Full Length Research Paper

Tunnels stability analysis using binary and multinomial logistic regression (LR)

R. Rafiee, M. Ataei and M. Kamali

Department of Mining, Oil and Geophysics Engineering, Shahrood University of Technology, Shahrood, Iran.

One of the most serious problems in tunneling projects are falling rock blocks. By considering this fact, the importance of stability predicting using some input parameters can be obviously understood. Among the existing rock mass classification systems for underground structures, rock mass rating (RMR) and Q are probably the most widely used ones this is rather unlikely to change, at least in the near future, frequently used and more available in tunneling projects, therefore establishing a proper and valid stability method based on such items would be useful. Since none of them (RMR and Q) can reflect the tunnel stability condition entirely and each has some lacks in rock mass properties defining, therefore both of them were used in this analysis which can provide the whole perspective of rock mass condition and stability. For this aim, data (RMR, Q, and hydraulic radius) from 104 cases of eight tunnel projects were gathered. By observing the stability condition in each tunnel, the data were classified in three categories: stable, potentially unstable and unstable. Two models next were defined and the related formulas were found using binary and multinomial logistic regression, at last the best predictor model would be selected by using the percent of correctly predicted cases in each model. The results of this paper show that the logistic regression (LR) is a robust tool to establish and develop predicting model for tunneling projects and can assist engineers to predict the stability condition of tunnels.

Key words: Logistic regression, tunneling, rock mass classification, stability.

INTRODUCTION

Rock falls and collapses in excavation projects result in considerable fatalities and damages in man sources and machinery. So for long time, many stability analyzes have been considered, and a lot of studies were conducted to predict and control these events in tunnels, stopes, pillars and, as a result different stability graphs were established. For example, The Mathews stability graph method for open-stope design was first proposed for mining at depths below 1000 m (Mathews, 1980), and during years it has been modified by Potvin (1988), Stewart and Forsyth (1995), Nickson (1992), and

Hadjigeorgiou et al. (1995) for example the modified Mathews stability graph by Stewart and Forsyth was shown in Figure 1. Other different stability graphs have been developed during years to estimate the stability condition and the probability of failure for stopes, pillars, tunnels such as; Laubscher's caving stability graph (Laubscher, 1990). Barton et al. (1974) stability graph method based on the NGI tunneling index Q, pillar stabilitygraph developed by Lunder (1994) and modified by Mah (1995).
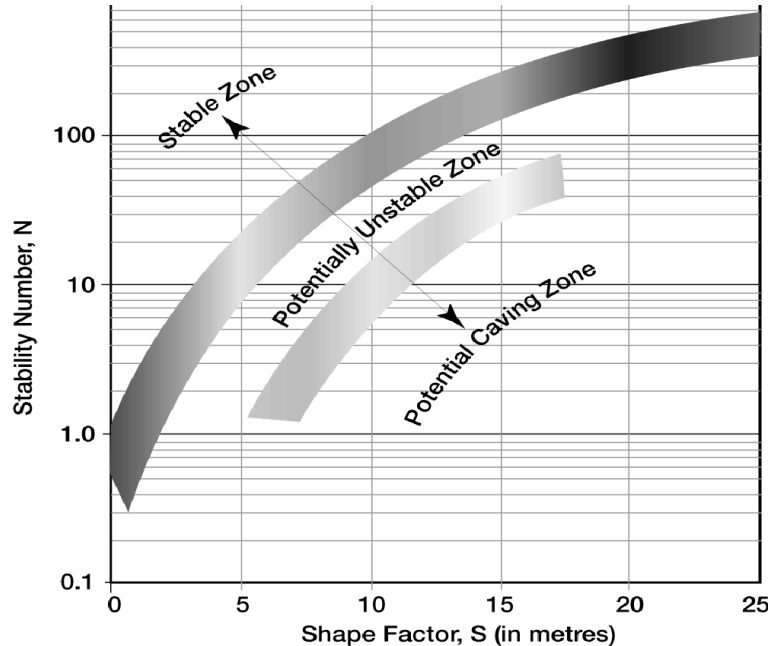
The purpose of this paper is to develop a stability

*Corresponding author. E-mail: raminlamezi@gmail.com.

