https://doi.org/10.1016/j.ijrmms.2022.105306

Statistical study of squeezing for soft rocks based on factor and regression analyses of effective parameters

Mohammadreza Akbariforouz[1,2*], Qi Zhao[2,*], Kewei Chen[3], Alireza Baghbanan[4], Roohollah Narimani Dehnavi[5], Chunmiao Zheng[1,3]

[1]EIT Institute for Advanced Study, Ningbo, China;
[2]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China;
[3] State Environmental Protection Key Laboratory of Integrated Surface Water-Groundwater Pollution Control, School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China;
[4] Department of Mining Engineering, Isfahan University of Technology (IUT), Isfahan, Iran;
[5]Mining Engineering and Mineral Economics, the University of Leoben, Leoben, Austria;

* Correspondences: mo2022.akbariforouz@connect.polyu.hk
qi.qz.zhao@polyu.edu.hk

Abstract: The time-dependent deformation of rocks due to stress released by excavation is referred to as squeezing. Accurate evaluation of the squeezing at the design stage can dramatically reduce technical problems and the financial costs of underground structures. Although various methods are presented to predict tunnel squeezing at the preliminary stage, being site-specific and incorporating incomplete databases are deficiencies of the available procedures. In this study, based on a comprehensive literature review, we prepared a database of tunnel squeezing for soft rocks, including possible effective parameters. Statistical processing methods such as univariate, reduction, and cleaning were employed to improve the statistical quality of the database. The statistically-processed datasets were also validated based on various scales such as accuracy, convergence, and usefulness. Significant predictors of squeezing are recognized as the ratio of strength to stress and the rock mass classification system. New squeezing criteria were developed using binary and multi-class regression methods to predict the squeezing occurrence and intensity of soft rocks. The results are confirmed by a Multilayer Perceptron Feed-Forward Neural Network and are compared to well-known empirical equations. The developed equations are more accurate comparing the empirical equations used to predict the squeezing of soft rocks. This methodology can be utilized at the design stage for another database to predict squeezing rocks for topographic-stress and tectonic-stress-based cases.

Keywords: squeezing; statistical processing methods; regression analyses; neural network; empirical equations

List of symbols

Q	Norwegian rock mass classification index	-
H	Overburden depth	m
Q_n	Modified Norwegian rock mass classification index	-
B	Tunnel span	m
K	Support stiffness	MPa
σ_v or P_0	Vertical stress (γH)	MPa
SSR or N_C	Strength stress ratio	-
γ	Unit weight of rock mass	MN/m^3
σ_{cm}	Rock mass uniaxial compressive strength	MPa
u_a	Tunnel convergence	m
a	Tunnel radius	m
ε_{cr}	Critical strain	%
RMi	Rock Mass index	-
$\sigma'_{critical}$	The critical effective vertical/horizontal stress for squeezing	-
σ'	The in situ vertical/horizontal effective stress	-
σ_θ	Tangential stress	MPa
FEM	Finite element method	-
FDM	Finite difference method	-
p_i	Support pressure	MPa
SVM	Support vector machine	-
ε	Strain (normalized convergence)	%
RQD	Rock quality designation	%
γ_{dry}	Dry intact rock unit weight	MN/m^3
γ_{sat}	Saturated intact rock unit weight	MN/m^3
UCS	Uniaxial compressive strength of the intact rock	MPa
GSI	Geological strength index	-
σ_t	Rock mass tensile strength	MPa
C	The cohesion of rock mass	MPa
E	Young’s modulus	GPa
ϕ	Friction angle of rock mass	°
ν	Poison ratio of rock mass	-
SRF	Stress reduction factor	-
BQ	Basic quality classification system	-
ε_θ^a	Peak tangential strain	%
ε_θ^e	Peak elastic strain	%
NPCA	Nonlinear Principal Component Analysis	-
CDF	Cumulative distribution function	-
p	Class probability of normal distribution	-
ϕ	The standard deviation of the normal distribution	-
OR	Odds ratio	-
i	The index of the class	-
γ_i	Cumulative probability of i^{th} class	-
MLE	Maximum likelihood estimation	-
MMSE	Minimum mean square error	-
β and βi	Coefficients	-
Link	Regression link function	-
Y	Value of the qualitative dependent parameter	-
Y'	Continuous latent variables	-
k	Number of predictors	-
β_0	Constant value	-
x_i	The predictors	-
SE	Standard error	-
r and c	The number of rows and columns, respectively	-
n	The number of parameters	-
$L(\hat{B})$ and $L(B^{(0)})$	log-likelihoods with and without predictor parameters	-
MLP-NN	Multilayer Perceptron Feed-Forward Neural Network	-
St	Statistic	-
df	Degree of freedom	-
Si	Significance	-

1. Introduction

Squeezing is defined as time-dependent deformations of rocks because of excavation-induced stresses that can impose extreme pressures on the support system of underground structures (Swannell et al., 2016). This phenomenon is related to creep triggered by high shear stresses (Chern et al., 1998). Squeezing can be continued even after the construction phase (Barla & Barla, 2008). Tunnel squeezing is well-known for relatively deep tunnels excavated in soft rocks. However, topographic or/and tectonic factors can activate over-time deformations that may affect weak rocks excavated at a shallow depth (Shrestha & Broch, 2008) or hard rocks at a deep depth (Malan, 1999; Hadjigeorgiou and Karampinos, 2017).

Even though squeezing rocks can be classified as soft (henceforth, tunnel squeezing in this paper) or hard rock, a specific criterion to recognize the squeezing of the hard rock and the soft rock is absent in the literature. Squeezing of hard rocks (e.g., quartzite and hard lava) is recorded at gold-bearing reefs of South Africa with uniaxial compressive strength of 200 and 400 Mpa, respectively (Malan, 1999).

Despite the tremendous progress in predicting squeezing during the past decades, some obstacles still exist during the design stage and construction phases of tunneling projects. The lack of accurate methods for evaluating the occurrence and intensity of squeezing rocks makes the proper design of the excavation procedure and support installation a challenge. Common predictive approaches contain some intrinsic disadvantages that can result in unrealistic results, e.g., some methods present a general, universal equation based on a database collected from a specific site condition (Panthi & Nilsen, 2007).

Squeezing predictive criteria are usually developed based on a database containing some (but not all) effective parameters. What are the effective parameters, i.e., the squeezing predictors, is still a million-dollar question. Moreover, many classification systems are binary, i.e., these methods can predict two conditions: squeezing and no-squeezing (Goel et al., 1995; Jimenez & Recio, 2011; Shafiei et al., 2012). Using a database with insufficient datasets may also affect the final results (Mahdevari & Torabi, 2012; Sun et al., 2018).

Therefore, some questions must be answered: (1) what are the effective independent parameters/predictors for squeezing prediction? (2) Which properties must contain a database to provide a valid predictive criterion? (3) How to access the validity of such a criterion? A thorough literature review was performed to answer these crucial questions, detect possible shortcomings of available predictive methods and determine the most significant predictors (**Section 2**). A comprehensive frequently-used database was collected from the literature. We utilized statistical data-processing methods (including reducing and cleaning) to improve the database quality. The statistical significance of common squeezing predictors was then evaluated.

This study aims to develop statistical-based models for predicting the occurrence and intensity of squeezing at the design stage of excavation in soft rocks. We utilized regression analyses to provide squeezing models considering possible correlations among predictors. The fitting quality, predicting power, and numerical problems were addressed to control the quality of the final equations. Finally, we compared the results of the developed criteria to some well-known empirical equations and a neural network approach using the same statistically-processed database.

2. Background studies of the squeezing phenomenon

Previous research works have introduced illustrative examples of squeezing-related problems at various intensity conditions (Thut et al., 2000; Khanlari et al., 2012). The type and extent of technical and economic problems can be estimated based on squeezing intensity (Hoek & Marinos, 2000; Panthi and Nilsen, 2007). These problems include support failure, delayed construction, floor heaves, financial costs, entrapped machines (Hoek & Guevara, 2009), severe flooding (Lyu et al., 2019a and 2019b), and even human injuries (Sun et al., 2018). Re-excavation or non-standard excavation procedures may be required to decrease the excessive convergences (Thut et al., 2000).

Multiple factors can affect squeezing intensity, such as the stress field, tectonic condition, discontinuity properties, water table, and the type and strength of rocks (Arora et al., 2020). Accurate evaluation of these parameters at the design stage is essential to predict the occurrence and intensity of squeezing. Methods including empirical (Singh et al., 1992) and semi-empirical equations (Hoek & Marinos, 2000), as well as analytical/numerical models (Debernardi & Barla, 2009; Hasanpour et al., 2014) and soft computing procedures (Sun et al., 2018; Chen et al., 2020), were employed to predict the squeezing intensity and occurrence. Empirical and semi-empirical methods are based on rock mass classification indices and recorded deformations around tunnels, respectively. Although these two approaches are considered attractive options (Singh et al., 1992; Goel et al., 1995), the prediction accuracy can be dubious due to their reliance on some local geotechnical parameters. For example, some methods are developed based on experimental data from a few laboratory samples collected from boreholes. Such a methodology may be highly influenced by the condition at the drill site and may result in erroneous field-scale decisions.

The efficiency of data and enough inclusion of all ranges of squeezing intensities are significant factors in developing a squeezing classification system. In some studies (**Table 1**), a database including lower than 100 datasets was used to provide predictive equations for the squeezing. **Table 1** summarizes the well-known empirical and semi-empirical predictive approaches. The predictors-including Norwegian rock mass classification indices (Q and Q_n), overburden (H), and vertical stress (σ_v)- are parameters used to predict squeezing conditions in each equation.

Table 1. Empirical and semi-empirical equations for predicting squeezing

Predictors	Equation (Reference)	Explanation
Q and H	$H \geq 350Q^{1/3}$ (Singh et al., 1992)	39 datasets
Q_n, B, and H	$H \geq (275Q_n^{0.33})\, B^{-0.1}$ (Goel et al., 1995)	72 datasets
Q and H	$H \geq 424.4Q^{0.32}$ (Jimenez & Recio, 2011)	62 datasets
Q, σ_V, and K	$\varepsilon=\left(\frac{0.0191\sigma_V Q^{-0.2}}{K+1}\right)^{-2}+0.0025 \geq 1$ (Dwivedi et al., 2013)	63 datasets
Competency factors (semi-empirical equation)	$N_C = \frac{\sigma_{cm}}{\gamma H} \leq 2$ (Jethwa et al. 1984)	$P_0=\gamma H$
	$SSR = \frac{\sigma_{cm}}{\gamma H} \leq 1$ (Barla, 1995)	----
	$SI = \frac{u_a/a}{\varepsilon_{cr}}$ (Aydan et al., 1996)	Datasets of 21 tunnels excavated in Japan were used.
	$C_g = RMi/\sigma_\Theta$ (Palmström, 2001)	σ_Θ is tangential stress.
	$\frac{\sigma_{cm}}{P_0} \leq 0.35$ (Hoek, 2001)	Datasets collected from 16 tunnels
	$\varepsilon = 0.15(1 - p_i/p_0)\, \frac{\sigma_{cm}}{P_0}^{-(3p_i/p_0 + 1)(3.8p_i/p_0 + 0.54)}$ (Hoek, 2001)	
	$\varepsilon_t=0.2\left(\frac{\sigma_{cm}}{P_0}\right)^{-2}$ or $\varepsilon = \left(0.002-\frac{0.0025p_i}{p_0}\right)\left(\frac{\sigma_{cm}}{P_0}\right)^{(2.4p_i/p_0-2)} \geq 1\%$	This equation provides a strain-based equation based on Monte Carlo simulations (Hoek & Marinos, 2000).

Stability factor	$\frac{\sigma_O}{\sigma_{cm}} \geq 1$ (Bhasin & Grimstad, 1996)	---
Squeezing number	$S = {\sigma'}/{\sigma'_{critical}}$ (Gutierrez & Xia, 2009; Arora et al., 2020)	For clay-rich rocks

B, *K*, σ_v (or P_0), *SSR* (or N_C), γ, σ_{cm}, u_a, a, ε_{cr}, RM_i, $\sigma'_{critical}$ and σ' are representatives of tunnel span, support stiffness, vertical stress, strength-to-stress ratio, unit weight of rock mass, rock mass uniaxial compressive strength, tunnel convergence, tunnel radius, critical strain, rock mass index, the critical effective vertical/horizontal stress for squeezing and the in situ vertical/horizontal effective stress, respectively.

Analytical models such as elasto-viscoplastic, convergence confinement, viscoelastic and plain strain theories were used to evaluate squeezing. Well-known analytical approaches for predicting squeezing are also introduced in **Table 2**.

The analytical methods exclude some key predictors and consequently may inaccurately evaluate the occurrence or intensity of squeezing. The assumptions for developing these analytical theories-e.g., using a circular span, plane strain, or hydrostatic stress condition- can limit their application. Moreover, the anisotropy of stress and heterogeneity of rock mass can impact squeezing. For example, assuming a plane strain condition can result in a predictive criterion and is only valid for a defined distance from the excavation face. But squeezing occurs adjacent to the excavation face. The three-dimensional nature of strain and stress also makes such assumptions invalid. Shortcomings of these analytical procedures, especially at the design phase, can be classified as: (1) relying on up-scaled properties and (2) producing significant uncertainties as well as being calibrated based on a few experimental (Sterpi & Gioda, 2009) or post-excavation in-situ (e.g., convergences) data (Guan et al., 2009; Boidy et al., 2002; Zhifa et al., 2001).

Table 2. Analytical models for predicting squeezing

Model	Reference	Description
Axisymmetric Elasto-Visco-plastic	(Fritz, 1984)	The time-dependent strain and stress (for excavated tunnels) were evaluated based on an axisymmetric assumption.
Time-dependent displacements	(Sulem et al., 1987)	The generalized convergence confinement approach was defined for isotropic homogeneous rocks to evaluate time-dependent deformations of circular tunnels.
Time-dependent viscoelastic	(Pan & Dong, 1991)	Time-dependent deformations of viscoelastic rocks were calculated considering the advancement rate.
Hyperbolic and power-creep laws	(Phienwej et al., 2007)	The time-dependent closure of circular tunnels for various support types and locations was evaluated.
Burger-MC	(Guan et al., 2008)	'Distance to failure' was introduced as a degradation scale of Mohr-Coulomb parameters related to a stress coefficient and time.
Elasto-visco-plastic	(Sterpi & Gioda, 2009)	The impacts of tertiary creep on the convergence were studied based on location and type of support equipment.
Stress-hardening elasto-visco-plastic	(Debernardi & Barla, 2009)	Triaxial creep deformations were evaluated by a stress-hardening constitutive law.

