - discontinuities rugosity, J_a – discontinuities alteration, SRF – stress state and J_w – underground water.

Most of the times there are some difficulties to apply these classifications systems. Some of the data required to their application may not be available, can lack of reliability or may be difficult/expensive to obtain. Also, the considered parameters may have different importance depending on the type of rock mass being analyzed.

In this work it is intended to develop new alternative models to calculate the RMR and Q indexes for the particular case of granite rock masses which are very important in the North of Portugal. They are intended to use only the most important parameters in the behavior of granite rock masses with a good predictive accuracy.

This study was carried out using a large database of the empirical systems application in an important underground structure built in a granite rock mass. On this database Data Mining techniques were applied to obtain the new models. Multiple regression techniques and artificial neural networks (ANN) were used. The first are simpler to use and analyze and allow having an insight of which parameters are the most important in the indexes prediction while the latter are more complex and suitable for highly non-linear problems.

2 KNOWLEDGE DISCOVERY IN DATABASES AND DATA MINING

Currently, there is a great expansion of information that needs to be stored. It is important to use computational tools to explore this data which often presents high complexity and can hold valuable information such as trends or patterns that can be very useful (Goebel & Gruenwald, 1999).

In the past, two major approaches have been used for this goal: classical statistics and knowledge from experts. However, the number of human experts is limited and they may overlook important details, while classical statistic analysis does not give adequate answer when large amounts of complex data are available. The alternative is to use automated discovery tools to analyze the raw data and extract new and useful knowledge (Hand et al., 2001).

Due to the awareness of the great potential of this subject there has been an increasing interest in the Knowledge Discovery from Databases (KDD) and Data Mining (DM) fields. These terms are often confused. KDD denotes the overall process of transforming raw data into knowledge and DM is just one step of the KDD process, aiming at the extraction of useful patterns from the observed data. The knowledge derived through DM is often referred to as models or patterns and it is very important that this knowledge is both novel and understandable.

The KDD process consists in the following steps (Figure 1):

- Data selection: the application domain is studied and relevant data is collected.
- Pre-processing or data preparation: noise or irrelevant data is removed (data cleaning) and multiple data sources may be combined (data integration). In this step appropriate prior knowledge can be also incorporated.
- Transformation: data is transformed in appropriate forms for the Data Mining process.
- Data Mining: intelligent methods are applied in order to extract models or patterns.
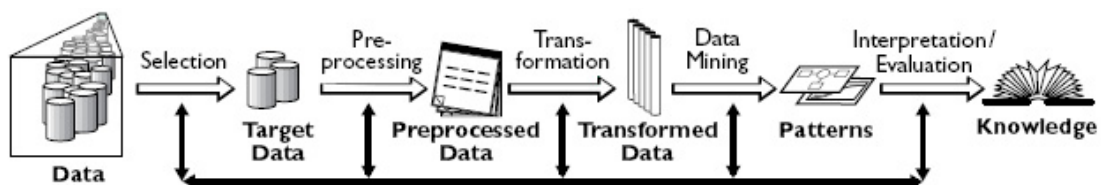- Interpretation: results from the previous step are studied and evaluated.

Figure 1 - Phases of the KDD process (Fayyad et al., 1996)

DM is a relatively new area of computer science that lies at the intersection of statistics, machine learning, data management, pattern recognition, artificial intelligence and others. DM is thus emerging as a class of analytical techniques that go beyond statistics and concerns with automatically find, simplify and summarize patterns and relationships within large data sets.

There are several DM techniques, each one with its own purposes and capabilities. Examples of these techniques include Multiple Regression Analysis, Decision Trees and Rule Induction, Neural and Bayesian Networks, Learning Classifier Systems and Instance-Based algorithms (Lee & Siau, 2001; Berthold & Hand, 2003).

3 MATERIALS AND METHODS

3.1 *The database*

The data for these models was assembled from the Venda Nova II powerhouse complex which is an important underground work recently built in the North of Portugal. The interested rock mass is a granite formation so the conclusions drawn in this study are only applicable to formations with similar characteristics. The overall process was carried out in the following steps:

- collect geotechnical data from Venda Nova II powerhouse complex;
- build and organize a database with the collected data;
- explore the data using DM techniques to induce the models.

The collected data was composed by applications of the empirical RMR and Q systems. After some data cleaning work it was then organized and structured in a database composed of 1230 examples and 21 attributes which are described in Table 1.

