

J. JAKUBOWSKI*[#], J.B. STYPULKOWSKI**[#], F.G. BERNARDEAU***[#]

**MULTIVARIATE LINEAR REGRESSION AND CART REGRESSION ANALYSIS OF TBM
PERFORMANCE AT ABU HAMOUR PHASE-I TUNNEL**

**ANALIZA WSKAŹNIKÓW WYDAJNOŚCI DRAŻENIA TARCZAMI TBM
DLA TUNELU ABU HAMOUR ETAP I, Z ZASTOSOWANIEM WIELORAKIEJ REGRESJI LINIOWEJ
I REGRESJI CART**

The first phase of the Abu Hamour drainage and storm tunnel was completed in early 2017. The 9.5 km long, 3.7 m diameter tunnel was excavated with two Earth Pressure Balance (EPB) Tunnel Boring Machines from Herrenknecht. TBM operation processes were monitored and recorded by Data Acquisition and Evaluation System. The authors coupled collected TBM drive data with available information on rock mass properties, cleansed, completed with secondary variables and aggregated by weeks and shifts. Correlations and descriptive statistics charts were examined. Multivariate Linear Regression and CART regression tree models linking TBM penetration rate (PR), penetration per revolution (PPR) and field penetration index (FPI) with TBM operational and geotechnical characteristics were performed for the conditions of the weak/soft rock of Doha. Both regression methods are interpretable and the data were screened with different computational approaches allowing enriched insight. The primary goal of the analysis was to investigate empirical relations between multiple explanatory and responding variables, to search for best subsets of explanatory variables and to evaluate the strength of linear and non-linear relations. For each of the penetration indices, a predictive model coupling both regression methods was built and validated. The resultant models appeared to be stronger than constituent ones and indicated an opportunity for more accurate and robust TBM performance predictions.

Keywords: EPB TBM, TBM performance, penetration rate, field penetration index, CART trees, machine learning, multivariate regression

Pierwszy etap budowy systemu odpływowego Abu Hamour został ukończony na początku roku 2017. Tunel o długości 9,5 km i średnicy 3,7 m przeprowadzono z zastosowaniem dwóch maszyn drążących z równoważeniem ciśnienia gruntu (EPB TBM), wyprodukowanych przez Herrenknechta. Przebieg pracy maszyn TBM był monitorowany i zapisywany przez automatyczny system zbierania danych.

* DEPT. OF GEOMECHANICS, CIVIL ENG. & GEOTECHNICS; AGH UNIVERSITY, KRAKÓW, POLAND

** CDM SMITH, WOODBURY, NY, USA

*** CDM SMITH, DOHA, QATAR

Corresponding author: jacek.jakubowski@agh.edu.pl

Autorzy połączyli te dane z dostępnymi informacjami o właściwościach masywu skalnego, oczyścili dane, uzupełnili zmiennymi wtórnymi oraz zagregowali tygodniami i zmianami roboczymi. Zbadano korelacje i statystyki opisowe. Metodami liniowej regresji wielorakiej i regresji CART zbudowano modele łączące wskaźniki wydajności drążenia (PR, PPR, FPI) z ich charakterystykami operacyjnymi oraz charakterystykami geotechnicznymi słabego masywu skalnego rejonu Doha, w którym prowadzono tunel. Obydwie zastosowane metody regresji dają interpretowalne modele oraz stosują odmienne algorytmy obliczeniowe, co pozwala na wzbogacenie wyników. Głównym celem analizy było znalezienie możliwie najlepszych podzbiorów zmiennych objaśniających oraz ocena siły znalezionych związków liniowych i nieliniowych. Dla każdego wskaźnika wydajności zbudowano też model predykcyjny wykorzystujący obydwie metody regresji. Zbudowane w ten sposób modele wynikowe okazały się silniejsze od modeli składowych. To wskazuje drogę możliwej poprawy dokładności i stabilności przewidywań wskaźników wydajności TBM.

Słowa kluczowe: EPB TBM, wydajność TBM, prędkość drążenia, połowy wskaźnik drążenia, drzewa CART, uczenie maszyn, regresja wieloraka

1. Introduction

Complex physical phenomena and interactions make TBM performance sensitive to multiple conditions. Understanding of those relations, specifically the ability to model and estimate TBM performance, is important at all stages of a project, including technology selection, tunnel design, resources planning and excavation process. It is crucial for the project costs and its timely completion.

CSM (Rostami, 1997) and NTNU (1998, Bruland, 1998, 2014) models are often used for estimating TBM performance. Other theoretical or empirical formulas and predictive models were proposed or evaluated by Hassanpour et al. (2010, 2011), Delisio and Zhao (2014), Gong and Zhao (2009), Benato and Oreste (2015), Ramezanzadeh (2005), Bieniawski et al. (2007), Palmstrom (1995), Barton (2000), Yagiz (2008), Hamidi et al. (2010), Jain et al. (2014), Salimi et al. (2016) and others. Physical and mixed methods consider cutting mechanism, disk cutter forces and wear, and are based on various field and laboratory tests and simulations. Empirical models rely on historical operational and field data.

Due to the sensitivity to geological and geotechnical conditions, predictions of TBM performance are accurate under a limited range of conditions, specifically rock mass quality conditions and technology details. TBM performance prediction models for hard rock, described comprehensively by Hassanpour et al. (2011), have been applied for the conditions of the weak/soft rocks of Doha by authors in an earlier publication (Stypulkowski et al., 2017). As expected, hard rock approach to FPI prediction does not give satisfactory results in the weak/soft rocks of Doha without major adjustments.

While TBM performance predictive models are quite common and perform well for a wide range of strong rock mass conditions, EPB TBM performance predictions and models for weak/soft rock conditions are not well documented and verified. Factors impacting the EPB TBM Penetration Rate (PR) in the conditions of soft ground are different than in hard rock. The PR in hard rock is linked to rock hardness. In weak/soft rock, no increase of penetration can be associated with thrust and torque increase typically observed in hard rock (Avunduk, 2012). Maher (2017) suggested that there is a relationship between PR and surfactants used in soft ground.

Authors have previously conducted a regression analysis of TBM performance with General Linear Regression and Neural Networks models for Earth Pressure Balance technology and weak/

soft rock of Doha. The dataset was aggregated by strokes, thus the dependences and variability had been different due to a smaller aggregation step and consequently much larger number of cases. Still, strong linear relations were observed and high predictive potential of the dataset processed by Neural Networks was revealed.

This paper continues research into the issue and describes relations linking performance indicators with EPB TBM operational and geotechnical features. We present explanatory Linear Regression and CART regression trees analysis for better understanding of the processes and relations within the dataset aggregated by weeks and shifts. Then we use these two analytical techniques to build and validate predictive models with the Averaged model approach.