Numerical modeling was also an alternative option suggested for evaluating squeezing (Aksoy et al., 2012). However, the intensity of squeezing is governed by several parameters such as rock type, discontinuity properties, fluid pressure, and fluid properties (Kovári & Staus, 1996). Many previous numerical studies disregarded interactions among these effective parameters; coupled numerical schemes are required to consider these interactions.

The results of complicated coupled numerical simulations are quantitative, which can be considered the main advantage of these methods. However, several factors may affect the accuracy of these approaches, such as a lack of valid geotechnical parameters (especially in the design phase), limited coverage area (e.g., a few drilled boreholes), and modeling simplifications. **Table 3** lists some numerical studies for predicting squeezing.

Table 3. Numerical simulations for predicting the squeezing

Approach	Reference	Description
Finite element method (FEM) analytical solution	(Kulhawy, 1974)	The finite element formulation was used under plane strain conditions for homogenous, linearly elastic rocks.
Axisymmetric FEM	(Ghaboussi & Gioda, 1977)	Kelvin model was utilized to evaluate the time-dependent behavior of rocks.
FEM creep model	(Gioda, 1981)	A non-linear creep model was developed based on finite element formulation.
FEM	(Gioda, 1981)	Time-dependent deformations of viscous rock mass were evaluated based on FEM simulations.
Axisymmetric FEM	(Shalabi, 2005)	Numerical modeling based on hyperbolic and power-law creep models (case study) was performed.
Discrete element simulations	(Yassaghi & Salari-Rad, 2005)	Two-dimensional modeling of tunnel convergence at an igneous contact zone was studied (a case study).
Axisymmetric numerical modeling	(Ramoni & Anagnostou, 2010, 2011)	Interactions between the TBM shield and squeezing intensity were modeled.
Elasto-visco-plastic simulations	(Barla et al., 2012)	An analytical model was implemented in numerical simulations to calculate squeezing intensity.
Finite Difference Method	(Hasanpour et al., 2014)	Influences of excavation rate on squeezing were evaluated.
Discrete element simulations	(Gao et al., 2015)	This study simulates the squeezing at a roadway tunnel due to stresses triggered by coal mining.
Finite element method (FDM) by FLAC3D (back-analysis)	(Tran Manh et al., 2015)	The over-time deformations and the anisotropy were implemented in the numerical model to evaluate the squeezing of the Saint-Martin-la-Porte gallery.
Quasi-3D modeling	(Wu et al., 2018)	A yielding support system is designed for severe squeezing conditions. Quasi-3D models were used to evaluate the cushion effects of this support system on large deformation.
FDM (back-analysis)	(Wang et al., 2021)	A case study (Tawarazaka Tunnel) was considered. Effects of internal friction angle, the ratio of horizontal stress, and bedding were modeled.

Soft computing approaches such as artificial intelligence and support vector machine (SVM) can process a database in a flexible non-linear way without requiring prior knowledge of a particular model (Alimohammadlou et al., 2014). However, chosen predictors, the quality of the database, and validating methods can significantly affect the final results. Most binary classification approaches only used *H* and *Q* as predictors (Jimenez & Recio, 2011; Shafiei et al., 2012; Ghasemi & Gholizadeh, 2019; Farhadian & Nikvar-Hassani, 2020). Therefore, considering the final results of soft computing methodology as a final solution is unreasonable. This approach requires being updated and improved by more datasets (Jimenez & Recio, 2011). The research works performed by soft computing and probabilistic methods to evaluate squeezing are listed in **Table 4**.

Table 4. Soft computing and probabilistic methods for predicting squeezing

Approach	Reference	Description
Uncertainty analysis	(Panthi & Nilsen, 2007)	Monte Carlo simulations were used. Predictors were P_0, ε, σ_{cm}, and p_i; the database included two tunnels excavated in Nepal.
Logistic regression	(Jimenez & Recio, 2011)	A linear classifier model with 62 datasets (including *H* and *Q*) was used.
Naive Bayes classifier	(Feng & Jimenez, 2015)	A Bayesian network based on the Junction Tree algorithm was employed to produce a binary method by 10-fold cross-validation. Predictors were *D*, *K*, *H*, *SSR*, and *Q*. The database included 166 incomplete datasets.
Bayesian network	(Hasanpour et al., 2019)	The risk of TBM jamming was evaluated. The database was prepared based on numerical modeling. Independent parameters were E_{rm}, shield properties (length and thickness), *D*, σ_{cm}, and over-excavation depth.
Support vector machine	(Sun et al., 2018)	Support vector machine and decision tree were combined for 117 datasets (*D*, *K*, *H*, and *Q*). Multi-class, 8-fold cross-validation and "one-against-one" methods were used.
	(Shafiei et al., 2012)	A binary classification was developed by SVM for 198 datasets (including *H* and *Q*). The leave-one-out method was used for validation.

	(Li et al., 2012)	SVM was combined with particle swarm optimization and chaotic mapping. Chaotic mapping was used to optimize kernel function and training parameters for SVM. A database of convergence for the Xiakeng Tunnel at Wan-Gan Railway was included (a case study).
	(Yao et al., 2012)	The genetic algorithm was used to optimize SVM (Hybrid method), and the results were compared with an artificial intelligence method. A database of strains was used from the Wuhan-Guangzhou railway (138 datasets).
	(Huang et al., 2021)	SVM was combined with backpropagation. A database (180 datasets of *D*, *K*, *H*, and *Q*) was used to produce binary results.
	(Farhadian & Nikvar-Hassani, 2020)	Kringing geostatistical method for 225 datasets (including *H* and *Q*) was utilized. Results were presented as a multi-class classification of squeezing.
Radial basis network	(Mahdevari & Torabi, 2012)	Radial basis function analysis was compared to multi-layer perceptron and multi-variable methods. The effects of different parameters were evaluated by a sensitivity analysis of convergence data. Parameters were *UCS*, *RQD*, *H*, *GSI*, ϕ, γ_{dry}, γ_{sat}, *C*, ν, *E*, σ_c, and σ_t for 60 datasets from Ghomroud Tunnel, Iran.
Logistic regression	(Ghasemi & Gholizadeh, 2019)	Binary logistic regression (BLR) and linear discriminant analysis (LDA) were used to develop two binary empirical equations based on 220 datasets, including *H* and *Q*.
Decision tree algorithm	(Chen et al., 2020)	A trained classifier was integrated into the Markovian geologic model. A database (154 datasets including *D*, *K*, *H*, *SSR*, and *BQ*) was used for binary and multi-class methods. 10-fold cross-validation was used.
Hybrid classifier ensemble	(Zhang et al., 2020)	Seven machine learning classifiers were combined using a weighted voting approach to provide a classifier ensemble. The weight and hyper-parameters of each classifier were tuned using the firefly algorithm. Missing parameters of datasets were replaced by imputation procedures. The predictors were *D*, *K*, *H*, *SSR*, and *Q*, and a database containing 166 incomplete datasets was collected. Results were presented as a binary method.

RQD, γ_{dry}, γ_{sat}, UCS, GSI, σ_t, C, E, ϕ, and ν depict the rock quality designation, dry intact rock unit weight, saturated unit weight, uniaxial compressive strength, geological strength index, rock mass tensile strength, the cohesion of rock mass, young's modulus, friction angle of rock mass, poison ratio of the rock mass. SRF and BQ are representatives of the stress reduction factor and rock mass quality, respectively.

The squeezing intensity is usually classified based on a qualitative description of the structure at the production phase, **Table 5** (Hoek & Marinos, 2000). This qualitative method is used for some of the studies introduced in **Tables 1-4**. The normalized strain (ε), the ratio of deformations to tunnel size, is also presented as a quantitative parameter to classify squeezing. In many research works, the normalized strain of 1% is considered a squeezing threshold for unsupported tunnels (Aydan et al., 1993; Hoek & Marinos, 2000; Sun et al., 2018; Zhang et al., 2020). However, various ranges of ε were introduced for each squeezing class in previous research works (Aydan et al. 1996; Sakurai 1997; Singh and Goel 1999; Hoek and Marinos 2000).

Table 6 shows various quantitative classifications of squeezing intensity based on considering four (Singh et al., 1992) or five squeezing categories (Aydan et al., 1996; Hoek & Marinos, 2000). Recorded deformations during excavation may be considered a predictor to classify the squeezing intensity (Li et al., 2012; Yao et al., 2012), but pre-excavation is a prerequisite for measuring ε accurately. Therefore, this predictive factor usually cannot be used at the design stage.

Table 5. A frequently-used qualitative classification of squeezing intensity based on rock behavior (Hoek & Marinos, 2000)

Squeezing intensity	Rock behavior	Stability condition
No squeezing	Elastic	Stable tunnel after the termination of face effects
Light	Strain-hardening	Stable tunnel and converged displacement when face effects stop
Moderate	Strain-hardening	Large displacement: converged displacement after face effects
Severe	High strain-hardening	Extreme displacements: do not converge even after face effects

Very severe	Rock flow & collapses	Extreme displacement: re-excavation & heavy support are required.

Table 6. Quantitative classifications of squeezing intensity based on normalized strain

Squeezing intensity	(Singh et al., 1992)	(Aydan et al., 1993)	(Hoek & Marinos, 2000; Hoek & Guevara, 2009)
No squeezing	ε≤1	$\varepsilon_\theta^a/\varepsilon_\theta^e \leq 1$	ε≤1
Light	1<ε≤3	$1< \varepsilon_\theta^a/\varepsilon_\theta^e \leq 2$	1<ε≤2.5
Moderate	3<ε≤5	$2< \varepsilon_\theta^a/\varepsilon_\theta^e \leq 3$	2.5<ε≤5
Severe	ε>5	$3< \varepsilon_\theta^a/\varepsilon_\theta^e \leq 5$	5<ε≤10
Very severe	---	$\varepsilon_\theta^a/\varepsilon_\theta^e >5$	ε>10

ε_θ^a and ε_θ^e depict the peak tangential strain and peak elastic strain.

Considering the summarized literature review for evaluating squeezing, although substantial studies were performed, the shortcomings of available predictive approaches confirm a necessity to continue research works in this field. In the previous research works, a clear distinction between the squeezing type for tectonic-based shallow squeezing or hard/soft rocks at a deep depth is absent. Most previous works studied the squeezing of soft geomaterials at deep depths. This study aims to improve the understanding of tunnel squeezing based on statistical methods.

3. Methodology

Based on the literature review, we defined some effective predictors to evaluate the occurrence and intensity of the squeezing and processed a collected database to present new predictive criteria. Statistical approaches such as factor analysis were used to improve the quality of the datasets.

We performed binary and multi-class regression analyses for the occurrence and intensity of squeezing, respectively. Statistical tests such as goodness-of-fitness, Chi-Square statistics, likelihood ratio, and statistical significance were used to analyze the accuracy of final equations.

3.1. Datasets and effective parameters

Although numerous studies covered squeezing, most presented databases were limited (i.e., including low datasets) or incomplete (lack of some main parameters). In this study, a comprehensive database (made up of 159 complete, frequently-used datasets) was collected from previous research works (Shrestha & Broch, 2008; Dwivedi et al., 2013; Feng & Jimenez, 2015; Sun et al., 2018; Chen et al., 2020) to study the squeezing for soft rocks. Five new datasets from the Golab tunnel excavated in Iran were included.

The database was controlled to include the main characteristics of squeezing (i.e., time-dependent deformation of soft rock mass and stability problems) and contain all the defined squeezing predictors. It was possible to compare them statistically, considering the in-common properties of these datasets. We included the following possible effective parameters as initial key predictors (i.e., the independent parameters) affecting squeezing (the dependent parameter) based on the previous prementioned studies:

- *H*: height of overburden overlying the underground structure
- *SSR*: percentage of the strength-to-stress ratio
- *BQ*: rock mass classification index (quality index)

- D: tunnel diameter
- K: support stiffness

Stress state could be recognized as the most used parameter influenced by several factors (Jethwa et al., 1984; G. Barla, 1995; Hoek & Marinos, 2000; Bhasin & Grimstad, 1996), such as orientation and depth of underground structures (the mechanical stresses) as well as the tectonic situation of rock mass (Mahdevari & Torabi, 2012).

In the case of tectonic-related stresses, some studies cited that tectonic conditions could only alter the stress field at a large-scale range, e.g., a 200 km area (Heidbach et al., 2010). Therefore, the total stress state could be considered only depth-dependent for most geo-engineering projects (with length < 200 km). Other research works mentioned that tectonic condition is partially understood and may alter based on the findings at the production phase (Guan et al., 2012). Therefore, these in-field characteristics cannot be included as a scale in the design phase (Sun et al., 2018).

In many cases, an approximately similar density is assumed for all squeezing-prone rocks, i.e., soft rocks. Therefore, tunnel depth (H) is used as a scale to evaluate the stress state (Sun et al., 2018). However, the occurrence of squeezing at shallow depths (Shrestha & Broch, 2008) has called *choosing overburden as a proper predictor* into question. Meanwhile, *SSR* implements the effects of both rock mass compressive strength (σ_{cm}) and vertical stress (γH)-**Table 1**. In many studies based on soft-computing methods, *SSR* was ignored (Sun et al., 2018; Huang et al., 2021), while the significance of this parameter was proved based on other works (Feng & Jimenez, 2015; Zhang et al., 2020).

An interesting point (especially in soft-computing approaches) is that *SSR* (as a well-known predictive scale of squeezing) has been simultaneously used with H (Feng & Jimenez, 2015; Chen et al., 2020; Zhang et al., 2020). The overburden is directly included in calculating *SSR* (**Table 1**); we evaluated the effects of the simultaneous inclusion of H and SSR on the statistical quality of the squeezing database.