Table 1 – Name and description of the attributes in the original database

Name	Description
RQD	*Rock Quality Designation*
J_w	Factor related with the underground water
J_n	Factor related with the number of discontinuities sets
J_r	Factor related with discontinuities rugosity
J_a	Factor related with the weathering degree of discontinuities
SRF	Factor related with the stress state in the rock mass
Q	Rock mass quality index proposed by Barton et al. (1974)
Q'	Altered form of the Q index (Q' = RQD/J_n * J_r/J_a)
RCU	Uniaxial compressive strength
P_1	Weight related with the uniaxial compressive strength of the intact rock
P_2	Weight related with the RQD
P_3	Weight related with discontinuities spacing
P_4	Weight related with discontinuities conditions
P_5	Weight related with the underground water conditions
P_6	Weight related with discontinuities orientation
P_{41}	Discontinuities conditions – persistence
P_{42}	Discontinuities conditions – aperture
P_{43}	Discontinuities conditions – rugosity
P_{44}	Discontinuities conditions – filling
P_{45}	Discontinuities conditions – weathering
RMR	Rock Mass Rating proposed by Bieniawski (1978)

Other attributes were added to the database in order to check their possible influence on the models. Globally, 9 new attributes were added and are presented in Table 2.

The histograms of some variables presented skewed distributions (Figure 2). This fact can influence the quality of the models specially those based on neural networks since this kind of algorithm can learn better the behaviour of variables with normal distributions. This way, and after some preliminary trial calculations, it was decided to proceed to the transformation of some variables in order to maximize their normality.

Table 2 – List of attributes added to the original database

| Name	Description
RQD/J_n	Ratio which represents the compartimentation of the rock mass
J_r/J_a	Ratio which represents the shear strength of discontinuities
J_w/SRF	Ratio which represents an empirical factor named "active stress"
logQ	Base 10 logarithm of the Q value
logQ'	Base 10 logarithm of the Q' value
GSI	Geological Strength Index proposed by Hoek et al., 2002
N	Altered form of the Q index ($Q' = RQD/J_n * J_r/J_a * J_w$)
RCR	Altered form of the RMR index ($RCR = P_2+P_3+P_4+P_5+P_6$)
RCU	Uniaxial compressive strength
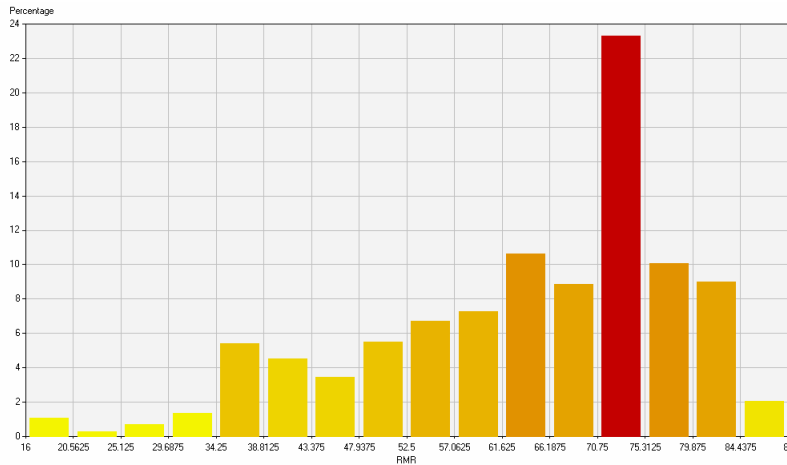
Figure 2 – Histogram of the RMR variable where is possible to observe the skewness of the distribution

The data is based on the results obtained in a granite rock mass with a good overall quality. More specifically, the main limitations that should be considered are high uniaxial compressive strength (>100 MPa), RQD values over 65% and slightly wet to dry rock mass. The models developed in this work should only be applied to rock masses with similar characteristics.

3.2 *Modelling and Evaluation*

The SAS Enterprise Miner, registered trademark of the SAS Institute Inc., was used as modelling tool. It performs DM tasks and combines statistical analysis with graphical interfaces and delivers a wide range of predictive models.

In the SAS Enterprise Miner the DM tasks are carried out programming and connecting nodes in a graphical workspace, adjust settings, and run the constructed workflow. In Figure 3 the workflow used in this work is presented.