2. General project description and tunneling accomplishment

Abu Hamour Surface & Ground Water Drainage Tunnel (Phase I) is a 9.5 km long, 3.7 m diameter tunnel, excavated about 30 m below the ground surface. The tunnel runs from an existing shaft to a retrieving shaft located within the footprint of a future pumping station on the coastline. The tunneling was carried out simultaneously by two similarly equipped Herrenknecht EPB TBMs. Phase I of the project also included 19 access shafts and drop shafts facilitating runoff inflows along the route of the tunnel. Offline shafts used sequential excavation tunneling methods to connect to the tunnel in 15 locations. There are also 3 online shafts, 1 existing shaft and 6 additional branch shafts facilitating connections to the main tunnel, accomplished by microtunnelling. Access Shaft AS11 located in the center of the tunnel alignment was the launching location for the eastern and western TBMs drives facilitating mining in opposite directions. It also provided access for tunnel construction for the duration of the project.

Tunneling was run in both drives simultaneously, 6 days a week, two 11-hour shifts a day. Eastern production runs started on May 13, 2014, and western on July 7, 2014. The breakthrough was reached on July 22, 2015, and July 29, 2015, respectively. The best weekly progress was 24.7 m/day in the western and 27.3 m/day in the eastern drive of the tunnel. The excavation was carried out by two TBMs with rotating cutter-heads fitted with cutting wheels and tools. The excavated material was collected by buckets and transported through the openings provided between the cutters into the excavation chamber. A screw conveyor extracted the spoil from the chamber and discharged it onto the belt conveyor installed on the TBM rear which in turn offloaded the material onto muck skips. The muck skips were operated by diesel powered locomotives, which traveled on the tunnel rail track. The permanent lining of the tunnel consisted of dowelled, pre-cast concrete segments reinforced with steel fibers. TBMs specification, overall tunneling performance and stroke performance statistics are presented in Tab. 1 and 2. For more information on the project, see Stypulkowski et al. (2013, 2017).

3. Geology

The Qatar region is geologically a part of the Arabian Gulf Basin between the Arabian shield and Iranian mobile belt. The post Cretaceous sedimentation is basically a sequence of shallow marine limestone with occasional shale in a shallow basin (Abu Zeid, 1991; LeBlanc, 2008).

TABLE 1

Specification of TBMs

Weights and Dimensions	Total length	124.5 m
	TBM weight	330 ton
	TBM core weight	172 ton
	Bore Diameter	4.52m
Design Parameters	Curve radius	300 m
	Gradient	0.05%
	UCS	2-65 MPa
	Hydrostatic pressure	2-3 bars
Segmental lining	Segments per ring	6+key
	Total rings installed	7198
	Ring width	1.3m
	Segment thickness	250mm
Cutterhead	Style	Mixed
	Cutters	17
	Scrapers	75
	Buckets	8
Cutterhead Drive	Cutterhead speed	0-4.5 min ⁻¹
	Nominal torque	2167 kNm
	Maximum thrust	20891 kN
Conveyor	Type	Screw conveyor
	Diameter	600 mm

TABLE 2

Tunneling performance

	TBM East	TBM West
Overall tunneling performance		
Advance per work days [meter/day]	17	15
Advance per calendar days [meter/day]	14	12
Rings per work days [ring/day]	13	12
Rings per calendar days [ring/day]	11	9
Average per stroke performance		
Tunneling time [min]	30	35
Liner erecting time [min]	22	27
Stop time [min]	36	42
Cutterhead torque [kNm]	1149	1226
Thrust force per cutter [kN]	341	382
Penetration rate [m/hr]	2.86	2.39

The formations encountered in Doha area comprise of Quaternary marine, aeolian and sabkha deposits: Rus of Lower Eocene, Lower Dammam and Upper Dammam of Middle Eocene and Lower Dam of Lower Miocene.

During the project we came across: Simsima Limestone which contains Dolomite of Upper Dammam formation, Midra Shale of Lower Dammam formation and Rus of Lower Eocene. The Dammam formation is usually divided into three main stratigraphic units: Simsima, Midra, and Rus. Based on findings in the shafts, specifically their strength, we further sub divided Simsima into Simsima A, Simsima B and Simsima C. Simsima C is typically fine to medium grained dolomitic limestone, loosely compacted, highly weathered, extremely weak to very weak, heterogeneous. It has been observed that compaction and strength increases when transitioning into Simsima B, which is moderately weathered, weak to medium strong. The grade of compaction increases and the ground mass becomes strong to very strong as it becomes Simsima A. Transition from Simsima to Midra is marked by layers of gypsum. Midra was encountered at different elevations along the alignment. Rus is characterized by the presence of horizontal layers with varying color, composition, grain size, weathering and strength. Fig. 1 reflects the geological profile along the tunnel alignment.

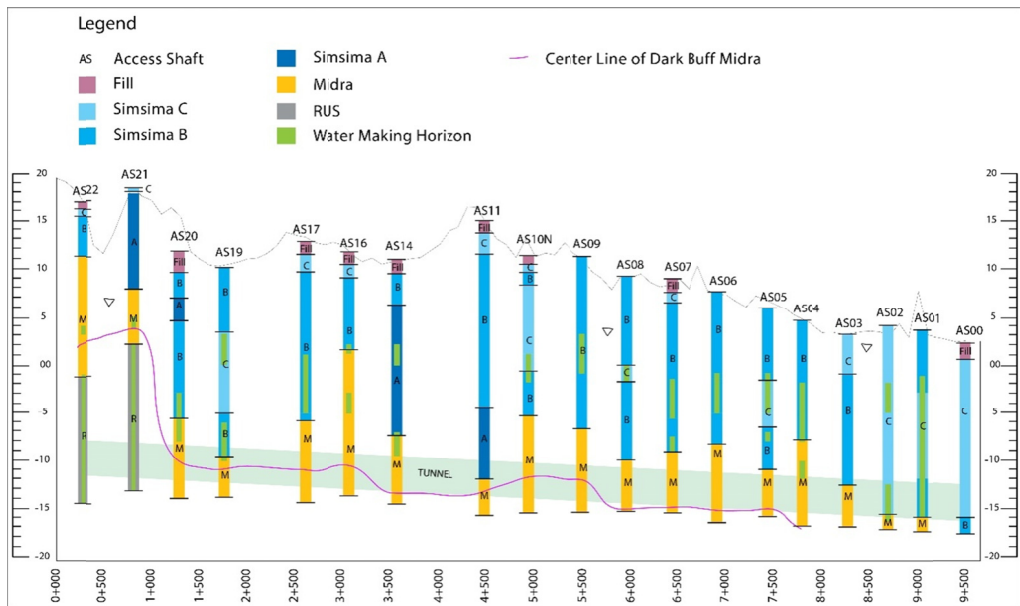


Fig. 1. Interpreted geology from shaft mapping. Scales in [m]

Shale, dolomitic limestone, pinkish dolomitic limestone, green clay, marl, clay stone and gypsum constitute the main rock mass. The thickness of layers varies from 2 cm to 1.5 m. Strength varies from medium strong, weak and very weak, to extremely weak rock. The silty/sandy limestone is only 1/3 while clay rich components are 2/3 at the tunneling horizon, at the shaft junctions. The characteristics of the ground mass along the alignment vary significantly both horizontally and vertically (Pathak et al. 2015, Stypulkowski et. al 2017).

