Proceedings of the 7th International Conference on Geotechnical and Geophysical Site Characterization
Barcelona, 18 - 21 June 2024

Toward probabilistic ground models for time and cost estimation of tunnel projects

Johan Spross[1#], *Jacob* Grasmick[2], and *Iason* Papaioannou[3]

[1]*KTH Royal Institute of Technology, Division of Soil and Rock Mechanics, Brinellvägen 23, 10044 Stockholm, Sweden*
[2]*Emprise Concepts LLC, 19553 Snowshoe Rd, Evergreen, Colorado, United States*
[3]*Technical University of Munich, Engineering Risk Analysis Group, Theresienstraße 90, 80333 Munich, Germany*
[#]*Corresponding author: johan.spross@byv.kth.se*

ABSTRACT

The construction time and cost of a rock tunnel project are highly dependent on the rock mass quality and encountered ground behaviour. In most rock tunnel projects, the knowledge about the ground conditions along the tunnel is limited, making it difficult to predict accurately the construction time and cost. The KTH model takes a probabilistic approach to address this problem; however, it does not account for the spatial variability of the ground conditions. This paper investigates an alternative probabilistic ground model to be used within the KTH model that enables accounting for the spatial variability through Markov random field theory. The new ground model employs a parametric approach to describe the properties of the Markov field, hence, enabling the simulation of the ground conditions with limited data, but does not consider the epistemic uncertainty in the model parameters. This will be the addressed in future research.

Keywords: tunnel; construction time; ground model; spatial uncertainty; Markov chain random field

1. Introduction

The construction time and cost of a rock tunnel project is highly dependent on the rock mass quality and encountered ground behaviour, as they affect the complexity of the excavation, amount of installed rock support, as well as the need for water-sealing measures. For most rock tunnel projects, however, the knowledge about the ground conditions along the planned tunnel route is limited, so the engineer must deal with a substantial epistemic uncertainty in the prediction of the project time and cost: The amount of rock support to be installed in a certain tunnel section will not be known until the section has been excavated and the actual ground conditions have been revealed. In addition, there are aleatory uncertainties related to occurrence of undesirable events, like excessive groundwater inflow or block fall. There are also aleatory uncertainties related to the performance of the contractor: Even if the ground conditions were completely known in advance, it would not be possible to predict the exact construction time and cost. These uncertainties – both epistemic and aleatory – are an important contributing factor to the apparent challenge of predicting accurately the final construction time and cost of tunnel projects (see e.g. Flyvbjerg et al. 2002, Miranda Sarmento and Renneboog 2017, Mohammadi et al. 2023a).

One way to address uncertainties in time and cost estimations is to apply a probabilistic model. Although this is not a heavily researched field in rock engineering, a few different models have been proposed:

- The Decision Aids for Tunnelling (DAT) (Einstein et al. 1999, Min et al. 2008, and several others) consists of a set of probabilistic modules that model the ground conditions and activities involved in a tunnelling construction project;
- The model by Špačková et al. (2013) uses a dynamic Bayesian network to model the tunnel construction process and assess its time and cost;
- The KTH model separates the tunnel into a number of segments and estimates construction time and costs through crude Monte Carlo simulation of the production efforts per segment (Isaksson and Stille 2005, Mohammadi et al. 2023b).

The KTH model provides a straightforward and easy-to-understand framework to assess the uncertainty in time and cost estimations. However, its approach to model the epistemic uncertainty about the ground conditions is rather simplified and mostly based on the engineer's subjective judgement. An attempt toward a less subjective approach was presented by Spross and Lidmar (2023), who suggested modelling the proportions of construction classes along the tunnel as a Dirichlet distribution, which fits well into the overall framework of the KTH model. While the Dirichlet distribution has some attractive features, it ignores the impact of spatial variability; moreover, it is based on an underlying assumption that the geotechnical investigations provide independent observations, which is not entirely realistic.

In this paper, we therefore investigate an alternative approach to model the uncertainty about the ground conditions in the KTH model. This approach considers spatial variability explicitly, through a Markov chain random field model of the ground conditions (Sartore 2013, Li 2007). Markov random fields have been previously proposed for use within the DAT model (Einstein et al. 1999) and the model by Guan et al. (2014). However, Markov chain random fields require the specification of the full transition probability matrix, which is often infeasible with limited data leading to subjective choices. In this work, we apply a continuous-lag Markov chain random field, which enables its specification with limited data.

In the following, we first provide the basic theory of the KTH model, followed by the new approach to model and incorporate uncertainty of ground conditions along tunnels within the KTH model. The approach is illustrated with an illustrative synthetic example. The paper closes with a discussion on the advantages and limitations of the proposed approach.

2. KTH model for time and cost estimation

The KTH model separates construction time into two uncertain components: normal construction time, T_{N}, and exceptional construction time, T_{E}. The final predicted construction time is the sum of these two components: $T = T_{\mathrm{N}} + T_{\mathrm{E}}$. Here, T_{N} considers the expected variability in time of the planned construction activities, while T_{E} considers delays caused by any disruptive events, which can be of both geological and human origin. Noting that the KTH model's cost estimation method is essentially a more complex variant of its time estimation method, we consider in this paper only time estimation.