The algorithms used for the regression models were multiple regression and ANN. The applied ANN was a multilayer feed-forward network with one hidden layer of six neurons. Focus was drawn to the multiple regression models because it was intended to obtain the explanatory physical knowledge behind the models. Moreover, these models are simpler to use and to implement. The neural network models were used mainly for comparison purposes and are an open issue for further research since it was not tried to optimize their behaviour.

In regression problems the goal is to induce the model which minimizes an error measurement e between real and predicted values considering N examples. The used error measures were the following:

$$\text{Mean Absolute Deviation: } MAD = \frac{\sum_{i=1}^{N}|e_i|}{N} \qquad (2)$$

$$\text{Root Mean Squared Error: } RMSE = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{N}} \qquad (3)$$
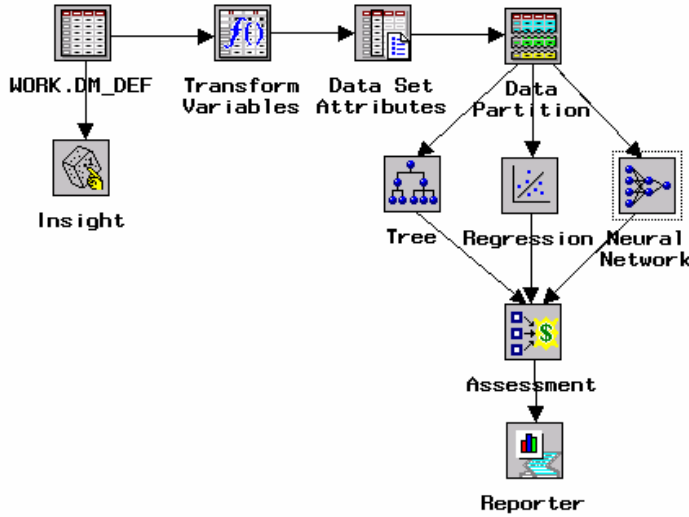

Figure 3 – Workflow used for the DM tasks

To validate and assess the models accuracy the holdout method was used. In this method data is randomly partitioned into two independent sets, a training set and a test set. In this case, 2/3 of data was used for training and 1/3 for testing. The training set is used to induce the model and its accuracy is estimated with the test set. For each model 10 runs were carried out randomizing the data within the training and testing sets. The mean and confidence intervals for the error measures were then computed considering the results of the 10 runs and a 95% confidence interval of a T-student distribution. These statistical measures define the range of expected errors for future predictions of the final model which is induced using all the data for training.

In addition to the error measures also the coefficient of determination (R^2), which is very common in many statistical applications, was used. This parameter is a measure of variability explained by the model but should not be used alone for it can lead to wrong conclusions. It varies between 0 and 1 and a value near 1 may mean that the model explains most of the data.

The regression models based on multiple regression were evaluated using the measures MAD and RMSE together with the determination coefficient (R^2). For the ANN only the RMSE was used due to computational limitations.

4 RESULTS

4.1 *RMR index*

This study started considering firstly all the variables to determine which were the most important ones in the prediction of RMR. This model itself is not relevant for prediction purposes since it uses more information than the original expression with no profit. In Figure 4, a plot of the relative importance of the main attributes for the RMR variable is presented.