Rock mass classification systems, especially the Q system, could be considered a representative index of rock mass quality and joint condition at underground structures (Singh et al., 1992; Jimenez & Recio, 2011; Dwivedi et al., 2013). However, applying Q could include two intrinsic shortcomings: (1) this parameter provides comparatively more accurate results for fractured blocky rocks but not necessarily for other rock mass conditions (Palmstrom & Stille, 2007). (2) Evaluating the stress reduction factor (*SRF*), as a required parameter for calculating Q, is a challenging task due to being related to the intensity of squeezing (Palmstrom & Broch, 2006). In one study, *SRF* was estimated based on the existence of weakness zones instead of squeezing intensity (Shrestha & Broch, 2008).

Other rock mass classification systems such as rock mass number (Q_n) (Goel et al., 1995; Dwivedi et al., 2013) or *BQ* (Shen et al., 2017; Chen et al., 2020) are used to determine rock mass condition in some methods to predict squeezing considering the prementioned shortcomings of using Q.

In this study, we utilized *BQ* as the classification index due to the prementioned disadvantages and because many datasets (mainly those related to Chinese tunneling projects included in our database) were presented based on *BQ*. For other cases, *BQ* was calculated based on other classification indices (Shen et al., 2017; Chen et al., 2020), e.g., by **Equations 1** (Yan-Jun et al., 2017) for the datasets including Q.

$$BQ = 63.029 \, \mathrm{Ln}(Q) + 327.5 \qquad \textbf{(1)}$$

Q and BQ depict the Norwegian Index and Basic Quality index, respectively.

Artificial intelligence-based studies confirmed that support stiffness and tunnel size could also affect squeezing, e.g., the accuracy of a prediction equation increased up to 11% and 4% after including K and D, respectively (Sun et al., 2018). Installing a stiff support system could decrease excessive deformations (Dwivedi et al., 2013; Feng & Jimenez, 2015), and the installation time could also be considered another effective predictor (Jethwa et al., 1984). The size of the underground structure might alter squeezing (Goel et al., 1995), and tunnel diameter could add these effects to a predictive formula of squeezing (Dwivedi et al., 2013; Feng and Jimenez, 2015; Sun et al., 2018).

Despite addressing many geotechnical parameters (the independent predictors), most available datasets in the literature were presented without describing the tectonics condition of the study sites. Therefore, it was impossible to include tectonic properties as a predictor. These restrictions resulted in the exclusion of tectonics in final models, similar to past soft-computing methods (**Table 4**). However, some squeezing cases could be considered tectonic-based in the collected datasets.

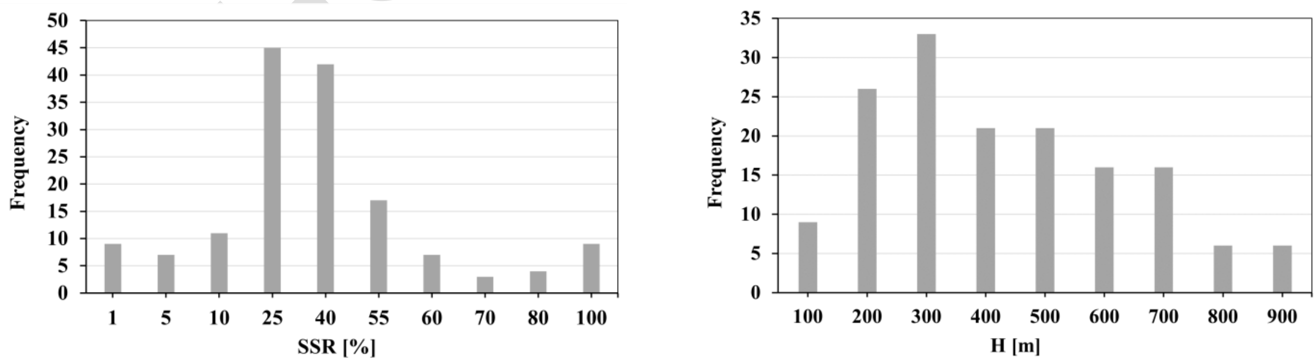
Table 7 and **Figure 1** show the statistical properties of the databases. Based on some statistical approaches, H was deleted from the effective parameters of database one. Then, Database 2 was processed to determine the significant predictors as the final fitting database (Database 3) for regression analysis.

Database 4 was used to control the ability of regression equations for datasets excluded in the fitting process. Some of these datasets were related to probable tectonic-based squeezing (a low H value and a high SSR value), especially the datasets of the Golab Tunnel excavated in Iran.

Table 7. The properties of squeezing databases used for regression analysis

Database	Datasets	Parameters					
		SSR [%]	*H* [m]	*BQ*	*K* [MPa]	*D* [m]	Intensity
1	154	√	√	√	√	√	√
2	154	√	×	√	√	√	√
3	154	√	×	√	×	×	√
4	10	√	×	√	×	×	√

(a) (b)



(c) (d)

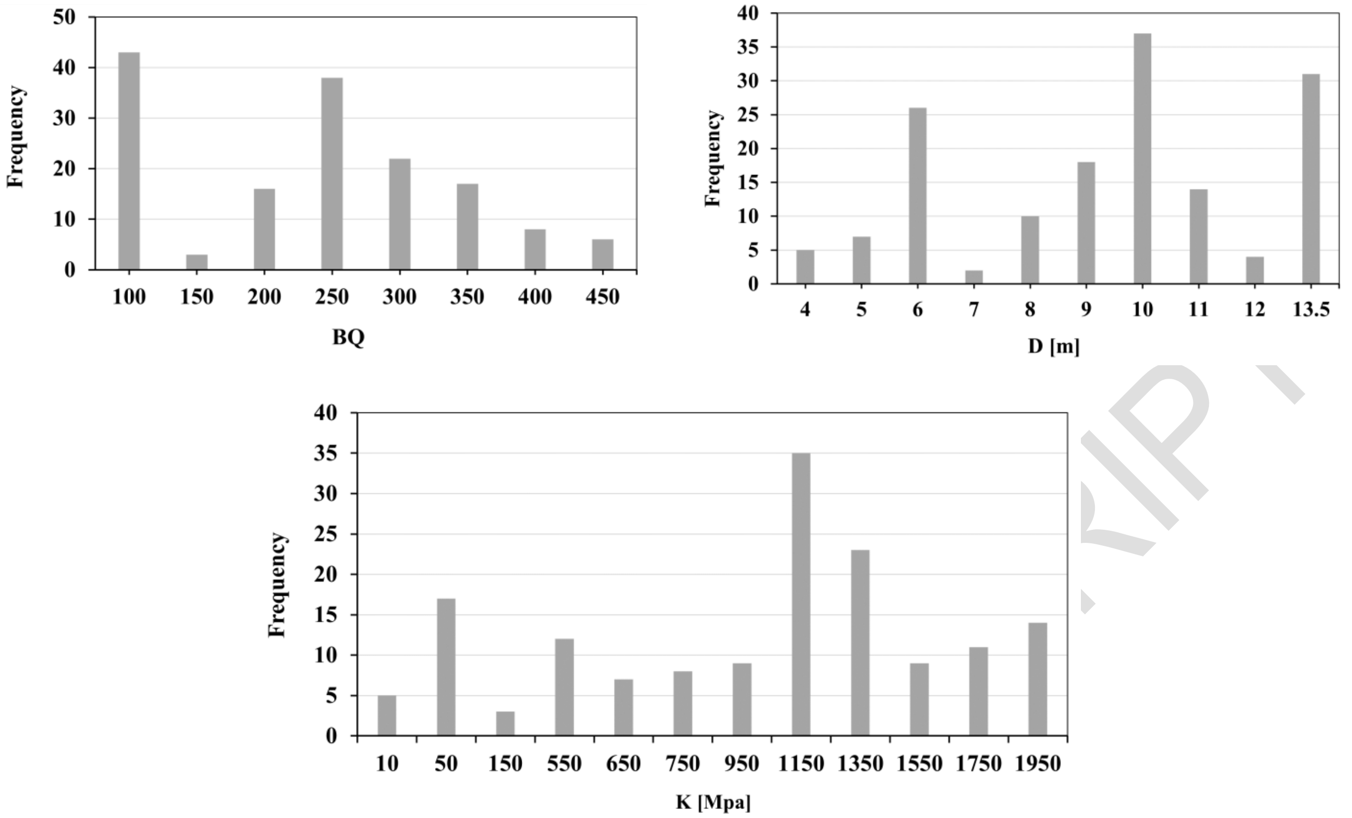


Figure 1. Frequencies and ranges of the predictors in the fitting database: (a) the strength-to-stress ratio (%), (b) overburden height (m), (c) *BQ* classification index, (d) structure diameter (m), and (e) support stiffness (MPa)

3.2. Data processing

The objective of regression analysis is to define a relationship between significant independent parameters and the dependent parameter. But, before performing a regression analysis, the statistical quality of the collected database should be improved by a series of statistical approaches. Several questions must be answered: first, how many datasets are required to begin a regression analysis? What predictors should be removed from the database, and what parameters are statistically significant? Do the values of the predictors follow a normal distribution? And finally, what are the outliers of each parameter?

Three groups of statistical approaches can be used to improve the quality of a database before regression analysis: (1) integrating or merging datasets from different sources, (2) reducing and transforming data, and (3) cleaning outliers by calculating noises and correcting inconsistencies (Acaps, 2016; Yang et al., 2006).

We utilized Nonlinear Principal Component (NLPCA) and factor analyses (for reducing the data) to increase the quality of the database collected from different sources (merging). Factor analysis uses maximum common variance to decrease the number of predictors (Wheelwright et al., 2020). Outliers, i.e., a value that lies at an abnormal distance from, were also detected and substituted to clean the database (data cleaning). This threshold can generally be defined between 1.5 to 3 times the standard deviation from the mean value (SPSS, 2010). **Figure 2** shows the processing methods used in this research work.

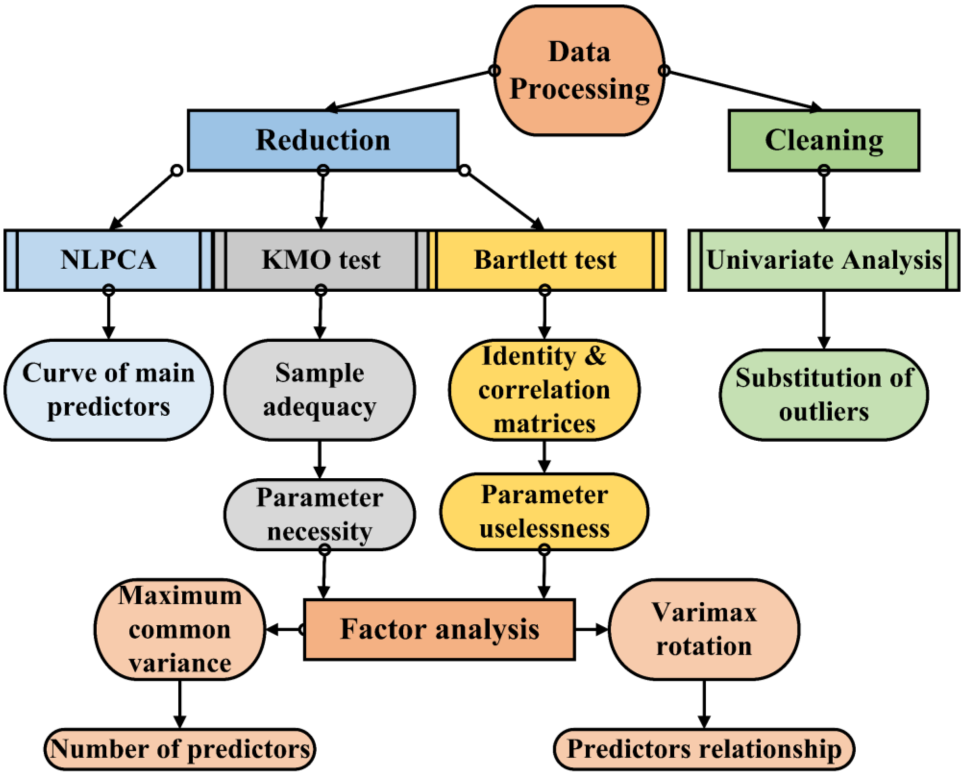


Figure 2. Flowchart of the data processing to improve the quality of the collected squeezing database

Before performing the factor analysis, the database efficiency and predictors compatibility were controlled by Kaiser-Meyer-Olkin (KMO) and Bartlett's Sphericity tests. The KMO test determines sample adequacy, while Bartlett's Sphericity test evaluates the uselessness of parameters by comparing identity and correlation matrices. Bartlett's test also controls if the variances are equal for all datasets (homogeneity of variances). The necessary number of predictors was selected based on Kaiser's rule. Varimax rotation determined the relations among the predictors in factor analysis, i.e., the independent parameters (Wheelwright et al., 2020).

Principal components analysis could define relations between significant predictors (Norusis, 2012), and final squeezing equations could result in more accurate outputs after performing this method. NLPCA provides curves of main predictors instead of straight lines. The factor extraction by NLPCA reduced the summation of differences between two squares between reproduced and observed correlation matrices (Scholz et al., 2008). The final results of these data-processing approaches are significantly less exposed to intrinsic limitations and assumptions (Kriegel et al., 2010).

Univariate or multivariate approaches could detect outliers. Intensity and occurrence of squeezing both could be considered univariate parameters because only one dependent variable was defined for each case: intensity class or occurrence possibility, respectively. The univariate approach substituted the outliers based on the standard deviation and mean values and the defined threshold of confidence intervals (Denis, 2018). In this research, we considered a confidence level of 95% and evaluated the outliers and their substitutes by SPSS software (SPSS, 2010). Univariate analysis was also used to determine the normality and equality of datasets (section 3.3).