Figure 1. Three stability zones of original Mathews's stability graph modified by Stewart and Forsyth (1995).

graph and method especially for tunnels based on both RMR and Q rock mass classification systems, because the two widely-used and available systems in tunnels are the Norwegian Geotechnical Institute's Q system, Barton et al. (1974) and the various versions of the rock mass rating (RMR) system, originally proposed by Bieniawski (1973). Interestingly, both systems trace their origin in tunneling. After all each of these rock mass classification systems has some lacks which cause not being able to reflect the tunnel stability condition entirely. For example the Q system does not take the rock material strength into account explicitly, although it is implicitly included in arriving at the stress reduction factor (SRF). The SRF in Q system may be significant depending on the depth. The orientation is also not taken into account. The RMR does not take account of the confining stress in the rock mass, nor explicitly the number of joint sets. Since both rock quality designation (RQD) and joint spacing are classification parameters considerable weight is given to block size and also the weighting of parameters in these two rock mass classification systems are different (Barton et al., 1974; Barton, 1991; Bieniawski, 1973, 1976, 1989).

Anyway, these two rock mass rating systems, that is, RMR and Q, were considered as candidates for assessing geotechnical conditions of these eight tunnels. But as mentioned just above, none of these systems adequately can describe all the observed geotechnical conditions and failure mechanisms. The final conclusion was that a hybrid system would provide the best results.

For this reason, both of these systems (RMR and Q) were used in the models. The two models including the two rock mass classifications were made using logistic regression, in which the probability of stability can be predicted by entering RMR, Q and hydraulic radius (R_h) of each case as input variables given in Table 1. The hydraulic radius of tunnel is defined as below:

$$\text{Hydraulic radius } (R_h) = \frac{S}{P} \tag{10}$$

Where S is the tunnel cross-section area and P is the tunnel cross-section perimeter. At last the best model was selected to estimate and classify the tunnel conditions using statistical result of each model.

LOGISTIC REGRESSION (LR) ANALYSIS

Logistic regression is useful for situations in which the purpose is to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables (independent). It is similar to a linear regression model but is suited to models where the dependent variable is discrete and discontinuous. Logistic regression is applicable to a broader range of research situations than discriminant analysis. For example, in mining engineering fields, the logistical regression analysis was used by Trueman et al. (2000) to

Table 1. Data of eight tunneling projects.

No.	RMR	Q	R_h (m)	Stability condition	No.	RMR	Q	R_h (m)	Stability condition
1	45	1	2.7	S	53	11	0.024	4.5	U
2	50	1.9	1.75	S	54	14	0.049	4.5	U
3	25	0.12	2.9	P	55	48	1.33	2	S
4	72	22.4	6.3	S	56	55	1.8	2	S
5	45	1	6.5	S	57	38	0.08	2	U
6	37.5	0.48	3.6	P	58	46	2.6	2	S
7	27.5	0.21	3.6	P	59	30	0.16	2	U
8	22.5	0.09	3.6	U	60	42	0.09	2	U
9	42.5	0.84	3.6	S	61	64	16.72	6	S
10	14	0.049	4.5	U	62	41.5	0.33	6	U
11	11	0.049	4.5	U	63	31	0.24	2.3	U
12	14	0.024	4.5	U	64	51	2.2	2.3	P
13	11	0.024	4.5	U	65	53	7.42	3.45	S
14	17	0.049	4.5	U	66	60	5.28	4	S
15	35	0.33	4.5	U	67	58	4	2	S
16	43	0.89	3.6	S	68	56	4.2	2	S
17	35	0.36	3.6	P	69	65	2.7	2	S
18	47.5	1.47	3.6	S	70	44	0.99	2	P
19	22.5	0.17	3.6	P	71	41	1.33	2	S
20	32.5	0.27	3.6	P	72	29	0.44	2	U
21	17.5	0.05	3.6	U	73	61	6.3	2	S
22	37.5	0.048	3.6	U	74	52	2.7	2	S
23	22.5	0.09	3.6	U	75	53	4.1	2	S
24	27.5	0.27	3.6	S	76	51	2.72	2	S
25	53	4.1	2	S	77	51	2.2	2	S
26	52	2.7	2	S	78	36	0.41	2	U
27	61	6.3	2	S	79	48	1.56	2.5	P
28	58	4	2	S	80	28	0.17	2.5	U
29	56	1.4	2	S	81	76	6.25	6	S
30	41	0.33	2	P	82	54	6.27	2.25	S
31	32	0.33	4.5	U	83	41	1.67	2.25	P
32	32	0.66	4.5	U	84	21	0.088	2.25	U
33	39	0.44	4.5	U	85	47	1.32	2.25	P
34	42	1.65	4.5	U	86	30	0.55	2.25	U
35	39	1.44	4.5	U	87	34	0.55	2.25	U
36	42	1.65	4.5	U	88	36	0.66	2.25	U
37	46	1.2	4.5	U	89	54	6.6	2.25	S
38	47	1.98	4.5	U	90	24	0.066	2.25	U
39	44	2.145	4.5	U	91	32.5	0.27	3.5	P
40	44	1.98	4.5	U	92	27.5	0.15	3.6	P
41	44	3.96	4.5	U	93	37.5	0.48	3.6	S
42	47	3.96	4.5	S	94	32.5	0.27	3.5	P
43	45	4.29	4.5	U	95	22.5	0.15	3.6	P
44	45	6.16	4.5	S	96	22.5	0.09	3.6	U
45	45	7.48	4.5	S	97	37.5	0.48	3.6	S
46	57	7.66	4.5	S	98	22	0.21	3.6	P
47	66	3.96	4.5	S	99	18.5	0.05	3.6	U
48	69	7.04	4.5	S	100	17.5	0.05	3.6	U
49	69	4.62	4.5	S	101	47.5	1.47	3.6	S
50	55	4.62	4.5	S	102	45.5	1.18	3.6	S