The model calculates T_{N} as the sum of the time it takes to construct a large number of small segments of unit lengths. Their positions along a tunnel of length L are denoted l. The construction times of segments are introduced as the parameter *production effort*, denoted Q_l [h/m]. Thus, the normal time can be assessed as

$$T_{\mathrm{N}} = \sum_{l=1}^{L} Q_l \tag{1}$$

Conceptually, each Q_l, can be considered a function of the prevailing ground conditions at that location, such that $Q_l = f(\mathbf{X}_l)$, where $\mathbf{X}_l$ collects the impacting conditions, for example rock mass quality, groundwater flow, and overburden, As the ground conditions vary along the tunnel, as does the crew's performance from one day to another, the Q_l will also vary accordingly. The engineering challenge in applying the KTH model is to represent this variation accurately in the determination of the Q_l along the tunnel.

Recent research by Mohammadi et al. (2023b) improved the KTH model with respect to the variation in the crew's performance in different construction activities for a set of predefined *construction classes*. Construction classes refer to a set of predetermined design variants for a tunnel excavated using drilling and blasting. Each design variant specifies, for example, type, spacing and length of rock bolts, as well as type and thickness of sprayed concrete, or other support measures as relevant. The design variants have been developed in the planning phase; the appropriate design variant for a particular tunnel segment is then selected based on the observed ground conditions after each blasting round.

It is rather straightforward to estimate the production effort for each construction class, Q_k, where k denotes the considered construction classes, say A–D. The Q_k only considers the aleatory variation in the crew's performance in each work task in that construction class. Thus, for each construction class,

$$Q_k = \sum_{m=1}^{n_{\mathrm{act}}} q_{k,m} \tag{2}$$

where n_{act} are the number of construction activities, and $q_{k,m}$ denote the assessed required times to perform each activity for 1 m of tunnel in the ground conditions of construction class k. The $q_{k,m}$ are random variables, often assigned triangular distributions for convenience, as they are easy for tunnel planners to work with.

To consider the epistemic uncertainty about the ground conditions at any segment l, its production effort Q_l is modelled as a mixture distribution of the different Q_k, weighted with respect to the assessed probabilities (proportions) $\pi_{k,l}$ of being in the respective construction classes at that segment:

$$Q_l = \sum_{k=1}^{K} \pi_{k,l} Q_k \tag{3}$$

where K is the used number of construction classes in the project.

In the original KTH model, the probabilities $\pi_{k,l}$ were interpreted as "deterministic" proportions of the respective classes along the tunnel, simplifying $\pi_{k,l}$ to be identical for every segment, such that all probabilities can be collected in one deterministic vector, $\boldsymbol{\pi} = [\pi_{\mathrm{A}}, \dots, \pi_K]$. As a consequence, the corresponding mixture distributions Q_l were also all identical. In Spross and Lidmar's (2023) approach, the uncertainty in the assessment of the proportions was considered by letting the construction class proportions $\boldsymbol{\pi}$ follow a Dirichlet distibution. In a Monte Carlo simulation, this implies that the N_{MC} simulated construction times T_{N} each have differently sampled proportions of construction classes, but the location of the classes along the tunnel is not addressed.

The improved modelling of $\pi_{k,l}$ is the focus of the present paper. Here we discuss an approach to model explicitly the spatial variation in $\pi_{k,l}$ along the tunnel. The $\pi_{k,l}$ should in this case be interpreted not as a proportion, but as the probability of tunnel segment l being in construction class k, which gives the following matrix for a tunnel of length L and a set of construction classes A, … , K:

$$\boldsymbol{\pi} = \begin{bmatrix} \pi_{\mathrm{A},1} & \cdots & \pi_{\mathrm{A},L} \\ \vdots & \ddots & \vdots \\ \pi_{K,1} & \cdots & \pi_{K,L} \end{bmatrix} \tag{4}$$

Thus, we arrive at different mixture distributions Q_l along the tunnel, instead of them being identical. This means for a Monte Carlo simulation of T_{N} that, for each segment l in the N_{MC} simulated tunnels, one first samples the construction class from $\boldsymbol{\pi}$, then samples the corresponding production effort (Q_l). The Q_l are then summed for each tunnel into T_{N} (eq. (1)).

As the present paper focuses on the modelling of normal time T_{N}, the exceptional time T_{E} is for simplicity not discussed further.

3. Modelling of spatial variability and uncertainty

The interpolation and estimation of geological and geotechnical conditions, including rock classes, between sample data generally entails a high degree of uncertainty. Simulation techniques are often employed to generate multiple equally probable estimations and evaluate the uncertainty. Each estimation, or realisation, is considered to have the same likelihood of representing the ground truth as all other realisations from the simulation. For categorical data like the construction classes A–D, such techniques include, but are not limited to, sequential indicator simulation (Journel 1983), Multiple-Point Statistics (Mariethoz and Caers 2014), Plurigaussian Simulation (Armstrong et al. 2011) and Markov Chain random fields (Carle and Fogg 1997).