3.3. The univariate analysis

Univariate analysis was performed to evaluate three properties for all squeezing classes: normality condition, mean equality, and distribution function equality of predictors. The normality condition confirms a normal distribution of datasets, while mean equality measures the potential of independent parameters (Norusis, 2012). For mean equity, correlation analysis was assessed among predictors, and different tests were chosen based on the number of classes. Two classes were used for squeezing occurrence and more than three for its intensity. **Figure 3** summarizes the process of univariate analysis.

Three methods were used to control the normality condition: (1) descriptive numerical approach (skewness for controlling asymmetry of the distribution and kurtosis for its tailedness), (2) theory-driven graphical methods including P-P plot for agreement of datasets based on cumulative distribution function (CDF), and Q-Q plot for agreement of datasets in terms of quintiles and (3) theory-driven numerical approaches including Shapiro-Wilk and Kolmogorov-Smirnov tests for normality of frequencies (Denis, 2018).

Four tests can evaluate mean equality: for parametric tests: the T-test (for two Independent Classes, i.e., squeezing occurrence) and F test (for a minimum of three independent classes, i.e., squeezing intensity class) were available options. The Mann-Whitney (for two Independent Classes) and Kruskal-Wallis (for a minimum of three Independent Classes) tests were utilized for non-parametric tests (Denis, 2018).

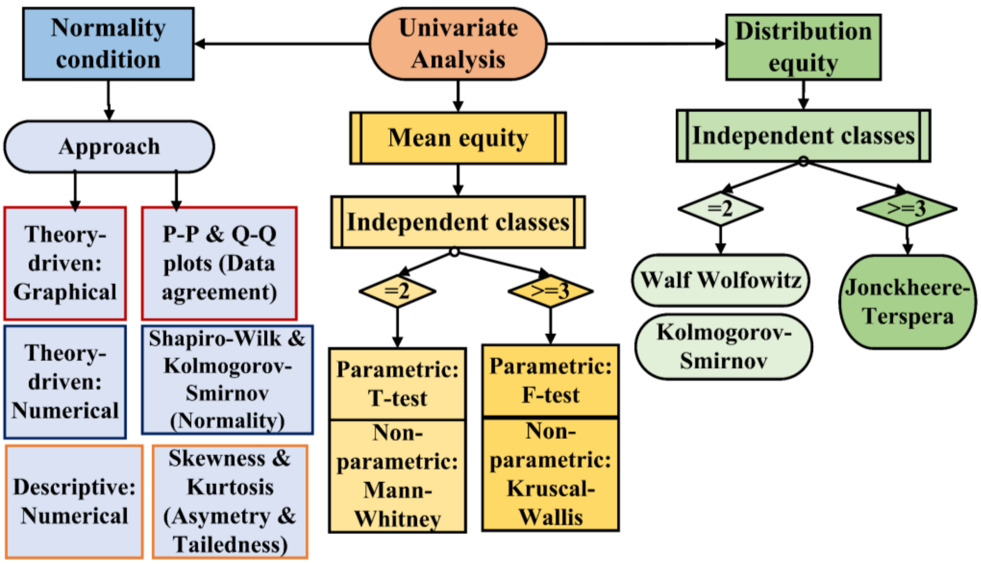


Figure 3. The flowchart of the univariate analysis for the squeezing database

Distribution function equality was also checked based on the number of independent classes: the Wald-Wolfowitz and Kolmogorov-Smirnov tests for two Classes and the Jonckheere-Terpstra test for at least three Independent Classes. The equality of the distribution function means that all parameters could be defined based on a distribution. Finally, the correlation coefficient was measured by the Pearson method for parametric and by the Spearman method for non-parametric tests (Noresis, 2012).

If the mean and distribution function of the predictor provided equality and the correlation with the dependent parameters was insignificant, its contribution to tunnel squeezing would be omitted. A predictor could be detected as negligible by comparing the results with and without including different predictors. A high significance value of independent parameters with significant inter-correlation (a correlation with other predictors) can result in the occurrence of a multicollinearity state (Denis, 2018). Therefore, these criteria were all used to evaluate the significant predictors of tunnel squeezing, and the results were prepared for regression analysis.

3.4. Regression analysis

Link functions for regression analysis of dependent parameters can be categorized as (1) two-class functions, including binary probit and binary logistic. (2) Multi-Class Multinomial Logistic approach. (3) multi-class ordinal methods. **Figure 4** presents the flowchart of regression analysis; validation of the results and correlation between significant predictors are also illustrated.

Binary and multi-class functions were used to predict the occurrence and intensity of squeezing, respectively. The following functions (**Equations 2** and **3**) are representatives of binary probit and binary logistic approaches. Odds ratio (*OR*)-i.e., $p \times (1-p)$ or the amount of correlation between two parameters- was also calculated (Fox, 2016).

(Fox, 2016):

Binary probit link= $\varnothing^{-1}(1-p)$ **(2)**

Binary logistic link=$\mathrm{Ln}(p/(1-p)$ **(3)**

ϕ and p depict the standard deviation and class probability of normal distribution.

For the multi-class nominal link function (**Equation 4**), the class with a minimum score (i.e., the no-squeezing class) was considered the reference class and compared with other classes using a series of binary logistic analyses. The minimum score class assumed zero membership probabilities, i.e., *OR* of the class divided by the sum of all classes' *OR* was considered zero, and the regression coefficients were evaluated for each comparison case (Fox, 2016). Index (i) depicts the number of each class.

Nominal link = $\mathrm{Ln}\left(p_i(1-p_i)\right) = \mathrm{Ln}OR_i$ **(4)**

P_i and ORi are representatives of class probability and odds ratio of class i.

In the case of multi-class ordinal link functions (**Equations 5-9**), all classes of the dependent parameter (except the last one) were fitted by an equation, and cumulative probabilities were evaluated by assuming to be equal to or less than a value for these classes. Cumulative probabilities of consecutive classes were differentiated to calculate independent probabilities. Multipliers of effective parameters were then assumed to be similar for all the squeezing Classes, and the correctness of this assumption was controlled by the test of parallel lines (Fox, 2016).

Logistic (**Equation 5**), probit (**Equation 6**), and Cauchit (**Equation 7**) equations and complementary (**Equation 8**) and negative (**Equation 9**) log-log link functions were all implemented for Database 3 (Norusis, 2012). We also compared the accuracy of these functions for predicting squeezing intensity:

Logistic link=$\mathrm{Ln}\left(\gamma_i(1-\gamma_i)\right)$ **(5)**

$Probit\ link = \emptyset^{-1}\gamma_i$ **(6)**

Cauchit link=$\tan(\tau(\gamma_i-0.5))$ **(7)**

$$\text{Complementary log\_log link} = \text{Ln}\left(-\text{Ln}(1-\gamma_i)\right) \quad \textbf{(8)}$$

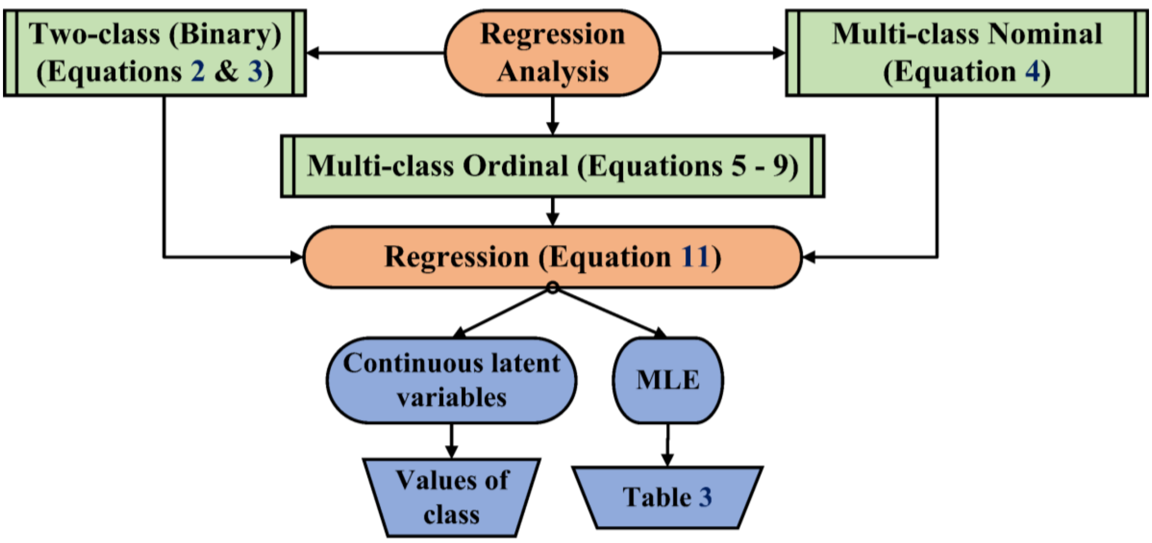
$$\text{Negative log\_log link} = -\text{Ln}(-\text{Ln}\gamma_i) \quad \textbf{(9)}$$

γ_i depicts the cumulative probability of ith class; $\varnothing$ is the standard deviation. Results of **Equations 5-9** can be similar in some cases. Since no prerequisite parameter (e.g., normal error distribution) was required for these regression approaches, maximum likelihood estimation (MLE) was used instead of minimum mean square error (MMSE) to develop predictive equations and to evaluate the multipliers of each predictor. i.e., MLE guarantees close prediction of the dependent parameter. The coefficients (β) were predicted in a way to produce the highest likelihood value based on y_i (the sample symbol) (**Equation 10**) (Fox, 2016):

$$\text{MLE} = \max \sum_{i=1}^{n} \ln(L(y_i \mid \beta)) \quad \textbf{(10)}$$

The outputs of equation **10** were controlled to include two properties: being probable and converging to a unique value. The predicted β values were used to calculate the membership probabilities of each class. The accuracy of each classifying equation was then evaluated by comparing the predicted and actual categories.

(a)



(b)

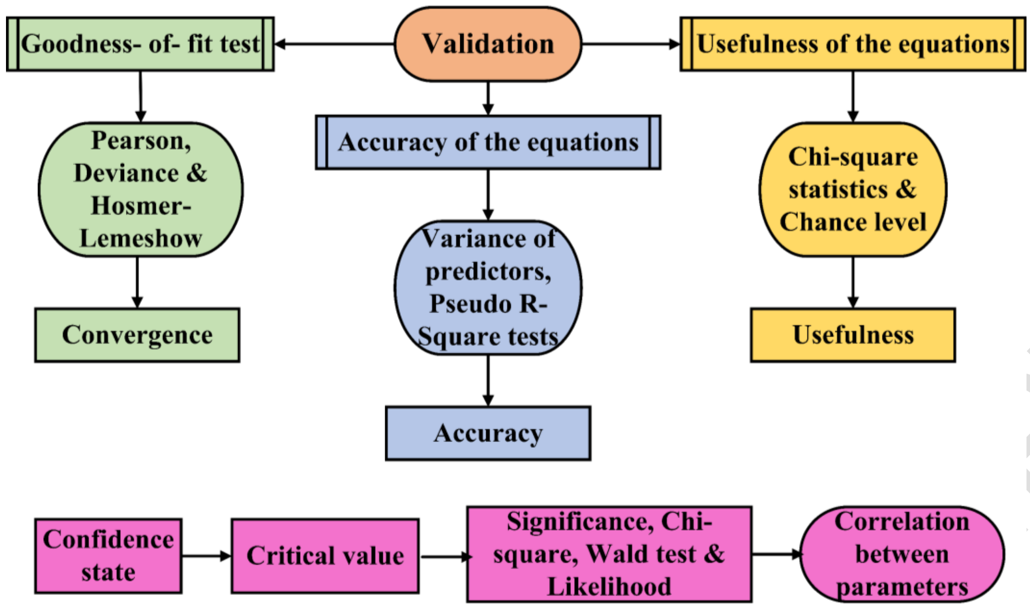


Figure 4. Flowchart of regression analysis for predictive equations of squeezing (**a**) and the validation process (**b**)

All link functions presented by **Equations 2-9** evaluated continuous latent variables (Y') and converted them to the value of classes (**Figure 4**). For example, Y' equal to 2.23 means that the squeezing intensity can be light (i.e., the squeezing class 2). Finally, the regression analysis for the dependent variables was performed using the following equation (Norusis, 2012):

$$\mathrm{Link(Y)} \equiv Y' = \beta_0 + \sum_{i=1}^{k} \beta_i x_i \quad \textbf{(11)}$$

Link, Y, Y', β_0, k, and β_i are the link function (**Equations 2-9**), the value of the qualitative dependent parameter, continuous latent variables, constant value, number of predictors, and multipliers of each variable (x_i). The fitting iteration was set to 300 for all regression models (SPSS, 2010). The correlations (between the dependent parameter with predictors and intercorrelation between predictors) were checked based on some statistical methods (i.e., using the likelihood ratio, Wald test, and Chi-square values), and a critical value was evaluated based on the selected confidence interval of 95 %.

At this stage, we developed various regression models for the occurrence and intensity of squeezing. However, the correlation among the parameters and validation of the fitting process (convergence of fitting, accuracy, and usefulness of presented equations) must be controlled.

Statistical significance was used to evaluate the rejection of the null hypothesis considering the correlation between the parameters. For this reason, values outside the critical areas were rejected, and the alternative (i.e., a higher Chi-Square value) was considered. The significance value of each function was equal to the Chi-Square value (the difference between log-likelihoods of the link function with and without considering the predictor) multiplied by -2. For the significance value of zero, the null hypothesis is true. Otherwise, where significance values were less than the presumed amount, a higher Chi-Square value was considered a scale to choose the most accurate function (Fox, 2016).

The quality of predictors was also checked by two tests based on the significance value (**Figure 4**): (1) assuming no relation between the dependent and independent parameters (initial significance value of 0), a likelihood ratio test was performed to calculate the possible relationships. (2) Presuming a zero multiplier for the independent predictor, the Wald test (**Equation 12**) calculates the significance value (Norusis, 2012).

$$\mathrm{Wald}(x_\mathrm{i})=(\beta_i/SE)^2 \qquad \textbf{(12)}$$

SE and β_i depict standard error and multipliers of the predictor x_i.