4. Intact rock and rock mass properties

Laboratory tests on rock samples obtained from borings before construction gave Uniaxial Compressive Strength (UCS) between 2 and 65 MPa. Tensile Strength from Brazilian splitting tests (BTS) ranged 0.1-9.1 MPa. Point Load Test strength estimate ranged 0.1-7.5 MPa. Elastic modulus from uniaxial compression tests ranged 0.4-49 GPa and from pressure meter testing in initial loading ranged 0.1-5.3 GPa. Down-hole seismic test for the dynamic elastic modulus ranged 1.0-4.3 GPa. The mapping results and analyses summarized in Tab. 3 indicate that rock quality was poor to extremely poor. Original rock mass quality assessment based on pre-design borings is generally conservative, when compared with shaft mapping results. Limited laboratory testing of the muck has been conducted. Test results indicate clay content 23-35%, silt content 9-16%, PI 36-57% and LL 10-13%.

TABLE 3

Rock mass quality assessment summary

Rock mass quality parameters	Borehole based	Shaft mapping max	Shaft mapping min
RQD	55	50	40
Joint Number	6	15	20
Joint Roughness	3	4	3
Joint Alteration	3	3	2
Joint Water	0.66	0.66	0.5
Q	1.23	2	0.79
UCS rating	2	2	1
RQD rating	13	8	3
Spacing rating	5	8	5
Condition of discont.	9	10	0
Ground water rating	10	10	7
RMR	34	38	18

5. Input dataset, explanatory and responding variables

The TBM excavation process was monitored by Herrenknecht Operation and Guidance Systems and the drive data collected by the TBM Data Acquisition and Evaluation System. The contractor kept their own, paper records. At the design stage of the project, general geology investigations and laboratory testing of rock properties were performed on cores from several boreholes along the axis of the tunnel. Additional information on geology, rock and rock mass properties came from mapping and were collected during access shafts excavation. The TBM drive data collected automatically and manually were coupled with available information on rock mass properties, cleansed and completed with secondary variables. All outlying values were reviewed case by case and corrected for typos and compared with other sources of data, or replaced with

nearby stroke values. Launching and learning curves data were omitted. The collected data were primarily integrated and aggregated by strokes. For the purpose of this analysis it was then aggregated by weeks and shifts, so each case represents a week of tunneling by one shift and one drive. As a result, the number of cases was reduced to 197. Aggregation comprised of mainly averaging, summing or evaluating a median, depending on variable meaning and property.

Three TBM performance indices were considered: penetration rate PR [mm/min], penetration per revolution PPR [mm/rev] and field penetration index FPI [(kN/cutter)/(rev/min)].

Operational and steering characteristics considered in this analysis included: Rotation [rev/min], Thrust and Jacking forces [kN], Torque [kNm] and secondary: Rotational Power [kNm/min] and Rotational Resistance coefficient. They also included over 20 other operational characteristics referring to Earth pressure sensor measurements, Foam delivery (pressure, air, surfactant volume), Grease seal pressure, Grouting injection pressures and volumes, Guidance measurements, Jacking force I/C ratio (force in invert to force in crown ratio), Drive labeling eastern or western drive of the tunnel. For some of those quantities both average and maximum values were considered in the analysis.

Operational characteristics (like steering, foam delivery) are adjusted automatically or manually by the TBM operator, in response to a change of tunneling conditions and performance. In effect, they impact the TBM performance and are dependent on the performance and other conditions at the same time. Quantities showing direct relation to PR (like muck weight or injected grouting volume) were eliminated from the analysis. Injection pressure and foam delivery measures, including surfactant consumption, have low or moderate statistical correlation with penetration rates and are considered among independent variables.

For Geological strata formations through which the tunnel is driven, the elevation of crown, axis and invert locations were identified along its length. It was interpolated from available core drills and nearby shafts mapping. Three types of Simsimi, one of Midra Shale and one of Rus were used to describe formations strata. UCS and BTS tests for each of the geological formations were performed. Coupling information on strata in crown, axis and invert and strength of each formation, an average UCS and BTS were estimated and assigned to a particular location along the tunnel.

Geotechnical quantities included the mechanical properties of rock in relation to strata geology: UCS [MPa], BTS [MPa], and some other geotechnical conditions: Overburden thickness [m] and Ground water pressure [m].

The regression methods select a narrow subset of variables as determined by a specific, automatic or semi-automatic variable selection rule. The two model-building methods applied differ in many aspects and the respective best subset criteria selections are different as well. The relation of the responding variable with explanatory variables can also differ, so the selection of the best subset varies for each of the responding variables. Finally, at the stage of model building the best subset of variables is selected by each analytical method and for each responding variable individually.

The final dataset consisting of 197 cases and about 40 manually preselected variables was used for the multivariate regression analysis. It was proceeded by an overview of basic statistics and correlations between dependent and explanatory variables.

6. Descriptive statistics and correlations for key variables

Descriptive statistics for the responding, major geotechnical and operational variables for eastern and western drive of the tunnel have been marked as boxes and whiskers in Fig. 2-4. The changes of the observed penetration, geotechnical and operational quantities along the tunnel axis are plotted in Fig. 5-7. Pearson correlations between dependent and independent variables are listed in Tab. 4.