Table 1. Contd.

| 51 | 60 | 4.62 | 4.5 | S | 103 | 42.5 | 0.84 | 3.6 | S |
| 52 | 58 | 4.62 | 4.5 | S | 104 | 37.5 | 0.48 | 3.6 | S |

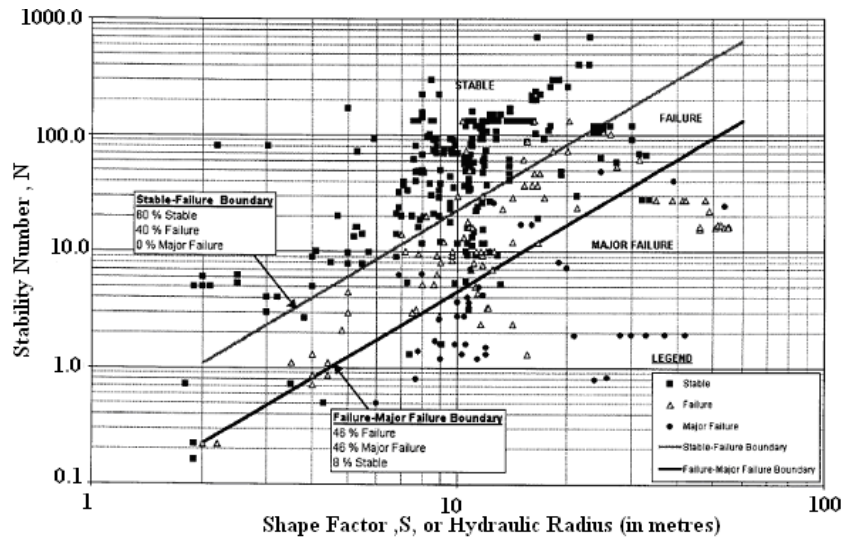R_h = Hydraulic radius, S = Stable, P = Potentially-unstable, U= Unstable.

Figure 2. Extended Mathews stability graph based on logistic regression (Mawdesley, 2001).

analyze databases of stable and collapsed slopes from Australian mines or by C. Mawdesley, R. Trueman and W. J. Whiten to extend the Mathews stability graph which is given in Figure 2 (Mawdesley et al., 2001) and also Molinda et al. (2000), performed simple regression analysis with some significant geotechnical variables like overburden, bolt strength, bolt capacity, grout length, density, entry width, coal mine roof rating (CMRR) and intersection span for predicting the roof fall rate. This type of analysis is appropriate to data sets with a binary or multinomial dependant variable and a number of numerical independent variables. Generally, logistic regression analysis has two basic forms; a) binary, b) multinomial that a sort summary of them have been brought as following:

Binary logistic regression

Binary logistic regression is most useful when the aim is to model the event probability for a categorical response variable (dependent) with two outcomes (dichotomous). For example: An engineer wants to know if a particular excavation with input parameters such as geotechnical properties of rock mass would fall in stable category or unstable. Since the probability of an event must lie between 0 and 1, it is impractical to model probabilities with linear regression techniques, because the linear regression model allows the dependent variable to take values greater than 1 or less than 0. The logistic regression model is a type of generalized linear model that extends the linear regression model by linking the range of real numbers to the 0 to 1 range. The general form of the logistical regression described by Charles (2002), is shown in Equation 2.

$$pr = \frac{1}{1 + e^{-Z}}$$

(2)

Where pr is the probability for the category with logit value 1, $Z = b_0 + b_1 x_1 + ... + b_k x_k$, k = Number of independent variables, x_i =independent variables (predictors), b_i = regression coefficients which are estimated by maximum likelihood method suggested by Cox (1970).

Multinomial logistic regression

Multinomial logistic regression is useful for situations in

Table 2. Parameter estimations and statistical summery.