Validation of regression analysis (**Figure 4**) was performed to avoid various problems. The divergence of regression analysis for the dependent parameter could be problematic. Goodness-of-fit tests- i.e., Pearson, Deviance, or Hosmer Lemeshow tests-could be used to control consistency between the observed and fitted parameters based on the significance of the dependent parameter for each squeezing class. The difference between expected (E_i) and calculated (O_i) frequency could create contingency matrices that can be approximately fitted by Chi-Square distribution. The statistic of the Hosmer Lemeshow test is (Norusis, 2012):

$$\mathrm{X}^2_{\mathrm{HL}}=\sum_{\mathrm{i}=1}^{\mathrm{g}}\frac{(O_\mathrm{i}\text{-}E_\mathrm{i})^2}{E_\mathrm{i}(1\text{-}\xi_i)} \qquad \textbf{(13)}$$

The freedom degree of the Hosmer Lemeshow test is (g-2), while ξ_i and g are the means of predicted probability and number of classes with equal size, respectively. For Pearson and Deviance tests, the following relationships were used (Norusis, 2012):

$$X_P^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{i,j}-E_{i,j})^2}{E_{i,j}} \qquad \textbf{(14)}$$

$$\mathrm{X}^2_{\mathrm{D}}=2\sum_{\mathrm{i}=1}^{\mathrm{r}}\sum_{\mathrm{j}=1}^{\mathrm{c}}O_{i,j}\mathrm{Ln}(\frac{O_{\mathrm{i,j}}}{E_{\mathrm{i,j}}}) \qquad \textbf{(15)}$$

r and *c* depict the number of rows and columns, and (*r*-1). (*c*-1) is the degree of freedom. Pseudo *R-square* scales, including McFadden (1974) (**Equation 16**), Cox and Snell (1989) (**Equation 17**), or Nagelkerke (1991) (**Equation 18**), can also be used to compare regression functions based on the variance of the dependent parameter. The higher the variance value is, the less accurate the link function is (Norusis, 2012):

$$R^2_{\mathrm{M}}=1\text{-}\frac{L(\widehat{B})}{L(B^{(0)})} \qquad \textbf{(16)}$$

$$R^2_{\mathrm{CS}}=1\text{-}(\frac{L(B^{(0)})}{L(\widehat{B})})^{\frac{2}{\mathrm{n}}} \qquad \textbf{(17)}$$

$$R^2_{\mathrm{N}}=\frac{R^2_{\mathrm{CS}}}{1\text{-}L(B^{(0)})^{\frac{2}{\mathrm{n}}}} \qquad \textbf{(18)}$$

n, $L(\widehat{B})$, and $L(B^{(0)})$depict the number of parameters and log-likelihoods with and without the predictor parameters. For the cases where a predictor was linearly estimated with enough accuracy by different link functions, multicollinearity (i.e., when the accurate, linear prediction of a predictor is possible using other predictors) occurs (Norusis, 2012).

Other numerical issues (such as the presence of classes with no cases or of a class separated from some predictors) could also occur when non-zero coefficients with high standard errors were calculated for predictors. Therefore, chance-level accuracy was calculated as the summation of the squared values of each parameter.

The usefulness of a link function was also controlled based on having an accuracy higher than the defined chance level. Therefore, the adequacy of link functions was checked by comparing pseudo-R-square scales, prediction accuracies, and chi-square statistics (Norusis, 2012).

We developed regression models for the occurrence and intensity of squeezing, which are comparatively accurate and converged based on the validating statistical methods. Henceforth, 1 and 0 depict squeezing and no-squeezing conditions for the occurrence; for intensity, 1, 2, 3, 4, and 5 show no-, light, moderate, severe, and very severe conditions, respectively (based on **Table 5**).

3.5. Comparison with alternative methods

We encountered several obstacles comparing the results of the regression models with empirical and soft-computing approaches: (1) data processing of previous studies was rarely addressed, (2) various independent parameters and different databases were used, (3) the squeezing types (soft rock, hard rock, and tectonic-based) generally were not addressed and (4) some statistically significant predictors such as rock mass classification systems were absent in the empirical equations (Barla, 1995; Aydan et al., 1996; Hoek & Marinos, 2000).

To tackle these issues, we used our processed database (Database 3) instead of datasets presented in the previous studies and evaluated the accuracy of these alternative approaches compared to the regression functions. In the case of the empirical equations, we chose frequently-used, well-known equations of Hoek & Marinos (2000) and Jiminez & Recio (2011) for intensity and occurrence of squeezing, respectively (**Table 1**).

A similar preprocessing was utilized for a multilayer perceptron feed-forward neural network (MLP NN). We selected two hidden layers with a hyperbolic tangent function as the activation function. 30 and 70 % of datasets were randomly chosen for training and validation tests. The optimization algorithm was scaled conjugate gradient, and '*the maximum steps without an error decrease*' was set to 10 (Hastie et al., 2017).

4. Results

We utilized the statistical methods to prepare a more efficient squeezing database. Factor analysis and NLPCA detected *SSR* and *BQ* as the main squeezing predictors (by processing Database 1) but deleted *H* from the database due to being an improper predictor (Database 2).

There were several reasons for removing *H* from the datasets: (1) based on the KMO test, the sample adequacy of the database was lower after including *H*. (2) *H* and *SSR* were recognized as correlated predictors. (3) As pre-mentioned, the intensity of squeezing can be severe even at shallow overburden (i.e., a low *H* value) due to the tectonic factors; it was visible for several datasets in the collected database. Therefore, severe squeezing cases can occur even at a low *H* value.

The outliers of each squeezing class are shown in **Table 8**. The outliers of *SSR* were substituted by 85.9 %in Databases 3 and 4 (SPSS, 2010; Norusis, 2012); i.e., for the no-squeezing class, *SSR* values higher than 85.9 % were replaced by 85.9 %. Since *H* was removed from all datasets in Databases 2 to 4, the outlier substitution of *H* values was unnecessary.

Based on the statistical analyses, *K* and *D* were recognized as minor predictors with low effects on the intensity and occurrence of squeezing. The reasons for this decision were having equal mean and distribution at different squeezing classes and being correlated insignificantly with the dependent parameter. No outlier was detected for these two parameters and *BQ*. The cases with a low *H* value and a high *BQ* could be considered a scale of possible tectonic-based squeezing.

Table 8. Outliers of the databases based on the statistical analyses

Predictors	Squeezing class	Outliers
SSR [%]	No-squeezing	90, 105.8, 102, 99, 90.5
H [m]	No-squeezing	1000, 1300, 1320
	Moderate	870
	Severe	1110

Having fitted various link functions to the processed database (Database 3 in **Table 7**), we controlled several factors: the classifying accuracy, chi-square statistic, pseudo-*R-square* scales, and numerical problems. *BQ* and *SSR* were chosen as significant independent predictors for all the developed models.

The binary link functions (**Equations 2** and **3**) predicted squeezing occurrence similarly. These accurate estimations (116 out of 116 cases) for binary logistic and probit methodologies seem satisfying. However, the performance of these models for predicting no-squeezing conditions was less accurate (e.g., 68.4 % or 26 out of 38 cases for the fitting database). This condition can be related to the included datasets of no-squeezing conditions based on BQ and SSR. Similar conditions were visible for the validating datasets (**Table 9**). **Tables 9** to **12** summarize the results of regression analyses. In these tables, NS, S, and T are representatives of no-squeezing, squeezing, and total cases, respectively.

Table 9. The binary link functions for predicting tunnel squeezing for fitting and validating databases

Links	Regression Equation	Fitting datasets			Correct predictions (/ depicts out of)					
					Fitted datasets			Validation datasets		
		NS	S	T	NS	S	T	NS	S	T
Logistic	$\mathrm{Ln}(OR) = 5.911 - 0.13\ SSR$	38	116	154	26/38	116/116	142/154	0/3	7/7	7/10
Probit	$\emptyset^{-1}(1\text{-}p)=3.26\text{-}0.072\ SSR$	38	116	154	26/38	116/116	142/154	0/3	7/7	7/10

We also compared the performance of all ordinal link functions. The Cauchit model (**Equation 7**) was the best ordinal classification function to predict the intensity (96 out of 154 cases for the fitting database). However, the model produced a little lower accuracy than the multi-nominal link function (**Equation 4** with 100 correct predictions out of 154). For the ordinal equations, the coefficients were similar according to the basic assumption of these methods.

The results of comparing the binary and multi-class regression models to empirical equations and MLP NN are shown in **Figure 5**, including the number of correct predictions for each squeezing class. We used a similar processed database for all methods, and *SSR* and *BQ* were considered significant predictors.

Table 10. The best-fitted multi-class (nominal and ordinal) link functions for predicting tunnel squeezing

Links	Regression Eq.	Correct predictions (/ depicts out of) For fitting datasets						Constants				
		NS	L	M	S	VS	T	NS	L	M	S	VS
Eq. 4	NS: Reference class $\mathrm{Ln(OR_i)} = ai - (bi * SSR) - (ci * BQ)$	28 /38	19/34	17/38	24/31	12/13	100 /154	Ref. class	a_1=3.802, b_1=0.107 $c1$=0	a_2=5.855, b_2=0.151 $c2$=0	a_2=8.073, b_2=0.264 $c3$=0	a_2=8.804, b_2=0.276 $c3$=0.014
Eq. 5	$Ln(\gamma_i(1-\gamma_i)) = Ti - 0.127\,SSR - 0.005\,BQ$	28 /38	20/34	20/38	18/31	6/13	92 /154	T_i=7.068	T_i=4.902	T_i=2.996	T_i=0.746	Reference class
Eq. 6	$\emptyset^{-1}\gamma_i = \; = Ti - 0.64\,SSR - 0.003\,BQ$	27 /38	23/34	21/38	19/31	5/13	95 /154	T_i=3.695	T_i=2.556	T_i=1.554	T_i=0.329	Reference class
Eq. 7	$\tan(\tau(\gamma_i - 0.5)) = Ti - 0.064\,SSR - 0.003\,BQ$	28 /38	23/34	19/38	20/31	6/13	96 /154	T_i=10.635	T_i=7.158	T_i=4.544	T_i=0.852	Reference class
Eq. 8	$Ln(-Ln(1-\gamma_i)) = Ti - 0.048\,SSR - 0.003\,BQ$	25 /38	21/34	14/38	22/31	7/13	89 /154	T_i=3.799	T_i=2.572	T_i=1.634	T_i=0.567	Reference class
Eq. 9	$-Ln(-Ln\gamma_i) = Ti - 0.074\,SSR - 0.003\,BQ$	29 /38	14/34	17/38	18/31	1/13	79 /154	T_i=3.436	T_i=2.354	T_i=1.218	T_i=0.315	Reference class

* NS, L, M, S, and VS depict no-, light, moderate, severe, and very severe squeezing intensities, respectively.

In the case of squeezing occurrence, the prediction accuracy of the chosen empirical equation (Jimenez & Recio, 2011) at no-squeezing conditions was almost similar to binary regression models (about 70 % in **Figure 5a**). However, the figure was lower for squeezing cases (72.6 % compared to 100 % of logistic and probit link functions).

For multi-class approaches (i.e., squeezing intensity), the Hoek equation (Hoek and Marinos, 2000) predicts 84 cases correctly (accuracy of 54%), while the accuracies of nominal and Cauchit ordinal approaches were 68.78 and 60.36 %, respectively. The accuracy of MLP was 68.8 %; however, the database was processed using statistical methods (factor analyses as well as NLPCA and univariate analysis) before performing neural network analysis.

Table 11. The quality of the best-fitted binary logistic link function for predicting tunnel squeezing

Data-base	Significance tests			Hosmer & Lemeshow test			*R-squares*		Significant parameters
	St	*df*	*Si*	*St*	*df*	*Si*	Nagelkerke	Cox & Snell	
3	79.25	2	0.000	5.917	8	0.657	0.661	0.442	*SSR*, *BQ*

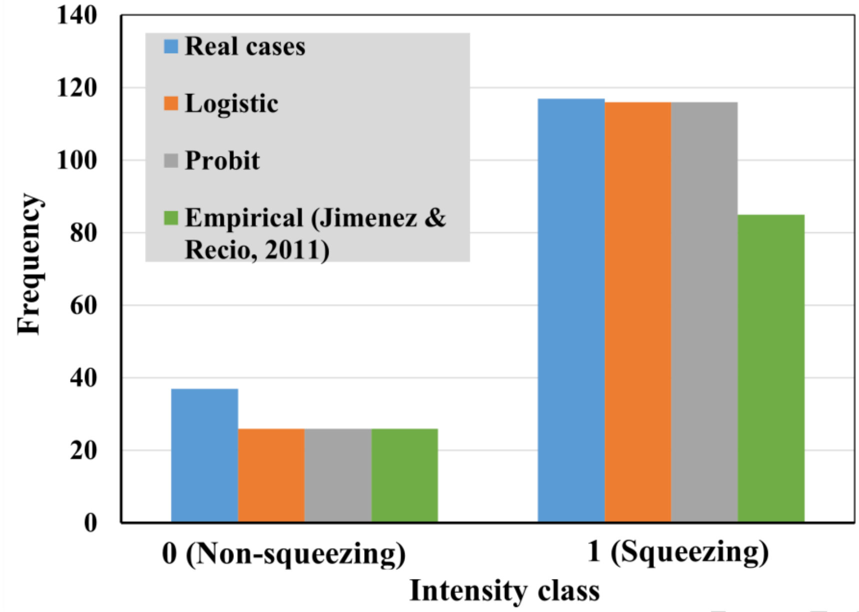
Table 12. The best-fitted probit and multi-class (nominal and Cauchit ordinal) link function for predicting tunnel squeezing

Equa-tion	Significance tests			Pearson goodness test			Deviance goodness test			*R-squares*			Significant parameters
	St	*df*	*Si*	*St*	*df*	*Si*	*St*	*df*	*Si*	Nagelkerke	Cox & Snell	McFadden	
3	89.41	2	0.000	135.6	141	0.962	80.4	141	0.57	0.665	0.446	-------	*SSR*, *BQ*
4	161.3	8	0.000	4158	564	0.000	320.8	564	1.000	0.678	0.649	0.333	*SSR*, *BQ*
7	173.9	2	0.000	632.7	570	0.035	308.2	570	1.000	0.707	0.677	0.359	*SSR*, *BQ*

Statistic=*St*; Significance=*Si*; *df*: degree of freedom

To better describe the sequence of the regression approach, the overall process for datasets 71 and 22 (two typical datasets) with recorded intensity classes of 1 (no-squeezing) and 3 (moderate squeezing) are presented (**Table 13**). For dataset 71 (Miyaluo tunnel), the SSR value was detected as an outlier and was substituted by 85.9 % (SPSS, 2010). The approximate zero binary value ($0.0053 \simeq 0$) depicted a no-squeezing condition. The Cauchit ordinal multi-class approach similarly calculated squeezing condition equal to 1 (i.e., no-squeezing intensity); the results for both regression methods were similar.