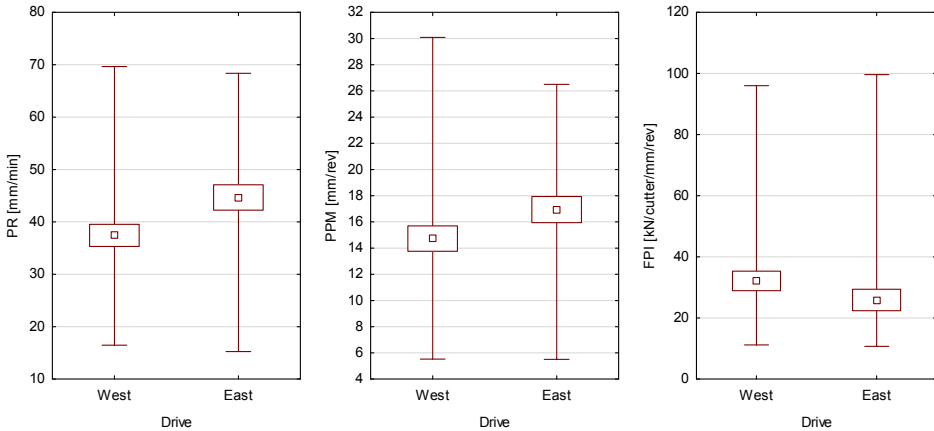


Fig. 2. Statistics of dependent variables: PR, PPM and FPI, for western and eastern drive of the tunnel. Box and whiskers plots showing mean, 95% conf. intervals for the mean, and min-max intervals

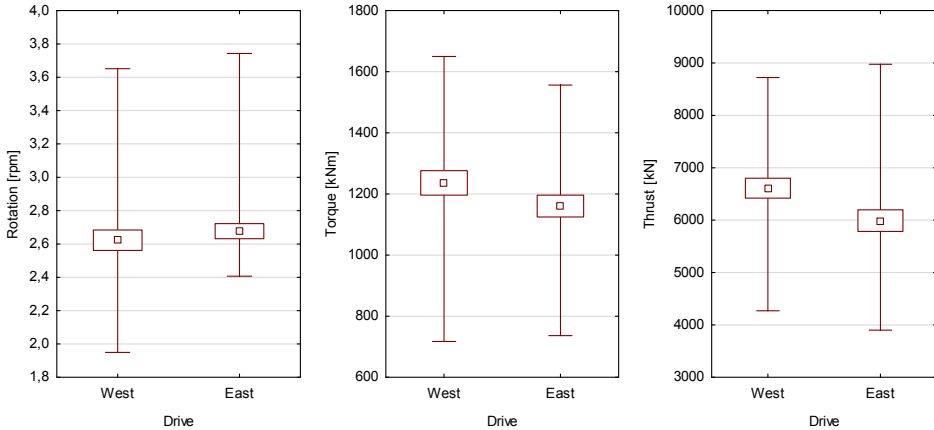


Fig. 3. Statistics of three operational characteristics: Rotation, Torque and Thrust for western and eastern drive of the tunnel. Box and whiskers plots showing mean, 95% conf. intervals for the mean, and min-max

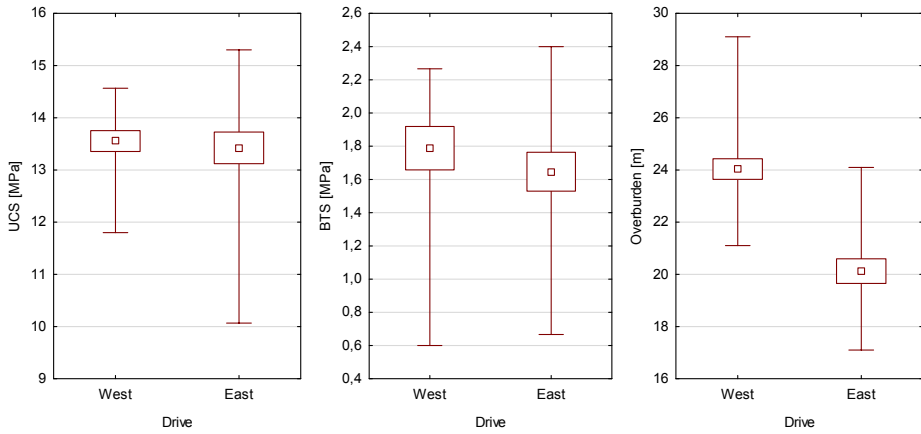


Fig. 4. Statistics of three geotechnical characteristics: UCS, BTS and Overburden, for western and eastern drive of the tunnel. Box and whiskers plots showing mean, 95% conf. intervals for the mean, and min-max

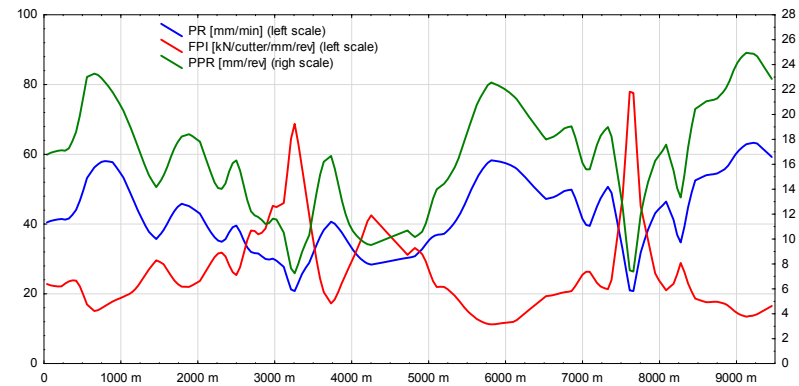


Fig. 5. Dependent variables' runs along the tunnel line, smoothed plots. Left axis: PR [mm/min], FPI [kN/cutter/mm/rev]. Right axis: PPR [mm/rev]

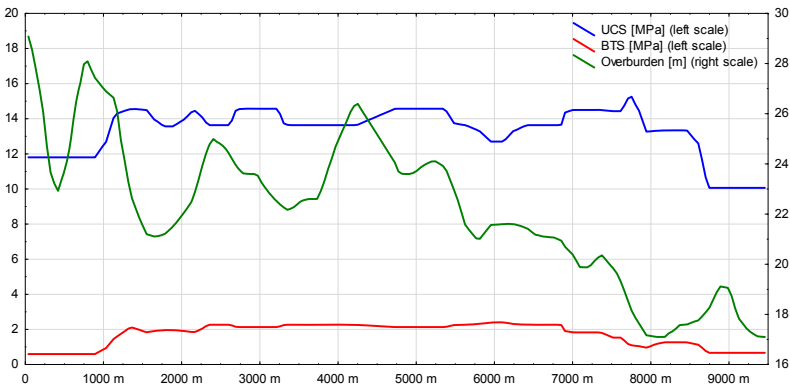


Fig. 6. Three geotechnical variables' runs along the tunnel line, smoothed plots. Left axis: UCS and BTS [MPa]. Right axis: Overburden [m]