Stability condition	Parameter	b_{ji}	Std. error	95% Confidence interval for Exp (B)	
	Intercept	7.736	2.687	Lower bound	Upper bound
0.00 (unstable)	RMR	-0.209	0.060	0.722	0.912
	Q	-0.125	0.239	0.552	1.409
	R_h	0.328	0.294	0.780	2.471
	Intercept	6.988	2.976		
	RMR	-0.147	0.068	0.756	0.986
0.50 (potentially-unstable)	Q	-0.606	0.518	0.198	1.507
	R_h	-0.211	0.353	0.405	1.619

which the purpose is to classify subjects based on values of a set of predictor variables. This type of regression is similar to binary logistic regression, but is more general because the dependent variable is not restricted to two categories. For example; an engineer wants to know if a particular excavation with input parameters such as geotechnical properties of rock mass would fall in stable category, potentially-unstable or unstable categories.

$$pr_i = \frac{e^{Z_i}}{e^{Z_1} + e^{Z_2} + ... + e^{Z_k}}$$

(3)

Where Pr_i is the probability of falling each case in category i, $Z_i = b_{0i} + b_{1i}x_1 + \cdots + b_{ki}x_k$ (k= Number of independent variables, X_i: independent variables (predictors), b_{ji}: regression coefficients for category i). As it stands, if you add a constant to each Z, then the outcome probability is unchanged. This is the problem of non-identifiably. To solve this problem, z_k is (arbitrarily) set to 0. The kth category is called the reference category, because all parameters in the model are interpreted in reference to it.

DATA ANALYSIS

In this paper the input parameters are RMR, Q and hydraulic radius (R_h) which are the independent variables (predictors). The dependent variable (response) is the stability condition which is a categorical variable divided into three categories; stable, potentially-unstable and unstable. These variables for each tunneling case were given in Table 1. Since the dependent variable (stability condition) has categorical nature, using logistic models is so practical therefore to well-defined models were considered as following and their applicability would be evaluated.

Model 1

In this model all cases have been classified into three categories namely; stable, potentially-unstable, unstable. The three logit values 1, 0.5 and 0 are assigned to stable, potentially-unstable and unstable categories, respectively. The nature of this model show that it should be analyzed by multinomial logistic regression, therefore for this aim, the SPSS statistical software was used and the parameter estimations and statistical summary results have been summarized in Table 2 in which b_{ji} are the regression coefficients in $Z_i = b_{0i} + b_{1i}x_1 + ... + b_{ki}x_k$ of Equation (3). In this analysis the reference category is stable (logit value 1.00). There are several ways to estimate the logistic model accuracy and reliability but among them using the classification results or goodness of fit is so usual. The goodness-of-fit results are given in Table 3. The reliability and goodness-of-fit table shows the practical results of using the multinomial logistic regression model in classifying the 104 cases. Cells on the diagonal are correct predictions and cells off the diagonal are incorrect predictions. For example, this table shows that there are 47 stable cases, of which 42 cases have been classified correctly into stable category and the 5 remained cases were incorrectly classified into the unstable category by the model. After all, the total percent of correctly classified cases in this model is 71.2%. However, the overall percentage is 71.2% but it can be seen that about potentially-unstable category the correct classification percent is not satisfactory because out of 17 potentially-unstable cases only one case was classified correctly and 11 cases were fell into the unstable category and the 5 remained cases have been classified into the stable category. Two important results could be extracted from Table 3, are as follows:

(1) This model doesn't classify the potentially-unstable cases satisfactorily and it is not a proper and reliable model for this category.

Table 3. Reliability and goodness-of-fit for model 1.

Observed	Predicted			
	0.00	0.50	1.00	Percent correct
0.00	31	2	7	77.5
0.50	11	1	5	5.9
1.00	5	0	42	89.4
Overall percentage	45.2	2.9	51.9	71.2

Table 4. Parameter estimations.

Parameter	R_h	Q	RMR	Constant
b_i	-0.140	0.222	0.192	-8.320

Table 5. Reliability and goodness-of-fit for model 2.

Observed		Predicted		
		Stability condition		Percentage correct
		0.00	1.00	
Stability condition	0.00 (unstable)	49	8	86.0
	1.00 (stable)	9	38	80.9
Overall percentage				83.7

(2) In order to integrate the potentially-unstable category with any of stable and unstable categories, it should be added to the unstable category because the logistic model tends to classify its cases into the unstable category.