As it was expected, the main parameters which influence the prediction of RMR are the ones related to its calculation even though P_1 appears only in an indirect way in the form of the unconfined compressive strength (defined as RCU in the plot). Among these parameters, the most important are the ones related with the discontinuities. In particular the parameters related with conditions (P_4) and orientation of discontinuities (P_6) are very good predictors of RMR. Moreover, in the scale of relative importance, the parameters of the Q system also related with discontinuities appear (J_n and J_r/J_a). This means that in granite formations data related to the discontinuities is a very good predictor of the overall quality of the rock.
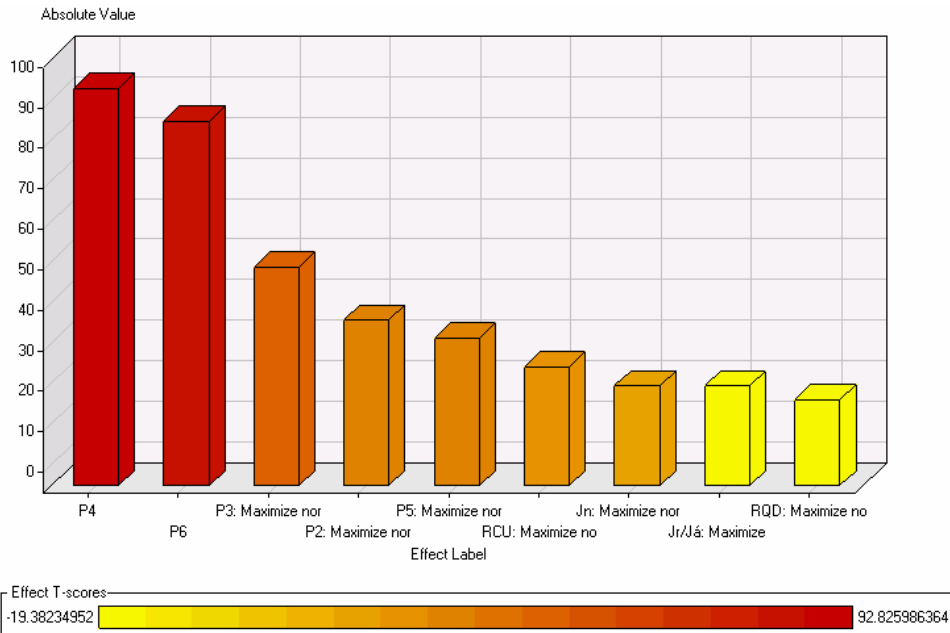
Figure 4 – Relative importance of the attributes for the prediction of the RMR variable

The next step was to induce models considering only the most important parameters: P_3, P_4 and P_6. The obtained regression model was the following:

$$RMR = 34.77 + 0.065 \times P_3^2 + 1.369 \times P_4 + 0.977 \times P_6 \qquad (4)$$

In Table 3 the results for the regression and ANN models are presented in terms of average errors and determination coefficient and correspondent T-student 95% confidence intervals. The results for the models which use all the attributes are presented only for comparison matters.

Table 3 – Results of the models for the RMR index

All attributes				P_3, P_4 and P_6			
Regression			ANN	Regression			ANN
R^2	MAD	RMSE	RMSE	R^2	MAD	RMSE	RMSE
0.995	0.650	1.094	1.070	0.944	2.565	3.522	2.857
± 0.001	± 0.050	± 0.073	± 0.070	± 0.005	± 0.083	± 0.169	± 0.114

The models which use all attributes are very accurate. The error measures are low and the de-termination coefficient is near 1. Using only the three main parameters, the error significantly increases. This is because only half of the parameters used in the original expression are ap-plied. Nevertheless, the error can be considered low for engineering purposes. Analysing the MAD and RMSE values a prediction error around 3 is expected. This means that, for instance, if a rock mass has a "real" RMR value of 65, a value within [62; 68] will be predicted which is ac-ceptable. This expression can be useful for preliminary stages of design or when only informa-tion about discontinuities is available or is reliable.

Considering the RMSE, the ANN slightly outperforms the regression models. Only for the ones with less attributes the difference can be considered significant. In this case the RMSE for the ANN is approximately 20% less than the correspondent value of the regression model. In Figure 5 the plot of real versus predicted RMR values is presented.

As it can be seen, the values lay near a 45 degree slope line which means that the prediction model shows a good accuracy. However, the deviations between real and predicted values in-crease with decreasing rock mass quality. For RMR values below 30-35 the prediction error in-creases and the model tends to overestimate the RMR. Since the model is based in the disconti-nuities characteristics this fact can be explained by the loss of importance of discontinuities for poorer rock masses.
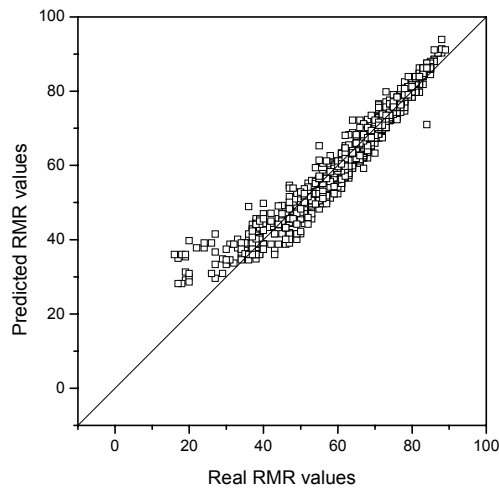
Figure 5 – Real versus Predicted RMR values for regression model with parameters P_3, P_4 and P_6