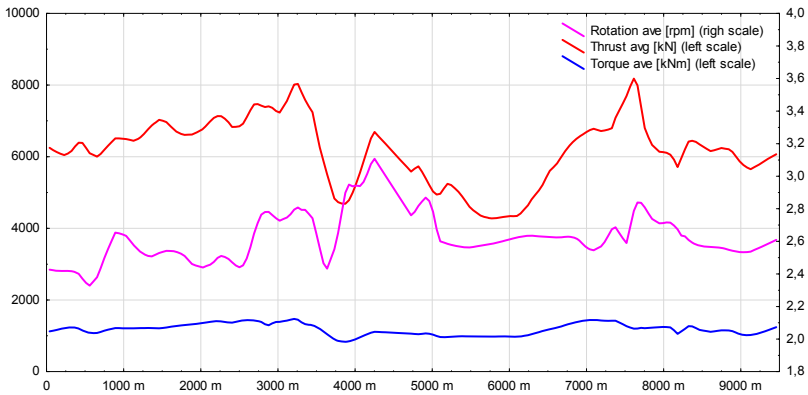


Fig. 7. Three operational variables’ runs along the tunnel line, smoothed plots. Left axis: Thrust avg [kN], Torque avg [kNm], Rotational power avg [kNm/min]. Right axis: Rotation avg [rpm]

TABLE 4

Linear correlation for dependent and some explanatory variables. $N = 197$.
Steering variables are excluded from later FPI models (respective values in brackets)

Variables	Bivariate Pearson correlation			
	PR	PPR	FPI	ln(FPI)
UCS [MPa]	-0.612	-0.616	0.420	0.499
BTS [MPa]	-0.401	-0.423	0.187	0.229
Overburden [m]	-0.188	-0.153	0.005	0.066
Ground water pressure [m]	-0.264	-0.162	0.089	0.160
Earth pressure from 4 sensors max [bar]	-0.129	-0.096	0.314	0.321
Earth pressure c1	0.229	0.194	-0.382	-0.397
Jacking force I/C ratio	0.361	0.305	-0.432	-0.442
Foam delivery pressure avg [bar]	-0.177	-0.135	0.384	0.319
Foam delivery air [m3]	-0.062	-0.007	0.072	0.116
Surfactant consumption	0.226	0.237	-0.387	-0.368
Grease seal pressure max [bar]	-0.338	-0.268	0.271	0.344
Guidance c4	0.413	0.413	-0.293	-0.356
Rotation max [rpm]	-0.529	-0.668	(0.541)	(0.579)
Torque avg [kNm]	-0.251	-0.227	(0.359)	(0.443)
Thrust avg [kN]	-0.522	-0.479	(0.725)	(0.767)
Thrust max [kN]	-0.488	-0.443	(0.696)	(0.734)
Rotational power avg [kNm/min]	-0.392	-0.481	(0.547)	(0.621)
Rotational power max [kNm/min]	-0.668	-0.722	(0.770)	(0.816)
Rotational resistance avg [-]	0.364	0.333	(-0.468)	(-0.449)
Rotational resistance max [-]	-0.105	-0.123	(-0.041)	(-0.014)

Multivariate linear and non-linear tree regression models of data acquired during tunneling were estimated using Statistica (Tibco, 2017), specifically with GLZ and C&RT workspace nodes.

7. General Linear Regression models

GLR regression models were constructed for each of the dependent variables (PR, PPR and FPI) in order to search for the set of independent variables explaining the three considered TBM penetration rates, and to measure the strength of the models and evaluate the part of its variability that can be explained with the collected dataset. Instead of the FPI variable itself, its natural logarithm was analyzed. Regarding all FPI models, steering variables were excluded from FPI regression models in order to avoid any trivial, predefined relations being reflected in descriptive models, but an alternative approach may probably also be considered.

Generalized Linear Model module with identity link function and normal distribution was used. Semi-automated, forward entry, and forward stepwise algorithms for explaining variable selection were applied. The significance and statistical assumptions of GLR models were checked with appropriate statistics, as required by classic statistical techniques. This ensures model's usability for interpretation and prediction. To evaluate the fitness or the strength of all the multidimensional regression models, Model response vs. Observations plot and coefficient of determination R^2 were used. The number of 197 cases gave 20+ cases per single variable allowing robust and stable model parameters estimation.

The GLR models specification and results summary are presented in Tab. 5. Explanatory variables selected using a forward entry algorithm and employed by these models are listed. The sequence is related to the importance of a specific variable for the model measured by the Wald statistic, while the +/- signs refer to the positive or negative contribution of a variable to the model. All models are statistically significant and all variables are significant in a model. Tab. 5 shows coefficients of determination for each of the considered linear models. The strongest linear relation and regression model is achieved for penetration per revolution PPR. To judge how much the multivariate approach improves models compared to any single variable regression, the highest R^2 value for a single independent variable model is also shown. Fig. 8, 9 and 10 left, show Observations vs. Model response graphs for GLR models.

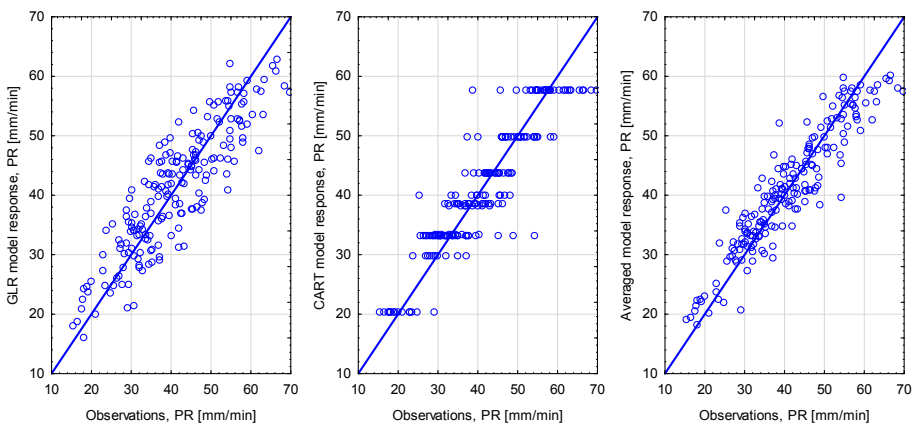


Fig. 8. Observations vs. model response graphs for Penetration rate PR [mm/min] models. GLR (left), CART regression (mid), Averaged model (right)