By using the presented Formula (3), the formulas of this model are as following:

$$Z_1 = 0 \tag{4}$$

$$Z_{0.50} = 6.988 - 0.147\,RMR - 0.606Q - 0.211R_h \tag{5}$$

$$Z_{0.00} = 7.736 - 0.209RMR - 0.125Q + 0.328R_h \tag{6}$$

$$probabilit \quad y_{unstable} = \frac{e^{7.736 - 0.209\ RMR\ -0.125\ Q + 0.328\ R_h}}{1 + e^{7.736 - 0.209\ RMR\ -0.125\ Q + 0.328\ R_h} + e^{6.988 - 0.147\ RMR\ -0.606\ Q - 0.211\ R_h}} \tag{7}$$

$$probabilit \quad y_{p-unstable} = \frac{e^{6.988 - 0.147\ RMR\ -0.606\ Q - 0.211\ R_h}}{1 + e^{7.736 - 0.209\ RMR\ -0.125\ Q + 0.328\ R_h} + e^{6.988 - 0.147\ RMR\ -0.606\ Q - 0.211\ R_h}} \tag{8}$$

$$probabilit \quad y_{stable} = \frac{1}{1 + e^{7.736 - 0.209\ RMR\ -0.125\ Q + 0.328\ R_h} + e^{6.988 - 0.147\ RMR\ -0.606\ Q - 0.211\ R_h}} \tag{9}$$

In these formulas, $Z_1 = 0$ because the reference category is stable with logit value 1.0.

Model 2

Next, the potentially-unstable and unstable categories were taken into one category called unstable, so there are two categories stable and unstable, this is a kind of binary problem. In this model, the two logit values 1, and 0 are assigned to stable and unstable categories, respectively. After analyzing by SPSS software the Tables 4 and 5 results were achieved. In Table 4, b_i are the regression coefficients in $Z = b_0 + b_1x_1 + ... + b_kx_k$ of Equation (3). From goodness of fit table, it can be seen

Table 6. Regression analysis summary and parameter estimates.

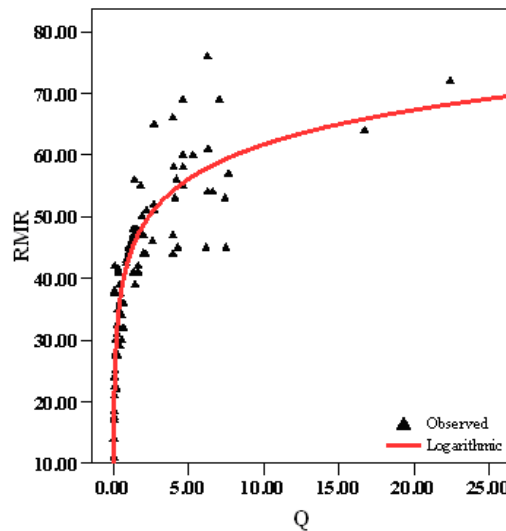| Equation	R Square	Constant	B (coefficient)
Linear	0.456	34.159	3.080
Logarithmic	0.835	43.082	8.090
Exponential	0.331	31.841	0.077

Figure 3. Correlation between RMR and Q rock mass classifications.

that out of 57 unstable cases, 49 cases were classified correctly, totally 86.0% of all unstable cases are classified correctly and just 8 cases have been classified wrongly. About stable category, we can see that out of all 47 stable cases, 38 cases were classified correctly and the total correct percentage is 80.9 which are reliable and satisfactory. Overall percentage of correctly classified cases in this model is 83.07. In this model any of categories have been classified satisfactorily and yet the total percent is reliable. By using the Formula (2), the related formulas are as following:

$$probabilit\ y_{stable} = \frac{1}{1 + e^{-z}} \tag{10}$$

$$Z = -8.320 + 0.222\ Q + 0.192\ RMR - 0.140\ R_h \tag{11}$$

$$probability_{stable} = \frac{1}{1 + e^{-(-8.320 + 0.222Q + 0.192RMR - 0.140R_h)}} \tag{12}$$

$$probabilit\ y_{unstable} = \frac{e^{-(-8.320 + 0.222Q + 0.192RMR - 0.140R_h)}}{1 + e^{-(-8.320 + 0.222Q + 0.192RMR - 0.140R_h)}} \tag{13}$$

An example

Consider a tunnel with these rock mass properties as: RMR = 45, Q= 6.16 and R_h = 4.5 m. The model 2 is used to predict the stability in this tunnel, the results are as following: Probability of stability=74% and probability of instability=26%. Since the stability probability is greater than the instability probability this case falls in stable category, field observation shows that this case is stable as obtained from analysis.