(a)



(b)

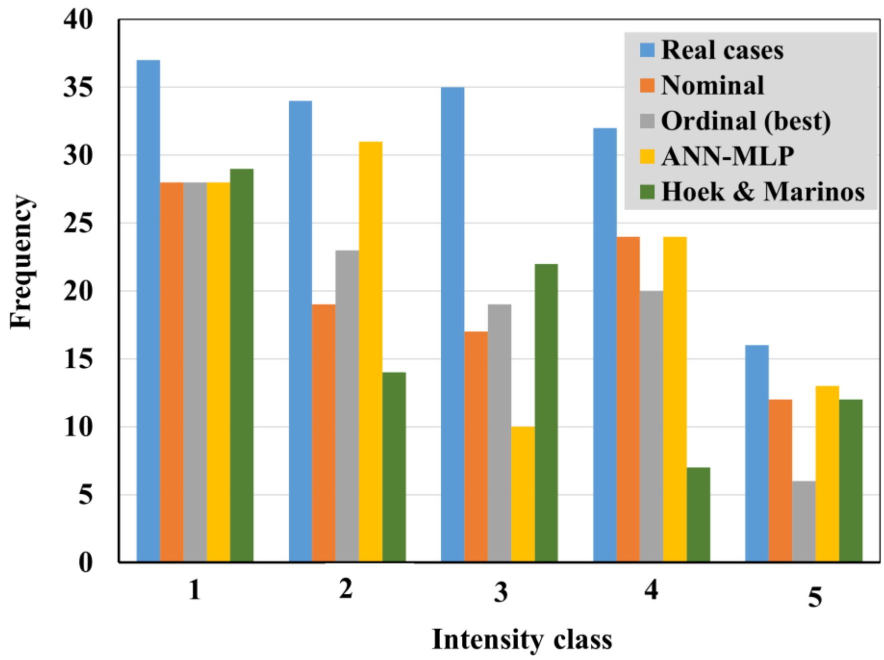


Figure 5. Accuracy of different binary and multi-class approaches to predict squeezing: (a) occurrence of squeezing; 0 and 1 depict no-squeezing and squeezing conditions (b) Intensity of squeezing; 1, 2, 3, 4, and 5 are representatives of no-, light, moderate, severe and very severe squeezing intensity respectively. The Hoek equation (Hoek and Marinos, 2000) is used for the multi-class method.

For dataset 22 (Miyaluo tunnel), the logistic binary resulted in squeezing occurrence ($Y' \simeq 1$) while the Cauchit approach produced 2.77 (rounded to 3, i.e., moderate squeezing intensity). This example confirms how binary methods can provide ambiguous results; the consequences for moderate intensity can be significantly different from light squeezing.

Figure 6 shows the relationship between *BQ* and *SSR* for Database 3. Correctly-predicted datasets are usually in specific parts of the chart for each squeezing class, with some scattered cases (related to incorrectly-predicted data). In such cases, with high *BQ* values, the squeezing can be tectonic-based or for hard rocks, while the developed criteria are mainly for soft materials. Then, the overburden can be considered to detect the squeezing possibility of soft materials at a low depth.

Table 13. The properties of squeezing databases used for regression analysis

Dataset	Database	Parameters						Regression (*Y′*)	
		SSR [%]	*H* [m]	*BQ*	*K* [MPa]	*D* [m]	Intensity	Binary	Multi-class
71	1	105.8	80.33	197.64	1833	9.98	1	---	---
71	2	85.9	×	197.64	1833	9.98	1	---	---
71	3	85.9	×	197.64	×	×	1	0.0053	1
22	1	24.3	600	74.29	26.2	8.7	3	---	---
22	2	24.3	×	74.29	26.2	8.7	3	---	---
22	3	24.3	×	74.29	×	×	3	0.9403	2.77

The *BQ-SSR* charts of no-, moderate and severe squeezing classes of the ordinal regression approach are presented to detect the accuracy of the regression analysis dividing the datasets into correctly- and incorrectly- predicted datasets (**Figure 7 a-c**).

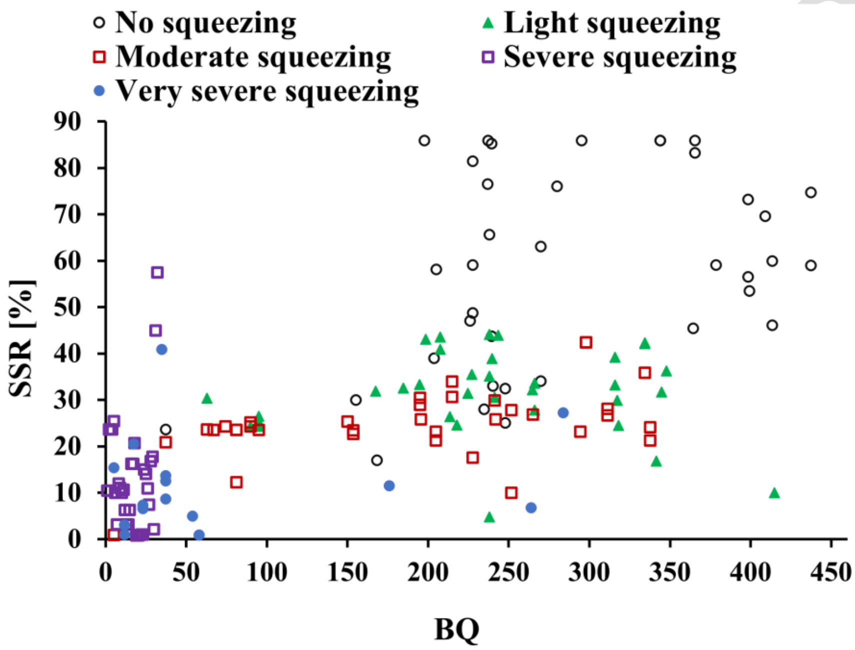


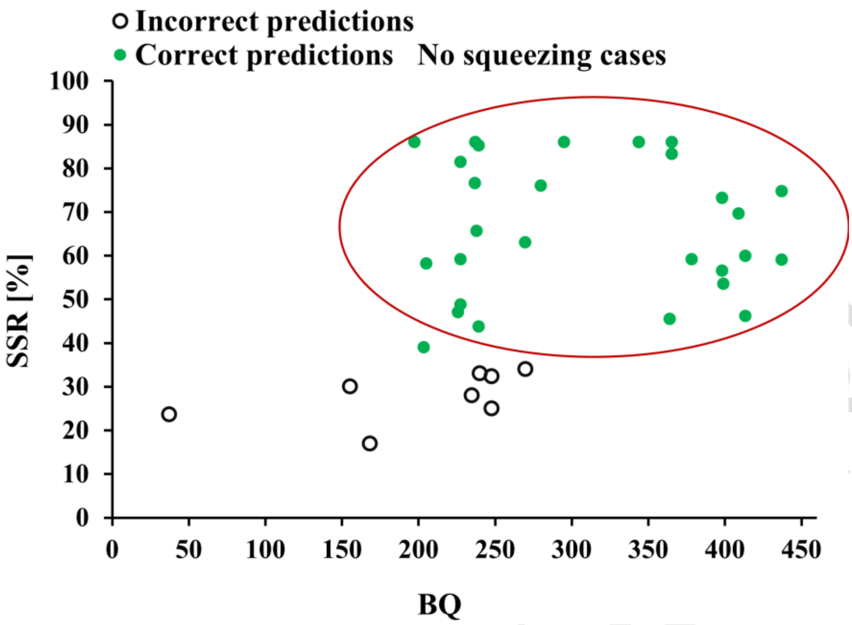
Figure 6. The *BQ-SSR* chart of the fitting database for all intensity classes (154 datasets)

According to **Figure 7a**, correctly-predicted datasets are all located inside the red oval area. A similar situation was visible for moderate, severe, and other intensity cases. Therefore, it can be suggested what combinations of *SSR* and *BQ* values are more probable to result in the correct prediction of squeezing. Some incorrectly-predicted data can be related to tectonic-based squeezing cases.

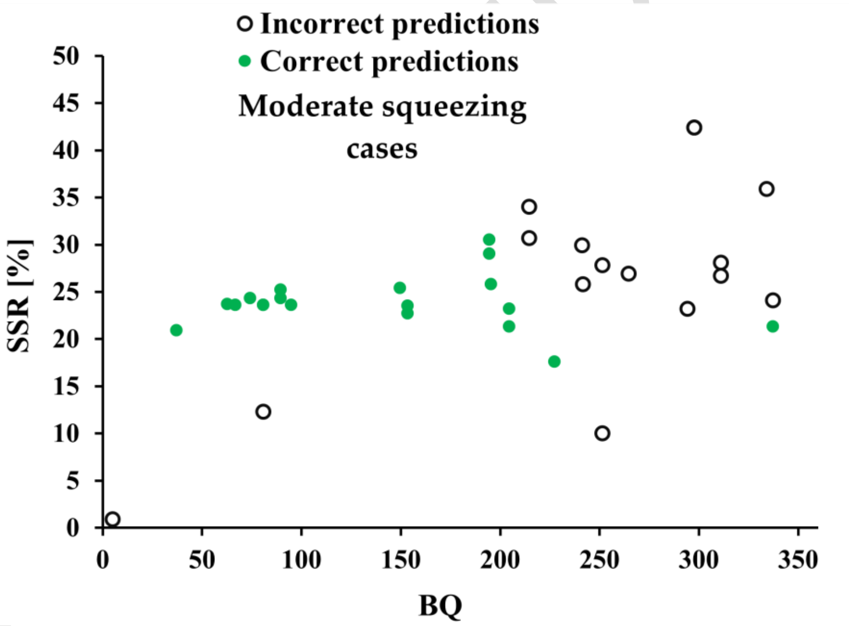
A priority of the regression methods compared to empirical equations is the amount of inaccuracy for the incorrectly-predicted datasets. For example, for datasets 9 and 39 of the fitting database with moderate squeezing intensity, the intensity was predicted as very severe by the Hoek equation. However, severe-intensity squeezing was estimated by the Cauchit ordinal function. This difference is also visible for some other datasets (**Figure 5b**) and can confirm that applying the regression links can provide more realistic results even for incorrect predictions. Such a difference impacts the financial burden of installing support systems during the design phase of tunneling projects.

Figure 8 compares the predictions of multi-class nominal regression for light squeezing intensity with the chosen multi-class empirical equation (Hoek equation). The percentages of under-estimated, correctly estimated, and over-estimated datasets (predicted by the regression function) were 14.7 %, 55.9 %, and 29.4 %, respectively. These proportions for the empirical approach were 26.5 %, 41.2 %, and 32.3 %; a higher underestimation is visible.

a)



(b)



(c)

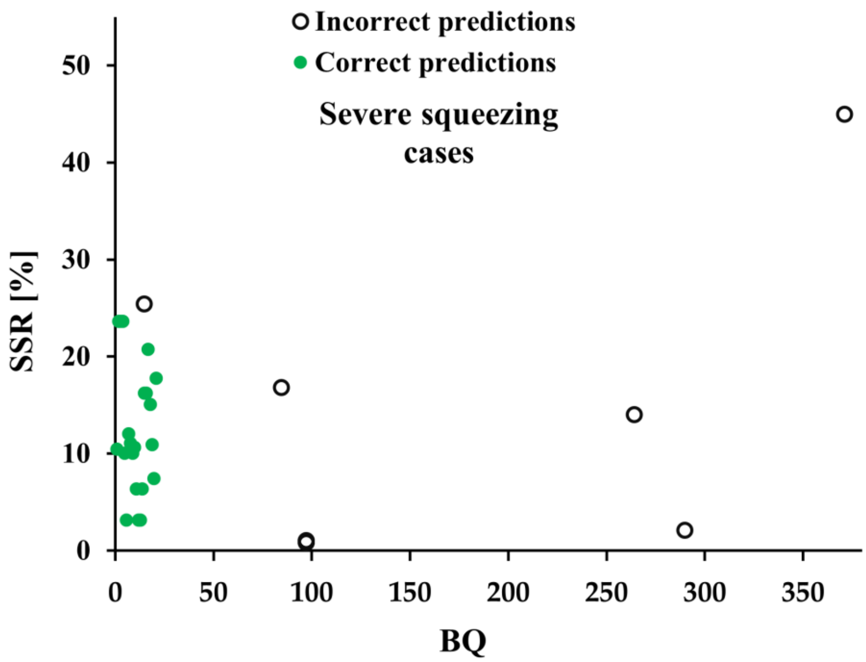


Figure 7. *BQ-SSR* charts of the fitting database for three intensity classes: (a) no squeezing (38 datasets), (b) moderate squeezing (38 datasets), and (c) severe squeezing (31 datasets). Green-filled circles are representatives of datasets that are correctly predicted by the Cauchit ordinal regression approach.

Using multi-class nominal regression, the proportion of under-estimated cases was 25.7 %, 43.7 %, and 25 % for moderate, severe, and very severe classes, respectively. Therefore, the average underestimation of regression models for all five intensity conditions was 21.82 %.