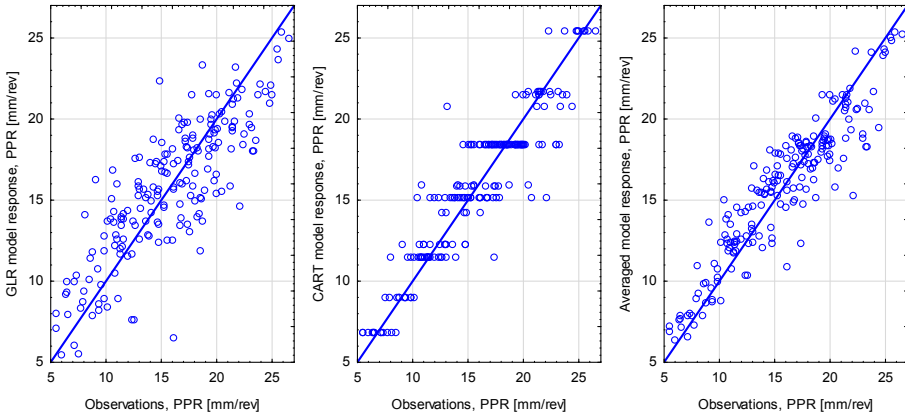


Fig. 9. Observations vs. model response graphs for Penetration rate per revolution PPR [mm/rev] models. GLR (left), CART regression (mid), Averaged model (right)

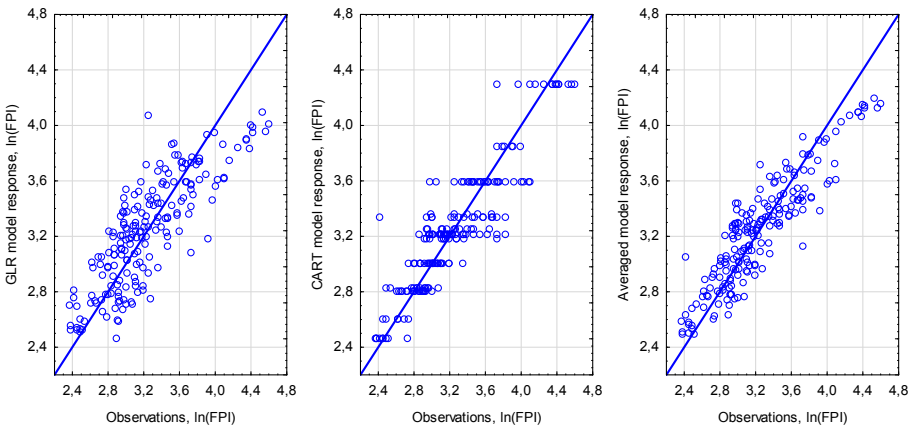


Fig. 10. Observations vs. model response graphs for Field penetration index ln(FPI) models. GLR (left), CART regression (mid), Averaged model (right)

TABLE 5

GLR models' specification and coefficient of determination R^2 . $N = 197$

Dependent variable	Explanatory variables	Model's R^2	Bivariate best R^2
PR	(+)RotResistance avg, (-)BTS, (-)RotPower max, (-)RotResistance max, (-)UCS, (-)Foam delivery air, (+)Jacking I/C ratio, (-)Ground water pressure, (+)Surfactant consumption	0.776	0.446
PPR	(-)RotPower max, (+)Torque avg, (-)Thrust avg, (+)Thrust max, (-)UCS, (-)Foam delivery air, (-)BTS, (+)Jacking I/C ratio	0.796	0.521
ln(FPI)	(+)UCS, (+)Earth pressure max, (+)Overburden, (-)Surfactant consumption, (+)BTS, (+)Earth pressure c1, (-)Jacking I/C ratio, (+)Foam delivery pressure,	0.679	0.249

8. Non-linear CART tree regression models

Classification and Regression Trees (CART) was the second regression method used. The idea behind this method is to reflect the nature of original data by splitting cases into discrete clusters with a set of simple logical conditions. In case of regression problems, each terminal node assigns a value to the elements it contains. Simple interpretation of results is one of the advantages of trees method. There are no rigorous assumptions specific for classic statistical regression methods. A major issue is to avoid over-fitting and provide high performance of the model at the same time.

The classic Classification and Regression Trees (CART) algorithm (Breiman et al., 1984) was used here. Stopping on variance pruning method was applied. Specifically, the following stopping parameters were applied: maximum of 5 tree levels, minimum of 5 % for splitting nodes, minimum 5 cases in child node. To validate CART models, a 10-fold cross-validation method was used and within-node variance monitored. The validation procedure gave satisfactory results and proved models' stability and usability.

CART trees models based on geotechnical and operational features were built for PR, PPR and ln(FPI). The results are presented in Tab. 6, Fig. 8-10 mid and Fig. 11-13.

TABLE 6

Coefficients of determination R^2 for three regression models. $N=197$

Dependent variable	R^2 by regression method		
	GLR	CART	Averaged model
PR	0.776	0.828	0.869
PPR	0.796	0.878	0.895
ln(FPI)	0.679	0.803	0.823

Coefficients of determination values are high, substantially higher than the respective linear multivariate models (Tab. 6). Observations vs. Model response graphs for CART models (Fig. 8, 9, 10 mid) show horizontally scattered, detached groups referring to terminal clusters of cases (leafs) in the respective tree structure (compare Fig. 8 mid and Fig. 11).

A tree structure directly reflects the model. The resulting tree structure for PR is presented in Fig. 11. The best discriminating variable and value is Rotation max. Cases with Rotation max ≤ 2.9 rpm have generally higher PR than these having Rotation max > 2.9 rpm. The higher PR cases node (Rotation max ≤ 2.9), is further split by UCS = 18.4 MPa value into higher PR node (UCS ≤ 13.3 MPa) and lower PR node (USC > 13.3 MPa). On the right-hand side of the graph the low PR node (Rotation max > 2.9) is best discriminated by Thrust max = 8268 kN. Following the tree splits, one can conclude, that the highest PR values in the dataset are observed for: (Rotation max ≤ 2.9 rpm) and (UCS ≤ 13.3 MPa) and (Ground water pressure ≤ 18.4 m). The lowest PR values in the dataset sample are observed for: (Rotation max > 2.9 rpm) and (Thrust max > 8268). Finally we have 9 non-terminal nodes and 10 terminal nodes in the tree. The size of the terminal nodes ranges from 8 to 35 cases, the variance within the nodes ranges from 8.5 to 39.3.