In order to establish curves between RMR and hydraulic radius (R_h) or Q and hydraulic radius (R_h), the relationship between these two rock mass classifications should be found. To find the relationship between RMR and Q rock mass classifications in these tunnels, three forms of regression were done between them which their results are given in Table 6. The results of this table show that the logarithmic regression is the best equation between them. The curves of three regression forms are shown in Figure 3.

$$RMR = 8.090 LnQ + 43.082 \tag{14}$$
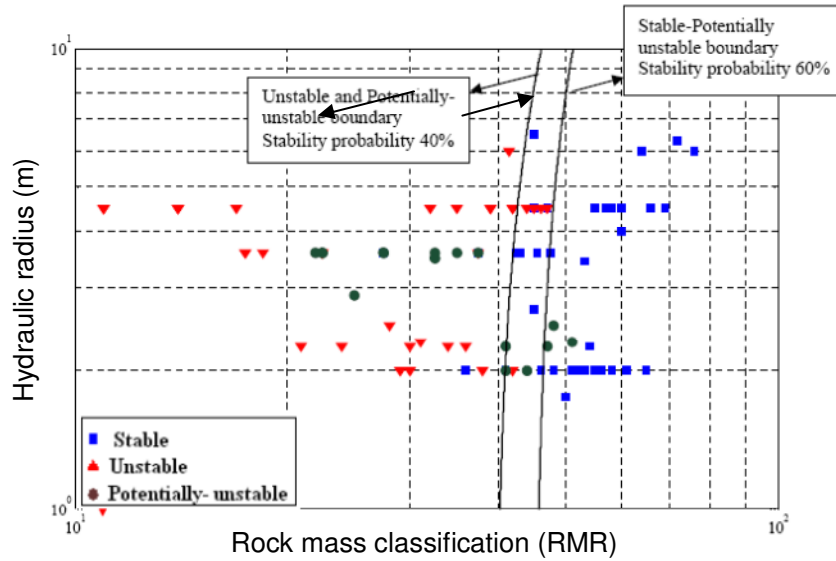
By using the Equation (14), the RMR or Q can be

Figure 4. Stable and potentially unstable and unstable regions.
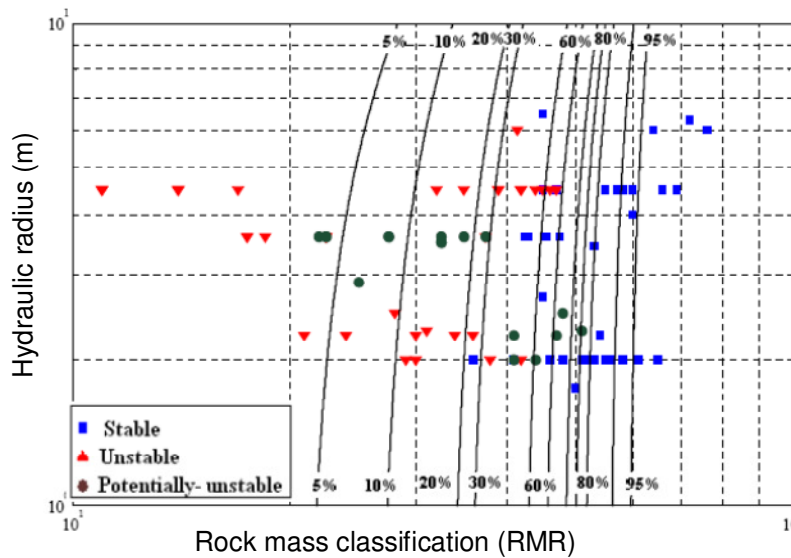

Figure 5. Isoprobability contours for stable tunnel based on logistic regression.

substitute by each other, in the probability formulas. Then the equation between RMR (and Q), R_h and probability of stability (or instability) could be obtained. And then the curves RMR versus R_h or Q vs. R_h can be drawn. By experience, engineering judgment and field observations the stable and potentially-unstable boundary is a line in which probability of stability is 60% and instability is 40% and also the boundary line between unstable and potentially-unstable is where the instability probability is 60%, on the other word in the stability probability is 40%.

These three different zones are shown in Figures 4 and 7. The isoprobability contour figures for two stable and unstable categories have been given in Figures 5, 6, 8 and 9.