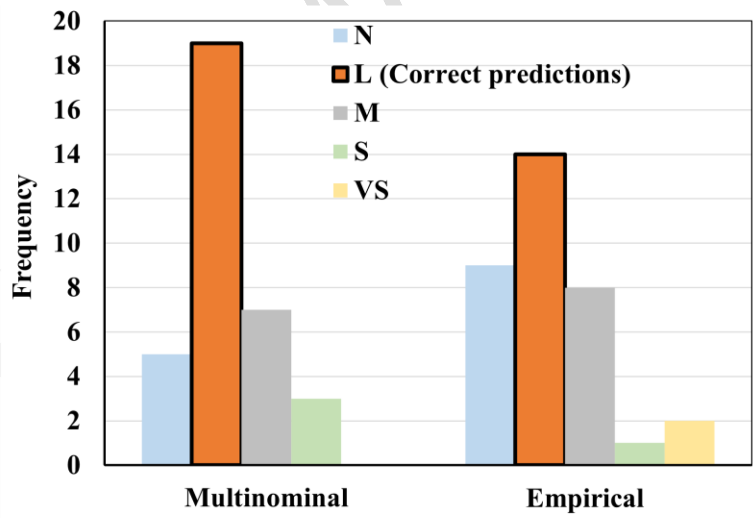


Figure 8. The results of the multi-class nominal regression and Hoek equation for the light intensity of squeezing (the number of datasets was 32); similar datasets were used for both methods, and the datasets were previously processed by statistical approaches presented in **section 3.2**.

In the case of the validation database, the percentage of under-estimated datasets was 40 %, and 40 % of datasets were correctly-predicted, according to **Figure 9**. Also, seven probable tectonic-based squeezing cases (including the data related to Golab Tunnel) were included in this database to control the quality of regression functions in such a condition.

The cases at the top part with high *SSR* values (the *SSR* value between 70 to 85 %) are representatives of these datasets. The incorrect predictions of tectonic-based squeezing are related to cases with high *BQ* (over 300) but a low *H* value. The reason for this inaccuracy can be tectonic factors that may be the main parameters that cause squeezing in such a condition. Such cases (high *BQ* and *SSR* values and a thick overburden) can be the squeezing in hard rock.

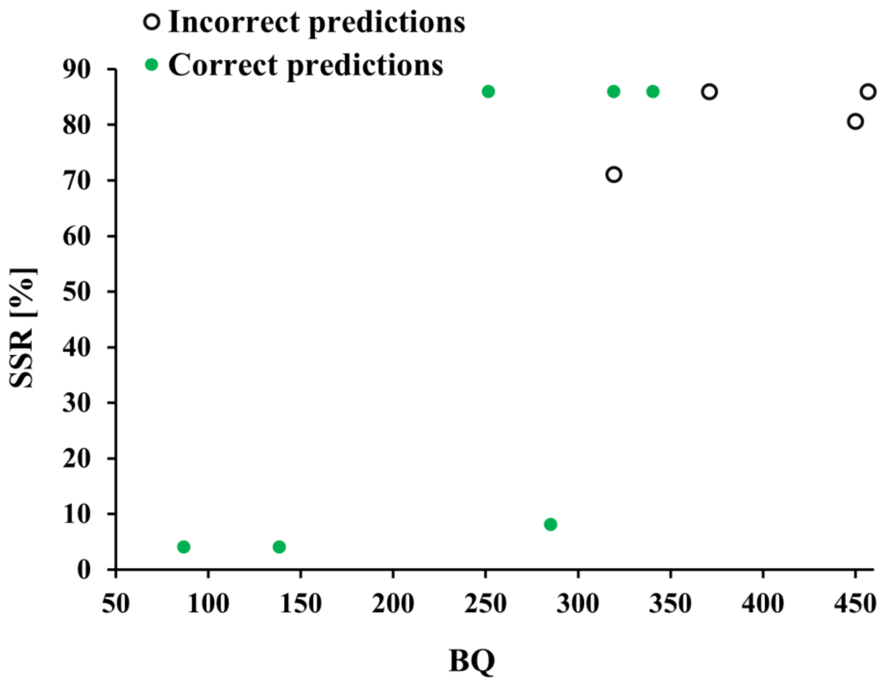


Figure 9. The *BQ-SSR* chart of validating cases (the number of datasets is 10)

5. Discussion

The statistical analyses of squeezing at underground sites, based on a comprehensive database of the squeezing in soft rocks, confirm the importance of strength-to-stress ratio and rock mass condition. NLPCA, univariate, and factor analyses show that *SSR* and *BQ* are the first and second main predictors based on the collected database.

We also detect *K* and *D* as minor independent parameters similar to the past findings (Sun et al., 2018). We eliminate H from our databases due to the technical and statistical disadvantages of including this predictor. The results confirm how squeezing intensity rises by decreasing *BQ* and *SSR*, which is also consistent with the previous studies (Feng & Jimenez, 2015; Chen et al., 2020).

The equations developed using the statistical processing/regression models can provide relatively more accurate models comparing previous approaches for predicting the intensity and occurrence of squeezing based on *BQ* and *SSR*. Over 60 % of cases are accurately predicted, and close overestimations (approximately 20 %) can also be considered acceptable.

The over-estimated predictions can result in designing an unnecessarily stiffer support system and a higher financial burden during the design phase. However, the possibility of tunnel collapse (with significant financial problems), injuries, and causalities will be fallen at the production phase after installing such a stiffer system.

Comparing the empirical equation (Jimenez & Recio, 2011) with our binary regression models to predict squeezing occurrence proves that the regression analyses can provide more accurate results for occurrence and intensity. We provide equations to evaluate squeezing by the strength/stress ratio and the condition of rock mass (*BQ*). The results of the MLP NN approach on the same statistically-processed database are also consistent with the regression functions.

One of the fundamental characteristics of regression models is providing a unique predicting equation for each class of the dependent parameter in contrast to the empirical approaches, which usually present a universal equation. The results of the nominal regression link are more accurate than all ordinal functions: for the Cauchit ordinal method, i.e., the most accurate ordinal equation, the accuracy is 3 % lower than the nominal function.

It is a conventional assumption that volcanic rocks can be considered hard rocks (Malan, 1999). However, in some cases, e.g., in the database used in this study, volcanic rocks are highly-weathered and cannot be categorized as hard rock.

The collected database includes approximately 60 and 30 % metamorphic and sedimentary rocks, respectively, with a maximum *SSR* value of 1.058 (105.8 %). The volcanic rocks in the database (10 %) cannot be considered hard rocks due to the low *SSR* and *BQ* values. Therefore, using the presented regression models to predict the squeezing of hard rocks can provide unreasonable results.

The *SSR-BQ* charts, **Figures 7** and **9**, can be utilized to evaluate the possibility of erroneous prediction. For example, for tectonic-based cases in **Figure 9**, the datasets with a high *BQ* value can result in overestimation. For severe squeezing cases (**Figure 7c**), most of the datasets with a *BQ* value higher than 80 are incorrectly predicted. In such a situation, it is necessary to consider the geological, tectonic, and geotechnical conditions and modify the designed support system at the production stage.

For complementary studies of predicting squeezing at underground sites, we offer a few suggestions to increase the accuracy of the regression models: (1) inclusion of other predictors, especially if it is possible for a tectonic factor, as independent parameters. (2) Increasing the quality of the processed database using more processed datasets, especially for the very severe class. The final regression models can be improved by including more processed datasets in the database. (3) Developing new equations for squeezing cases of hard rocks at high depths. Similar to the previous methods based on empirical and artificial neural procedures, the collected database cannot cover the squeezing of hard rocks. The mere existence of such datasets is the main drawback of providing such a database.

5. Conclusions

Soft-squeezing rocks at underground sites can cause technical issues and financial burdens. This research aims to evaluate the occurrence and intensity of squeezing based on a set of statistical approaches. We utilize statistical analyses and regression functions to predict the squeezing of soft rocks and provide the following conclusions:

- The statistical analyses detect *BQ* and *SSR* as significant predictors for squeezing in soft rocks.

- In contrast to some previous studies, overburden cannot be considered a reliable predictor for estimating squeezing.
- Although *K* and *D* can affect the intensity and occurrence of squeezing, the results confirm that these effects can result in a minor change.
- In the case of the tectonic-based squeezing of soft rocks, while we incorporate no tectonic-based predictor, the database includes some datasets related to such a squeezing condition. However, the number of correctly-predicted cases is lower for the tectonic-based squeezing.
- The results of regression models are proved more accurate compared to the previous well-known empirical approaches.
- The higher the number of used datasets is, the more accurate the results of regression models will be (provided that the database is processed by proper statistical procedures).
- The percentage of correctly- and over-estimated cases by the regression links is close to 80 % (in total). Therefore, designing a support system based on the presented regression functions can avoid squeezing problems. The *BQ-SSR* charts can also help estimate possible erroneous predictions.
- Considering the accuracy of the developed regression models to predict the validation database, i.e., the datasets unavailable in the fitting database, the presented methodology can be trusted to evaluate the squeezing condition of a new database.

6. Acknowledgements

We thank the editor and reviewers whose valuable comments and suggestions have led us to improve this manuscript. M. Akbariforouz is supported by the National Natural Science Foundation of China (No. 41890852), awarded to C. Zheng. Q. Zhao is supported by the FCE Start-up Fund for New Recruits at the Hong Kong Polytechnic University (P0034042) and the Early Career Scheme of the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 25220021).

References

Acaps. (2016). Data Cleaning: Dealing with Messy Data. 1–19.

Aksoy, C. O., Ogul, K., Topal, I., Ozer, S. C., Ozacar, V., & Posluk, E. (2012). Numerical modeling of no-deformable support in swelling and squeezing rocks. *International Journal of Rock Mechanics and Mining Sciences*, *52*, 61–70. https://doi.org/10.1016/j.ijrmms.2012.02.008

Arora, K., Gutierrez, M., Hedayat, A., & Xia, C. (2020). Tunnels in squeezing clay-rich rocks. *Underground Space (China)*, *XXXX*, 1–14. https://doi.org/10.1016/j.undsp.2020.07.001.

Aydan, O., Akagi, T., & Kawamoto, T. (1993). The squeezing potential of rocks around tunnels; Theory and prediction. *Rock Mechanics and Rock Engineering Journal*, *26*(2), 137–163. https://doi.org/10.1007/BF01023620

Aydan, O., Akagi, T., & Kawamoto, T. (1996). The squeezing potential of rock around tunnels: Theory and prediction with examples taken from Japan. In *Rock Mechanics and Rock Engineering Journal, 29 (3),* 125–143. https://doi.org/10.1007/BF01032650

Barla, G., & Barla, M. (2008). Innovative tunneling construction methods in squeezing rocks, *Ingegneria Ferroviaria*, *63*(12), 1017–1031.

Barla, G., Debernardi, D., & Sterpi, D. (2012). Time-Dependent Modeling of Tunnels in Squeezing Conditions. *International Journal of Geomechanics*, *12*(6), 697–710. https://doi.org/10.1061/(ASCE)gm.1943-5622.0000163

Bhasin, R., & Grimstad, E. (1996). The Use of Stress-Strength Relationships in the Assessment of Tunnel Stability. *Tunneling and Underground Space Technology*, *11*(1 SPEC. ISS.), 93–98. https://doi.org/10.1016/0886-7798(95)00047-X

Chen, Y., Li, T., Zeng, P., Ma, J., Patelli, E., & Edwards, B. (2020). Dynamic and Probabilistic Multi-class Prediction of Tunnel Squeezing Intensity. *Rock Mechanics and Rock Engineering Journal*, *53*(8), 3521–3542. https://doi.org/10.1007/s00603-020-02138-8

Chern, J. C., C. W. Yu, & H. C. Kao. (1998). Tunneling in squeezing ground. *Fourth International Conference on Case Histories in Geotechnical Engineering*, 793–796.

Debernardi, D., & Barla, G. (2009). New viscoplastic model for design analysis of tunnels in squeezing conditions. *Rock Mechanics and Rock Engineering Journal*, *42*(2), 259–288. https://doi.org/10.1007/s00603-009-0174-6

Denis, D. J. (2018). SPSS Data Analysis for Univariate, Bivariate, and Multivariate Statistics. In *SPSS Data Analysis for Univariate, Bivariate, and Multivariate Statistics*. https://doi.org/10.1002/9781119465775.

Doerffel, K., 1967. Die Statistische Auswertung Von Analysenergebnissen. In: Schormuller, I. (Ed.), *Handbuch lebensmittelchemie, 2, Springer, Berlin*, 1194–1246

Dwivedi, R. D., Singh, M., Viladkar, M. N., & Goel, R. K. (2013). Prediction of tunnel deformation in squeezing grounds. *Engineering Geology*, *161*, 55–64. https://doi.org/10.1016/j.enggeo.2013.04.005

Farhadian, H., & Nikvar-Hassani, A. (2020). Development of a new empirical method for tunnel squeezing classification (TSC). *Quarterly Journal of Engineering Geology and Hydrogeology*, *53*(4), 655–660. https://doi.org/10.1144/qjegh2019-108

Feng, X., & Jimenez, R. (2015). Predicting tunnel squeezing with incomplete data using Bayesian networks. *Engineering*

Geology, *195*, 214–224. https://doi.org/10.1016/j.enggeo.2015.06.017.