The resulting tree structure for PPR and ln(FPI) are presented in Fig. 12 and 13, respectively. The highest PPR values in the dataset are observed for: (Rotation max ≤ 2.9 rpm) and (UCS ≤ 13.3 MPa) and (Ground water pressure ≤ 18.4 m) and (Guidance to roll end ≤ -1.36 deg). The lowest PR values in the dataset sample are observed for: (Rotation max > 2.9 rpm) and

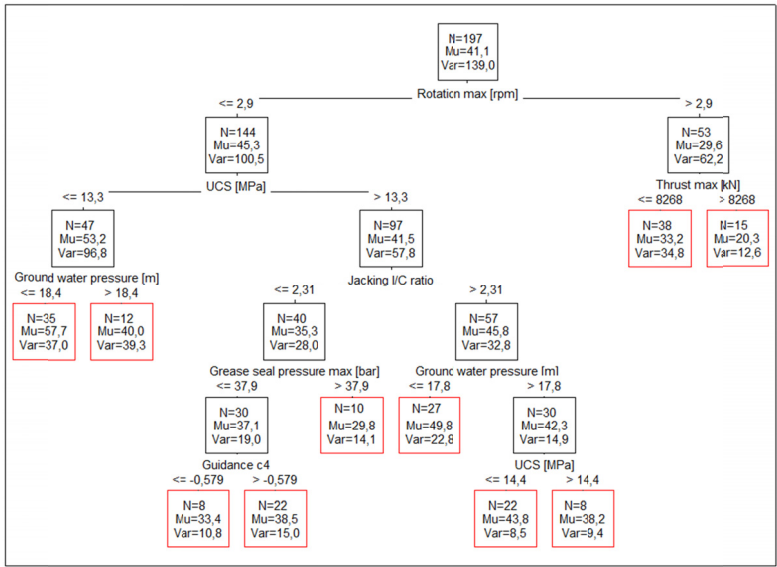


Fig. 11. CART tree structure representing PR model

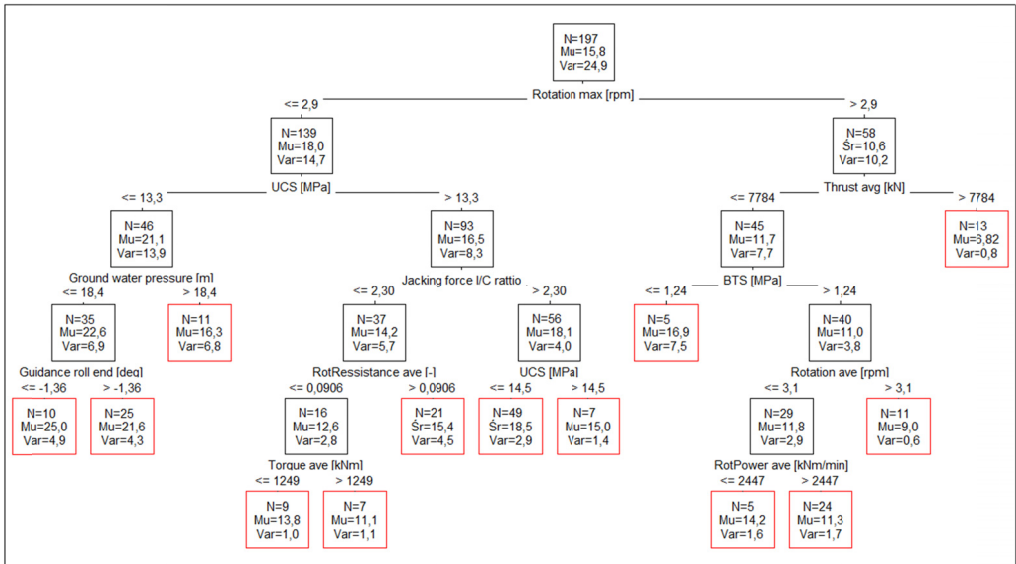


Fig. 12. CART tree structure representing PPR model

(Thrust max > 7784). There are 12 non-terminal and 13 terminal nodes in the tree. The highest FPI values in the dataset (Fig. 13) are observed for: (Jacking force I/C ratio ≤ 1.90 rpm) and (UCS > 13.7 MPa) and (Overburden ≤ 22.7 m). The lowest PR values in the dataset sample are observed for: (Jacking force I/C ratio > 1.90 rpm) and (Earth pressure max $\leq 0,25$). There are 12 non-terminal nodes and 13 terminal nodes in the tree.

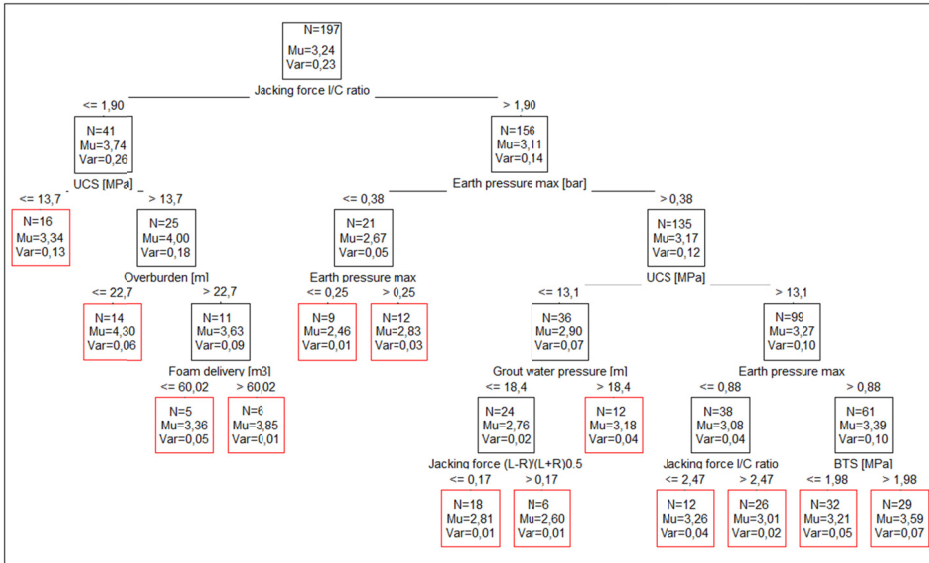


Fig. 13. CART tree structure representing ln(FPI) model

9. Averaged models

Averaged model response is the arithmetic mean of GLM and CARD model responses. Unlike each constituent models, it cannot be interpreted directly. Instead, the combination of two different modeling approaches can improve the robustness and predictive power of the averaged model.

GLR and CART model validation accounts for the averaged model being applicable, still the averaged model was further validated. The dataset was split in proportion 3:1 training and validation sample. The training sample was used for building an averaged model which was then deployed with the validation sample. Each of the 3 averaged models (PR, PPR and ln(FPI)) proved to be sufficiently stable.

Some narrowing of the model response range is observed, but all the averaged models perform considerably better than their constituent models. Observation vs. Averaged model response graphs (Fig. 8, 9 and 10 right) show that averaged models are less scattered around diagonals than respective GLR and CARD models, have lower variation and prediction errors. Tab. 6 confirms stronger relations.

10. Summary

Common availability of TBM drive monitoring and automated acquisition of data along with investigated geology and rock mass properties allows for producing reliable, empirical models of TBM penetration indices for various TBM technologies and rock mass conditions.