The plot of Figure 5 shows a tail with an almost quadratic trend. In order to minimize this fact a transformation of the RMR variable was performed and calculations were repeated using the squared value of RMR. The obtained regression model is presented in equation 5 and the results are resumed by Table 4 and Figure 6.

$$RMR^2 = 1036.7 + 7.148 \times P_3^2 + 166.3 \times P_4 + 116.7 \times P_6 \tag{5}$$

Table 4 – Results for the multiple regression model considering parameters P_3, P_4 and P_6 and using RMR^2

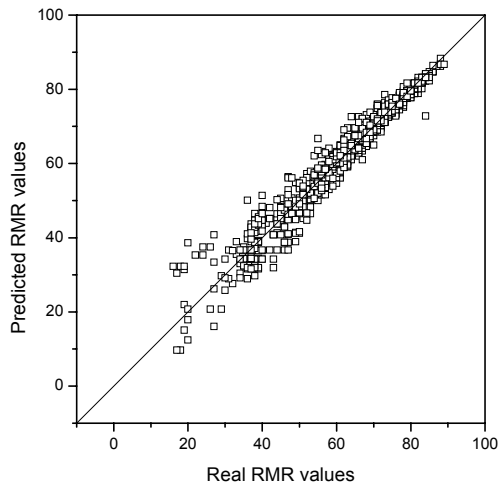| Regression		
R^2	MAD	RMSE
0.954 ± 0.004	2.179 ± 0.081	3.172 ± 0.119

Figure 6 – Real versus Predicted RMR values for regression model with parameters P_3, P_4 and P_6 and considering the transformation RMR

This transformation led to a slight reduction on the error measurements (approximately 0.4 for each) and a small increase on R^2. In Figure 6, a loss of accuracy for lower RMR values can still be observed. However, this only happens for RMR values below 30 and the overestimation trend is no longer observed has in the previous model. The points are almost equally distributed along the 45 degree slope line which means that the mean prediction error is close to 0.

As it was already referred, the Q system related parameters J_n and J_r/J_a are also important to the RMR prediction. These attributes were added to this model and calculations were again performed. However, only marginal increased performance was achieved.

4.2 *Q index*

The preliminary runs for the Q variable using all attributes indicated that the use of the base 10 logarithm of Q (logQ) led to a significant improvement of the results. Table 5 shows the results for the models which use all the attributes and the most important ones. As can be observed in Figure 7, the most important attributes are the J_r/J_a ratio and the SRF and J_n variables. This regression model is translated by the following expression:

$$\log Q = 2.00 + 0.47 \times \ln\left(\frac{J_r}{J_n \times J_a \times SRF^{1.07}}\right) \qquad (6)$$

Table 5 – Results of the models for the Q index

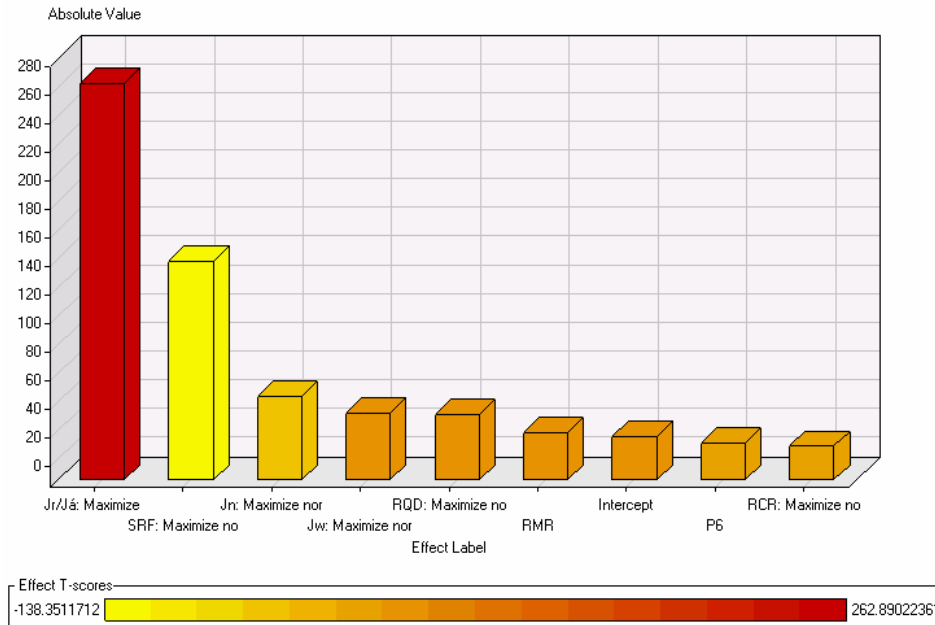| All attributes				J_r/J_a, J_n, SRF			
Regression			ANN	Regression			ANN
R^2	MAD	RMSE	RMSE	R^2	MAD	RMSE	RMSE
0.997	0.016	0.031	0.030	0.989	0.049	0.075	0.075
± 0.000	± 0.001	± 0.003	± 0.003	± 0.001	± 0.002	± 0.004	± 0.005