CONCLUSION

Due to the increasing number of tunnels which are being developed, this paper focused on the evaluation of eight tunnels stability and establishing a stability graph based on logistic regression models and rock mass classi-
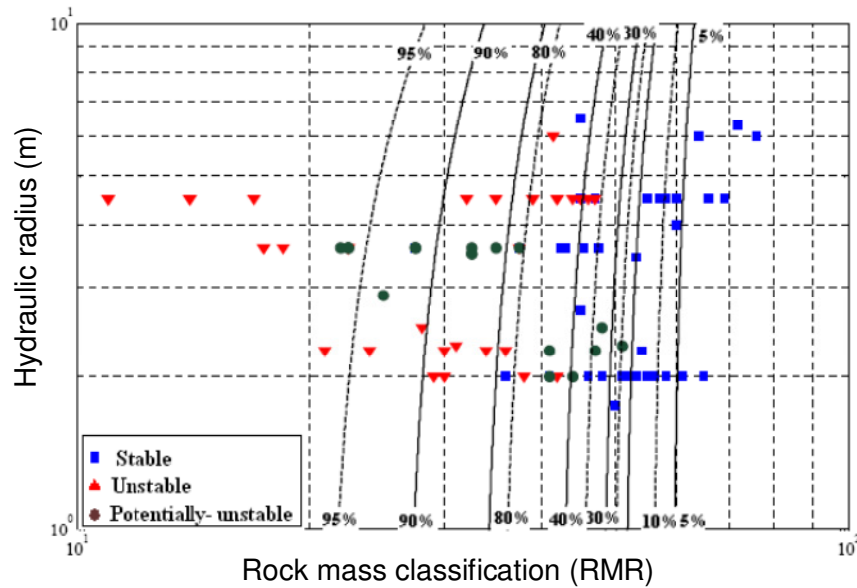
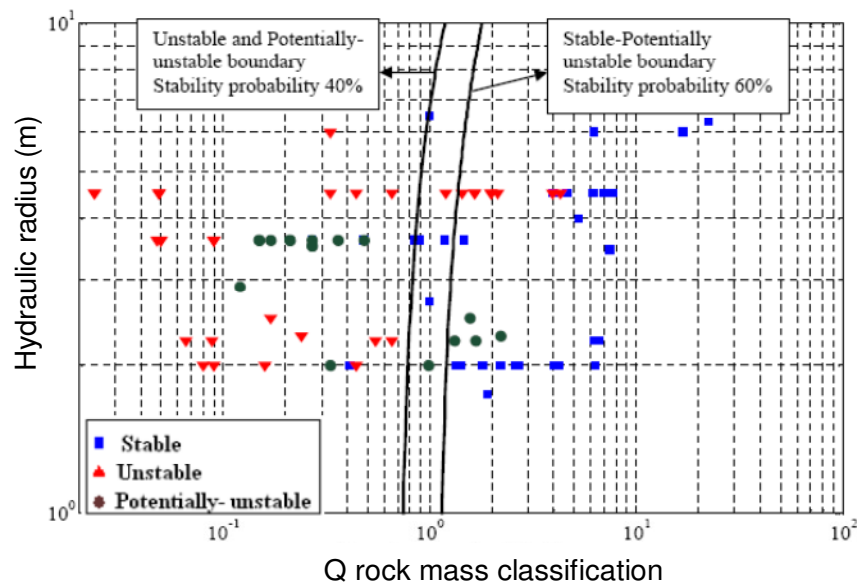**Figure 6.** Isoprobability contours for unstable tunnel based on logistic regression.

Figure 7. Stable and potentially unstable and unstable regions.

fication systems. For this aim the data of 104 cases form eight developing tunnels including RMR, Q, hydraulic radius and stability condition, were gathered and then two well-defined logistic models were considered. After analyzing with these models, it was understood that considering all the cases into two separate stable and unstable categories using binary form is so satisfactory. This model classifies cases 83.7% correctly. From engineering judgment point of view, this precision is

adequate and reliable.