Fox, J. (2016). Applied regression analysis and generalized linear models. *3rd Edition. SAGE Publications, Inc.*

Fritz, P. (1984). An analytical solution for axisymmetric tunnel problems in elasto-viscoplastic media. *International Journal for Numerical and Analytical Methods in Geomechanics*, *8*(4), 325–342. https://doi.org/10.1002/nag.1610080403

Gao, F., Stead, D., & Kang, H. (2015). Numerical Simulation of Squeezing Failure in a Coal Mine Roadway due to Mining-Induced Stresses. *Rock Mechanics and Rock Engineering Journal*, *48*(4), 1635–1645. https://doi.org/10.1007/s00603-014-0653-2

Ghaboussi, J., & Gioda, G. (1977). On the time-dependent effects in advancing tunnels. *International Journal for Numerical and Analytical Methods in Geomechanics*, *1*(3), 249–269. https://doi.org/10.1002/nag.1610010303

Ghasemi, E., & Gholizadeh, H. (2019). Development of Two Empirical Correlations for Tunnel Squeezing Prediction Using Binary Logistic Regression and Linear Discriminant Analysis. *Geotechnical and Geological Engineering*, *37*(4), 3435–3446. https://doi.org/10.1007/s10706-018-00758-0

Gioda, G. (1981). A finite element solution of no-linear creep problems in rocks. *International Journal of Rock Mechanics and Mining Sciences And*, *18*(1), 35–46. https://doi.org/10.1016/0148-9062(81)90264-3

Goel, R. K., Jethwa, J. L., & Paithankar, A. G. (1995). Tunneling through the young Himalayas - A case history of the Maneri-Uttarkashi power tunnel. *Engineering Geology*, *39*(1–2), 31–44. https://doi.org/10.1016/0013-7952(94)00002-J

Guan, Z., Deng, T., Du, S., Li, B., & Jiang, Y. (2012). Markovian geology prediction approach and its application in mountain tunnels. *Tunneling and Underground Space Technology*, *31*, 61–67. https://doi.org/10.1016/j.tust.2012.04.007

Guan, Z., Jiang, Y., & Tanabashi, Y. (2009). Rheological parameter estimation for the prediction of long-term deformations in conventional tunnels. *Tunneling and Underground Space Technology*, *24*(3), 250–259. https://doi.org/10.1016/j.tust.2008.08.001

Guan, Z., Jiang, Y., Tanabashi, Y., & Huang, H. (2008). A new rheological model and its application in mountain tunneling. *Tunneling and Underground Space Technology*, *23*(3), 292–299. https://doi.org/10.1016/j.tust.2007.06.003.

Gutierrez, M., & Xia, C. C. (2009). Squeezing potential of tunnels in clays and clay shales from normalized undrained shear strength, unconfined compressive strength, and seismic velocity: I*n Proc. 6th Intl. Symp. Geotechnical Aspects of Underground Construction in Soft Ground*, Shanghai, China.

Hadjigeorgiou, J., & Karampinos, e. (2017). Design tools for squeezing ground conditions in hard rock mines. *Eighth International Conference on Deep and High-Stress Mining – J Wesseloo*. Australian Centre for Geomechanics, Perth, ISBN 978-0-9924810-6-3. doi:10.36487/ACG_rep/1704_47_Hadjigeorgiou

Hasanpour, R., Rostami, J., & Ünver, B. (2014). 3D finite difference model for simulating double shield TBM tunneling in squeezing grounds. *Tunneling and Underground Space Technology*, *40*, 109–126. https://doi.org/10.1016/j.tust.2013.09.012.

Hasanpour, R., Rostami, J., Schmitt, J., Ozcelik, Y., & Sohrabian, B. (2019). Prediction of TBM jamming risk in squeezing grounds using Bayesian and artificial neural networks. *Journal of Rock Mechanics and Geotechnical Engineering*. https://doi.org/ 10.1016/j.jrmge.2019.04.006

Hastie, T., Tibshirani, R., & Friedman, J. (2017). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Springer, New York, NY*.

Hoek, E., 2001. Big tunnels in bad rock. *Journal of Geotechnical and Geoenvironmental Engineering, 127*, 726–740. http://dx.doi.org/10.1061/(asce)1090-0241(2001)127:9(726).

Hoek, E., & Guevara, R. (2009). Overcoming squeezing in the Yacambu-Quibor tunnel, Venezuela. *Rock Mechanics and Rock Engineering Journal*, *42*(2), 389–418. https://doi.org/10.1007/s00603-009-0175-5

Hoek, Evert, & Marinos, P. (2000). Predicting tunnel squeezing problems in weak heterogeneous rock masses. Part 1: Estimating rock mass strength. *Tunnels and Tunnelling International*, *Part 1-2*(2), 1–20. http://www.rockscience.com/hoek/references/H2000d.pdf

Huang, Z., Liao, M., Zhang, H., Zhang, J., Ma, S., & Zhu, Q. (2021). Predicting Tunnel Squeezing Using the SVM-BP Combination Model. *Geotechnical and Geological Engineering volume, 40*, 1–27.

Jimenez, R., & Recio, D. (2011). A linear classifier for probabilistic prediction of squeezing conditions in Himalayan tunnels. *Engineering Geology*, *121*(3–4), 101–109. https://doi.org/10.1016/j.enggeo.2011.05.006

Khanlari, G., meybodi, R. G., & Mokhtari, E. (2012). Engineering geological study of the second part of water supply Karaj to Tehran tunnel with emphasis on squeezing problems. *Engineering Geology*, *144–145*, 9–17. https://doi.org/10.1016/j.enggeo.2012.06.001

Kovari, K., & Staus, J. (1996). Basic considerations on tunnels in squeezing ground. *Rock Mechanics and Rock Engineering Journal*, *29*(4), 203–210. https://doi.org/10.1007/bf01042533

Kriegel, H.-P., Kröger, P., Schubert, E., & Zimek, A. (2010). Outlier Detection Techniques. *IEEE Communications Surveys*, *12*(2), 159–170.

Kulhawy, F. H. (1974). Finite element modeling criteria for underground openings in rocks. *International Journal of Rock Mechanics and Mining Sciences And*, *11*(12), 465–472. https://doi.org/10.1016/0148-9062(74)91996-2

Li, S., Zhao, H., & Ru, Z. (2012). Deformation prediction of tunnel surrounding rock mass using CPSO-SVM model.

Journal of Central South University, *19*(11), 3311–3319. https://doi.org/10.1007/s11771-012-1409-3

Lyu, H. M., Shen, S. L., Zhou, A. N., & Zhou, W. H. (2019). Flood risk assessment of metro systems in a subsiding environment using the interval FAHP-FCA approach. *Sustainable Cities and Society*, *50*(June), 101682. https://doi.org/10.1016/j.scs.2019.101682

Lyu, H. M., Shen, S. L., Zhou, A., & Yang, J. (2019). Perspectives for flood risk assessment and management for the mega-city metro system. *Tunneling and Underground Space Technology*, *84*(September 2018), 31–44. https://doi.org/10.1016/j.tust.2018.10.019

Mahdevari, S., & Torabi, S. R. (2012). Prediction of tunnel convergence using Artificial Neural Networks. *Tunneling and Underground Space Technology*, *28*(1), 218–228. https://doi.org/10.1016/j.tust.2011.11.002

Malan, D. F. (1999). Time-dependent behavior of deep-level tabular excavations in hard rock. *Rock Mechanics and Rock Engineering Journal*, *32*(2), 123–155. https://doi.org/10.1007/s006030050028

Palmstrom, A. (2001). *Characterizing Rock Masses*. *16*(3), 1–26.

Palmstrom, A., & Broch, E. (2006). Use and misuse of rock mass classification systems with particular reference to the Q-system. *Tunneling and Underground Space Technology*, *21*(6), 575–593. https://doi.org/10.1016/j.tust.2005.10.005

Palmstrom, A., & Stille, H. (2007). Ground behavior and rock engineering tools for underground excavations. *Tunneling and Underground Space Technology*, *22*(4), 363–376. https://doi.org/10.1016/j.tust.2006.03.006

Pan, Y. W., & Dong, J. J. (1991). Time-dependent tunnel convergence-II. Advance rate and tunnel-support interaction. *International Journal of Rock Mechanics and Mining Sciences And*, *28*(6), 477–488. https://doi.org/10.1016/0148-9062(91)91123-9

Panthi, K. K., & Nilsen, B. (2007). Uncertainty analysis of tunnel squeezing for two tunnel cases from Nepal Himalaya. *International Journal of Rock Mechanics and Mining Sciences*, *44*(1), 67–76. https://doi.org/10.1016/j.ijrmms.2006.04.013

Phienwej, N., Thakur, P. K., & Cording, E. J. (2007). Time-Dependent Response of Tunnels Considering Creep Effect. *International Journal of Geomechanics*, *7*(4), 296–306. https://doi.org/10.1061/(ASCE)1532-3641(2007)7:4(296)

Ramoni, M., & Anagnostou, G. (2010). Tunnel boring machines under squeezing conditions. *Tunneling and Underground Space Technology*, *25*(2), 139–157. https://doi.org/10.1016/j.tust.2009.10.003

Ramoni, M., & Anagnostou, G. (2011). The interaction between shield, ground, and tunnel support in TBM tunneling through squeezing ground. *Rock Mechanics and Rock Engineering Journal*, *44*(1), 37–61. https://doi.org/10.1007/s00603-010-0103-8

Scholz, M., Fraunholz, M., & Selbig, J. (2008). No-linear Principal Component Analysis: Neural network models and applications, *Lecture Notes in Computational Science and Engineering*, *58*, 44–67. https://doi.org/10.1007/978-3-540-73750-6_2

Shafiei, A., Parsaei, H., & Dusseault, M. B. (2012). Rock squeezing prediction by a support vector machine classifier. *46th US Rock Mechanics / Geomechanics Symposium 2012*, *1*(June), 489–503. https://doi.org/10.13140/RG.2.1.3836.3040

Shalabi, F. I. (2005). FE analysis of time-dependent behavior of tunnels in squeezing ground using two different creep models. *Tunneling and Underground Space Technology*, *20*(3), 271–279. https://doi.org/10.1016/j.tust.2004.09.001

Shen, Y., Yan, R. X., Yang, G. Sh., Xu, G. L., & Wang, Sh. Y. (2017). Comparisons of Evaluation Factors and Application Effects of the New [BQ] GSI System with International Rock Mass Classification Systems. *Geotechnical and Geological Engineering 35(4), 1-26*. DOI: 10.1007/s10706-017-0259-z

Shrestha, G. L., & Broch, E. (2008). Influences of the valley morphology and rock mass strength on tunnel convergence: With a case study of Khimti 1 headrace tunnel in Nepal. *Tunneling and Underground Space Technology*, *23*(6), 638–650. https://doi.org/10.1016/j.tust.2007.12.006

Singh, B., Jethwa, J. L., Dube, A. K., & Singh, B. (1992). Correlation between observed support pressure and rock mass quality. *Tunneling and Underground Space Technology Incorporating Trenchless*, *7*(1), 59–74. https://doi.org/10.1016/0886-7798(92)90114-W

SPSS. (2010). *IBM SPSS Statistics 19 Brief Guide*. 171.

Sterpi, D., & Gioda, G. (2009). Visco-Plastic behavior around advancing tunnels in squeezing rocks. *Rock Mechanics and Rock Engineering Journal*, *42*(2), 319–339. https://doi.org/10.1007/s00603-007-0137-8

Sulem, J., Panet, M., & Guenot, A. (1987). An analytical solution for time-dependent displacements in a circular tunnel. *International Journal of Rock Mechanics and Mining Sciences And*, *24*(3), 155–164. https://doi.org/10.1016/0148-9062(87)90523-7

Sun, Y., Feng, X., & Yang, L. (2018). Predicting Tunnel Squeezing Using Multi-class Support Vector Machines. *Advances in Civil Engineering*, *2018*, 17–20. https://doi.org/10.1155/2018/4543984

Swannell, N., Palmer, M., Barla, G., & Barla, M. (2016). The geotechnical risk management approach for TBM tunneling in squeezing ground conditions. *Tunneling and Underground Space Technology*, *57*, 201–210. https://doi.org/10.1016/j.tust.2016.01.013

Thut, A., Naterop, D., Steiner, P., & Stolz, M. (2000). Tunneling in squeezing rock-yielding elements and face control. *Solexperts AG Mettlenbachstrasse 25 CH – 8617 Monchaltorf Switzerland*.

Tran Manh, H., Sulem, J., Subrin, D., & Billaux, D. (2015). Anisotropic Time-Dependent Modeling of Tunnel Excavation in Squeezing Ground. *Rock Mechanics and Rock Engineering Journal, 48(6)*, 2301–2317. doi:10.1007/s00603-015-0717-y.

Wang, X., Iura, T., Jiang, Y., Wang, Z., & Liu, R. (2021). Deformation and Mechanical Characteristics of Tunneling in Squeezing Ground: A case study of West Section of the Tawarazaka Tunnel in Japan. *Tunneling and Underground Space Technology, 109, 103697*. doi:10.1016/j.tust.2020.103697.

Wheelwright, M., Levine, Z., Garfinkel-Castro, A., Bushman, T., & Brewer, S. C. (2020). Principal Component and Factor Analysis. In *Advanced Quantitative Research Methods for Urban Planners*. https://doi.org/10.4324/9780429325038-5.

Wu, G., Chen, W., Tian, H., Jia, S., Yang, J., & Tan, X. (2018). *Numerical evaluation of a yielding tunnel lining support system used in limiting large deformation in squeezing the rock. Environmental Earth Sciences, 77(12).* doi:10.1007/s12665-018-7614-0.

Yan-jun, S., Rui-xin, Y., Geng-she, Y., Guang-li, X., & Shan-yong, W. (2017). Comparisons of Evaluation Factors and Application Effects of the New $[BQ]_{GSI}$ System with International Rock Mass Classification Systems. *Geotechnical and Geological Engineering, 35(6), 2523–2548.* doi:10.1007/s10706-017-0259-z .

Yang, L., Schopf, J. M., Dumitrescu, C. L., & Foster, I. (2006). Statistical data reduction for efficient application performance monitoring. *Sixth IEEE International Symposium on Cluster Computing and the Grid, 2006. CCGRID 06*, 327–334. https://doi.org/10.1109/CCGRID.2006.97

Yao, J. B., Yao, B. Z., Li, L., & Jiang, Y. L. (2012). A hybrid model for displacement prediction of tunnel surrounding rock. *Neural Network World*, *22*(3), 263–275. https://doi.org/10.14311/NNW.2012.22.015

Yassaghi, A., & Salari-Rad, H. (2005). Squeezing rock conditions at an igneous contact zone in the Taloun tunnels, Tehran-Shomal freeway, Iran: A case study. *International Journal of Rock Mechanics and Mining Sciences*, *42*(1), 95–108. https://doi.org/10.1016/j.ijrmms.2004.07.002

Zhang, J., Li, D., & Wang, Y. (2020). Predicting tunnel squeezing using a hybrid classifier ensemble with incomplete data. *Bulletin of Engineering Geology and the Environment*, *79*(6), 3245–3256. https://doi.org/10.1007/s10064-020-01747-5