The presented work includes the results of a descriptive regression analysis linking TBM penetration rate (PR), penetration per revolution (PPR) and field penetration index (FPI) with

several TBM operational and geotechnical features, specifically for Earth Pressure Balance technology and weak/soft rocks of Doha. In order to search for the best set of variables explaining the considered TBM penetration rates, and to evaluate the part of their variability that can be explained with the collected data, the General Linear Regression was primarily used. Significant relations explaining considerable portion of variability were captured. Instead of linear equations of GLR, regression trees establish a set of simple classification conditions grouping cases and assigning them a value. CART regression trees were also used to show dependencies within the dataset and explain its structure.

Taking into account complex phenomena and interactions accompanying the process of EPB TBM operations, and rough, interpolated data on geology, the performed data analysis showed fairly strong and significant relations. The presented attempt to explain TBM performance indicators PR, PPR and FPI with operational and geotechnical data may be considered advisable and applicative.

The Averaged response modeling approach was used to probe the predictive potential of the dataset aggregated by weeks/shifts and to check the strength of averaged models. Combining two different regression methods appeared to be effective for building predictive models of TBM performance, and will be further investigated.

References

- Abu Zeid M.M., 1991. *Lithostratigraphy and framework of sedimentation of the subsurface Paleogene succession in northern Qatar*. Arabian Gulf. N. Jb. Geol. Palaeont. Mh.: 191-204.
- Avunduk E, Copur H., Bilgin N., Balci C., Tumac D., 2012. *Effect of Some Geotechnical Properties on TBM Performance*. Eurock 2012 ISRM International Symposium, 28-30 May 2012, Stockholm, Sweden.
- Barton N., 2000. *TBM Tunneling In Jointed and Faulted Rock*. Balkema, Rotterdam, p. 173.
- Benato A., Oreste P., 2015. *Prediction of penetration per revolution in TBM tunneling as a function of intact rock and rock mass characteristics*. Int. J. Rock Mech. Min. Sci. 74, 119-127.
- Bieniawski Z.T., Celada B., Galera J.M., 2007. *TBM excavability: prediction and machine-rock interaction*. In: Proceedings, Rapid Excavation and Tunneling Conference, pp. 1118-1130.
- Breiman L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. Chapman & Hall/CRC.
- Bruland A., 1998. *Prediction model for performance and costs*. Norwegian TBM Tunnelling, Publication No. 11. Norwegian Tunnelling Society.
- Bruland A., 2014. *The NTNU prediction model for performance*. Norwegian Tunnelling Technology, Publication No. 23. Norwegian Tunnelling Society.
- Delisio A., Zhao J., 2014. *A new model for TBM performance prediction in blocky rock conditions*. Tunnel. Undergr. Space Technol. 43, 440-452.
- Gong Q.M., Zhao J., 2009. *Development of a rock mass characteristics model for TBM penetration rate prediction*. Int. J. Rock Mech. Mining. Sci. 46 (1), 8-18.
- Hamidi J.K., Shahriar K., Rezaei B., Rostami J., 2010. *Performance prediction of hard rock TBM using rock mass rating (RMR) system*. Tunn.
- Hassanpour J., Rostami J., Khamsehchiyan M., Bruland A., Tavakoli H.R., 2010. *TBM performance analysis in pyroclastic rocks, a case history of Karaj Water Conveyance Tunnel (KWCT)*. J. Rock Mech. Rock Eng. 4, 427-445.
- Hassanpour J., Rostami J., Zhao J., 2011. *A new hard rock TBM performance prediction model for project planning*. Tunnel. Undergr. Space Technol. 26, 595-603.
- Jain P., Naithani A.K., Singh .TN., 2014. *Performance characteristics of tunnel boring machine in basalt and pyroclastic rocks of Deccan traps – a case study*. Int. J. Rock Mech. Geo. Eng. 6 (1), 36-47.

- LeBlanc J., 2008. *A Fossil Hunting Guide to the Tertiary Formations of Qatar. Middle East*. First Edition, March 2008.
- Maher J.A., 2017. *Machine Learning Approach to Predicting and Maximizing Penetration Rates in EPB TBMs*. <https://src.acm.org>, access 2017.
- NTNU, 1998. *Report IB-98 Hard Rock Tunnel Boring – Advance Rate and Cutter Wear*. Department of Civil and Transport Engineering, Trondheim, Norway.
- Palmstrom A., 1995. *RMI, a rock mass characterization system for rock engineering purposes*. Chapter 7, Ph.D. thesis, Oslo University, Norway.
- Pathak A.K., Stypulkowski J.B, Bernardeau F.G., 2015. *Super-vision of Engineering Geological Activities during Construction of Abu Hamour Surface and Ground Water Drainage Tunnel Phase-I Doha, Qatar*. [In:] International Conference on Engineering Geology in New Millennium at IIT, New Delhi 27-29 Oct 2015. Special Publication, J. of .EG. October 2015, pp. 467-486.
- Ramezanzadeh A., 2005. *Performance analysis and Development of new models for performance prediction of hard rock TBMs in rock mass*. Ph.D. thesis, INSA, Lyon, France.
- Rostami J., 1997. *Development of a force estimation model for rock fragmentation with disc cutters through theoretical modeling and physical measurement of crushed zone pressure*, Ph.D. thesis, Colorado School of Mines, Golden, Colorado, USA, p. 249.
- Salimi A., Rostami J., Moormann C., Delisio A., 2016. *Application of nonlinear regression analysis and artificial intelligence algorithms for performance prediction of hard rock TBMs*. Tunn Undergr Space.
- Stypulkowski J.B., Siyam A.A.F.M., Bernardeau F.G., Al Kuwari N.G., 2013. *Abu Hamour Drainage Tunnel. First TBM Mined Tunnel in Doha, Qatar*. [In:] The First Arabian Tunnelling Conference & Exhibition, Dubai, pp. 300-314.
- Stypulkowski J.B., Bernardeau F.G., Jakubowski J., 2017. *Abu Hamour Tunnel Phase I the First TBM Tunnel in Qatar: The art of tunneling in a new world*. [In:] Engineering Challenges for Sustainable Underground Use. GeoMEast 2017. Springer, Cham.
- Tajduś A., Cała M., Tajduś K., 2012. *Geomechanics in underground construction : design and construction of tunnels* (in Polish). AGH, Kraków.
- Tibco Software Inc., 2017. *Statistica* (data analysis software system), version 13.
- Yagiz S., 2008. *Utilizing rock mass properties for predicting TBM performance in hard rock condition*. Tunn. Undergr. Space Technol. 23 (3), 326-339.