Figure 7 – Relative importance of the attributes for the prediction of the LogQ variable

As it happened for the RMR, the parameters related with discontinuities have a significant effect on the prediction of this quality index together with the parameter related with the stress state. This point corroborates the previous conclusion that the discontinuities characteristics are good predictors of the overall rock mass quality. Analysing the values of R^2 in Table 5 it can be seen that the values are very high for both models. The error values are low considering that the target variable ranged approximately from -1.85 to 2.13. Figure 8 shows the plot of real against predicted values and a good relation can be observed.

Since it was concluded that the parameters related with the discontinuities are very much related with both studied indexes, two more sets of variables were tested: one using only the variables J_r/J_a and J_n and other using these variables together with the parameters related with the discontinuities of the RMR system (P_3, P_4 and P_6). The latter is justified since once it is possible to obtain information for the J_r/J_a and J_n variables it is not difficult to deduce values for P_3, P_4 and P_6. The regression models are translated by equations 7 and 8 and the overall results are presented in Table 6 and in Figure 9.

$$\log Q = 2.17 + 0.57 \times \ln\left(\frac{J_r}{J_n^{1.03} \times J_a}\right) \qquad (7)$$

$$\log Q = 1.27 + 0.43 \times \ln\left(\frac{J_r}{J_n^{0.95} \times J_a}\right) + 0.0015 \times P_3^2 + 0.015 \times P_4 + 0.0094 \times P_6 \quad (8)$$
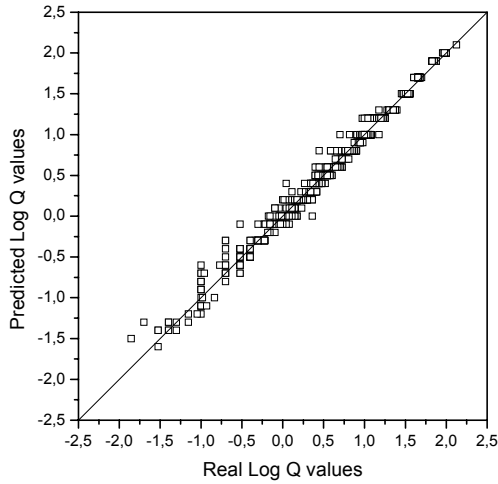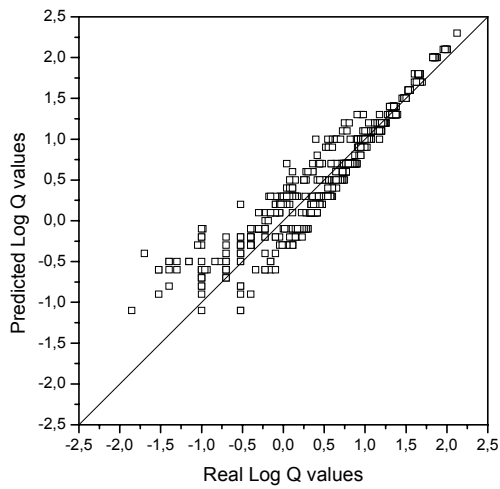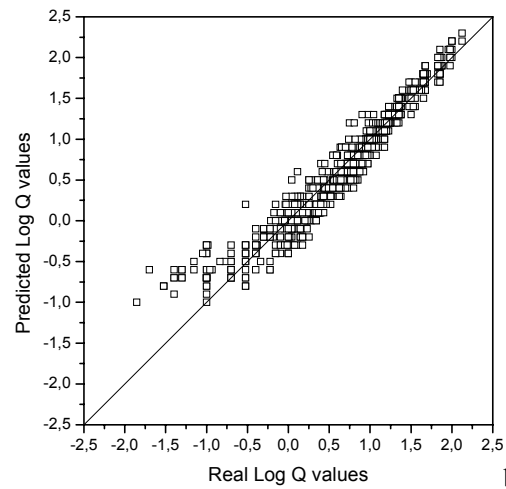