To plot the graphs of RMR vs. hydraulic radius or Q vs. hydraulic radius and the isoprobability contours, some regression models between data of RMR and Q were done and the best equation between them was obtained. By using this relationship the related curves have been drawn. The results of this paper show that the LR is a reliable and proper tool to establish and develop the predicting model for tunneling projects and can assist
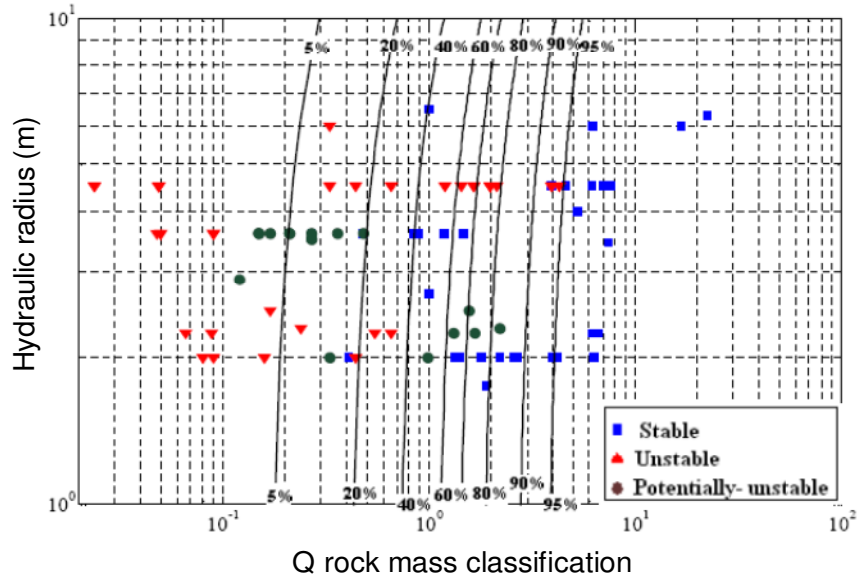
Figure 8. Isoprobability contours for stable tunnel based on logistic regression.
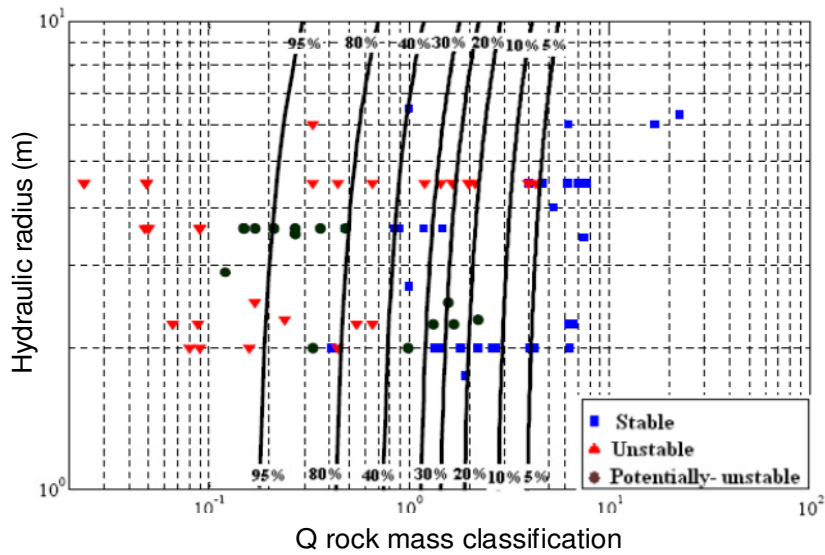

Figure 9. Isoprobability contours for unstable tunnel based on logistic regression.

engineers to predict the stability condition of tunnels. The approach presented in this paper has been employed as a practical tool for decision making process in order to arrive at a workable solution.

REFERENCES

Barton N (1991). "Geotechnical Design" WT Focus. World Tunneling.
Barton N, Lien R, Lunde J (1974). "Engineering classification of rock masses for the design of tunnel support". Rock Mech. pp.189-236
Bieniawski ZT (1973). "Engineering classification of jointed rock mass" Trans South Afr. Ins, pp. 335-344.

Bieniawski ZT (1976). "Rock mass classification in rock engineering applications," Proc. Symp. On Exploration for rock eng., Balkema, Rotterdam 12:97-106.

Bieniawski ZT (1989). "Engineering Rock Mass Classifications", John Willy & Sons.

Charles MF (2002). "Logistical regression" (on line),www.shsu.edu/~icc_cmf/ newCJ742/6LOGITnew.doc.

Cox DK (1970). "The analysis of binary data". Chapman and Hall, London.

Hadjigeorgiou J, Leclair J, Potvin Y (1995). "An update of the stability graph method for open stope designs". Paper presented at CIM Rock Mechanics and Strata Control session, Halifax, Nova Scotia.

Laubscher DHA (1990). "Geomechanics classification system for the rating of rock mass in mine design". Jl S. Afr. Inst. Min. Metall. 257-73.

Lunder P (1994). "Hard Rock Pillar Strength Estimation: An Applied Empirical Approach". MASc Thesis, University of British Columbia, p.166.

Mah P (1995). "Development of Empirical Design Technques in Burst Prone Ground at the A.W. White Mine", CANMET Project No: 1-9180

Mathews KE (1980). "Prediction of stable excavation spans for mining at depths below 1000 m in hard rock". Report to Canada Centre for Mining and Energy Technology (CANMET), Department of Energy Resource.

Mawdesley C, Trueman R, Whiten WJ (2001). "Extending the Mathews stability graph for open-stope design, Paper published in Trans". Instn Min. Metall. January–April 2001.

Molinda GM, Mark C, Dolinar D (2000). "Assessing coal mine roof stability through roof fall analysis". Information Circular, Center for Disease Control and Prevention, NIOSH.

Nickson SD (1992). "Cable support guidelines for underground hard rock mine operations". M.App.Sc. thesis, Uni of British Columbia.

Potvin Y (1988). "Empirical open stope design in Canada", Ph.D. thesis, Uni of British Columbia.

Stewart SBV, Forsyth WW (1995). "The Mathews method for open stope design". CIM Bull.pp. 45-53.

Trueman R, Mikula P, Mawdesley C, Harries N (2000). "Experience inAustralia with the application of the Mathews' method for open stopedesign".CIM Bulletin, pp.162-167.