Figure 8 – Real versus Predicted LogQ values for regression model with parameters J_r/J_a, SRF and J_n

Table 6 – Results for the models considering the J_r/J_a, J_n and J_r/J_a, J_n, P_3, P_4, P_6 attributes

J_r/J_a, J_n				J_r/J_a, J_n, P_3, P_4, P_6			
Regression			ANN	Regression			ANN
R^2	MAD	RMSE	RMSE	R^2	MAD	RMSE	RMSE
0.908	0.149	0.214	0.204	0.933	0.128	0.184	0.152
± 0.009	± 0.007	± 0.013	± 0.010	± 0.005	± 0.004	± 0.009	± 0.007

Figure 9 – Real versus Predicted logQ values for regression models with: a) parameters J_r/J_a and b) J_n and J_r/J_a, J_n, P_3, P_4 and P_6

Even though the R^2 value is still within acceptable values, for the simpler model the errors significantly increase. This is especially true again for poorer rock mass conditions. Figure 9 shows high dispersion for logQ values approximately below -0.5 (Q<0.3 or RMR<35). This is also due to the loss of discontinuities importance for rock masses with low geomechanical characteristics as discussed before and shows the importance of the stress parameter consideration. The results also show that the behaviour of the models is significantly enhanced with the inclusion of the discontinuities parameters of the RMR system resulting in reduced dispersion and error values. A thorough discretization about the discontinuities minimizes the lack of information about the stress state parameter.

5 CONCLUSIONS

The empirical geomechanical classification systems like the RMR and Q are still very much used in practice. In this paper new alternative models for the calculation of these indexes were developed using data collected during the Venda Nova II hydroelectric scheme construction. The interested rock formation was a granite with good overall geomechanical quality so the induced models should only be used in rock masses with similar characteristics.

The models were developed applying DM techniques to the data previously organized and structured. DM is a recent are in computer science which uses a set of different tools from areas like machine learning and artificial intelligence among others to automatically find new and relevant knowledge from raw data.

Regression models for the RMR and Q indexes were developed using multiple regression and ANN. In all cases it was possible to induce accurate and reliable models that can be useful for practitioners and researchers using different sets of parameters. They have the advantage of using less information than the original formulations maintaining high accuracy levels. Moreover, they allowed drawing some conclusions about the physical aspects and main phenomena behind the behaviour of granite rock masses.

One interesting issue was the fact that the most important parameters for prediction were the ones related with the discontinuities. This means that in good quality granite formations this data is a very good predictor of the overall quality of the rock masses. The prediction models loose accuracy for rock formations which lay in the border between hard-soil and soft rock due to the loss of discontinuities importance.

REFERENCES

Barton, N., Lien, R., & Lunde, J. 1974. Engineering Classification of Rock Masses for the Design of Tunnel Support. *Rock Mechanics, Springer-Verlag*, Vol. 6, pp. 189-236.

Berthold, M. & Hand, D. 2003. Intelligent Data Analysis: An Introduction. *Springer*, Second Edition.

Bieniawski, Z. T. 1989. Engineering Rock Mass Classifications. *John Wiley & Sons*, 251p.

Fayyad, U.; Piatesky-Shapiro, G.; Smyth, P. 1996. From Data Mining to Knowledge Discovery: an overview. In Fayyad et al. (eds) *Advances in Knowledge Discovery and Data Mining*. AAAI Press / The MIT Press, Cambridge MA, pp 471-493.

Goebel, M. & Gruenwald, L. 1999. A survey of Data Mining and Knowledge Discovery Software Tools. *SIGKDD Explorations*, 1 (1): 20-33, ACM SIGKDD.

Hand, D.; Mannila, H.; Smyth, P. 2001. Principles of Data Mining. *MIT Press*, Cambridge, MA.

Lee, S. & Siau, K. 2001. A review of data mining techniques. *Industrial Management & Data Systems*, 101 (1): 41-46, MCB University

Miranda, T. 2003. Contribution to the calculation of geomechanical parameters for underground structures modelling in granite formations. *MCs thesis*, UM, Guimarães, 186p (in Portuguese).