In this study, we employ a continuous-lag 1-D Markov chain random field to model the constructions classes that enables the Monte Carlo simulation of the construction classes along the tunnel (Sartore 2013, Li 2007). This method was selected based on its advantages for handling complex probability distributions and its ability to generate reasonable estimates from limited observations.

In this approach, the spatial variability of the categorical dataset is assumed to be characterized by a second-order stationary model defined through the transition probability matrix:

$$\mathbf{T}(h) = \begin{bmatrix} t_{11}(h) & \cdots & t_{1K}(h) \\ \vdots & \ddots & \vdots \\ t_{K1}(h) & \cdots & t_{KK}(h) \end{bmatrix} \tag{5}$$

where h is the lag separation distance, and the entries are the conditional probabilities defined as

$$t_{jk}(h) = \Pr\left\{ \begin{matrix} (\text{category } k \text{ is at } l+h) \mid \\ (\text{category } j \text{ is at } l) \end{matrix} \right\} \tag{6}$$

where j,k = A,…., K, denote the different possible categories and l is the location along the tunnel.

In spatial applications, it is convenient to employ parametric transiograms expressed as continuous-lag functions, such that the transition probability can be computed for any lag distance. This function is assumed to take an exponential form such that

$$\mathbf{T}(h) = \exp(\mathbf{R}h) \tag{7}$$

where $\mathbf{R}$ is the transition rate matrix,

$$\mathbf{R} = \begin{bmatrix} r_{AA} & \cdots & r_{AK} \\ \vdots & \ddots & \vdots \\ r_{KA} & \cdots & r_{KK} \end{bmatrix} \tag{8}$$

Entries r_{jk} are the autotransition rates denoting conditional rates of change per unit length from category j to category k. The autotransition rates are based on mean lengths $\bar{L}_j$ and transition frequencies of embedded occurrences f^*_{kj} for the respective category. The diagonal entries are estimated as $r_{jj} = -1/\bar{L}_j$ and the off-diagonal entries as $r_{jk} = f^*_{ij}/\bar{L}_i$.

There are several ways to compute $\bar{L}_j$, including taking the arithmetic mean or median of the observed lengths of each category. A more robust approach when stratum lengths are not well-understood is to use the maximum likelihood approach where lengths are assumed to be independent realizations from a log-normal distribution (Sartore et al. 2016). Calculation of f^*_{kj} is performed using a 'maximum entropy' method that involves the following iterative proportion filling algorithm (Goodman 1968):

1. Initialise: f_j with $p_j/\bar{L}_j$
2. Compute: $f^*_{jk} = f_j f_k$
3. Compute:
   $f_j = \left(p_j \sum_{i=1}^{K} \sum_{k\neq j}^{K} f^*_{ik}\right) / \left(\bar{L}_j \sum_{k\neq j}^{K} f^*_{jk}\right)$
4. Repeat the second and the third step until convergence.

where p_j is the assumed proportion of class j.

The prediction or simulation of categories at unsampled locations is performed using Markov chain random field (MCRF) theory (Li 2007). In this approach, a random path is taken for each realization of the sequential simulation. At the first randomly picked location, the nearest neighbours are used to estimate the conditional probability distribution (CPD) from which a specific category is drawn. This newly simulated location is added to the sample data set for conditioning in subsequent simulations of other unknown locations. This process is repeated until all unknown locations are visited and assigned a simulated value.

4. Simulation of encounter probabilities of construction classes

To illustrate the application of incorporating spatial variability modelling into probabilistic construction time estimation, a synthetic dataset for a 1000 m long tunnel constructed in a rock of variable quality is presented. Four construction classes are deemed possible in the geological setting. Classes A and B represent good quality hard crystalline rock mass with few (A) to moderate (B) number of rock joints. Classes C and D represent weaker zones and faults, where Class D is particularly challenging to tunnel excavation due to its severely crushed, almost soil-like material.

In this synthetic dataset, 26 observations have been made within the tunnel horizon. We assume no prior information of the geological deposition history. The positions of the observations were, however, decided by an engineering geologist based on some geological understanding of the area, slightly favouring areas deemed more likely to be challenging (classes C and D). Thus, the sampling cannot be considered independent. The observed classes are presented in Table 1.

Fig. 1 presents the transition rate matrix (Eq. (8)). It can be observed that the model estimates a high transition rate for Class D relative to other classes, which corresponds to higher frequency transitions to other states over a short distance. Class A has the lowest transition rates, indicative of continuous sections of this class over longer distances.

For the random field modelling, N = 1000 different samples using unique random paths were drawn. For each location, the two nearest neighbours are used to estimate the conditional probability distribution (CPD) from which a specific category is drawn. From the 1000

Table 1. Position of observed construction classes along 1000 m tunnel.

Position (m)	Observed class	Position (m)	Observed class
20	A	536	D
52	A	542	D
75	B	553	C
130	A	671	A
144	B	732	A
215	B	754	B
247	C	798	A
262	C	850	A
345	A	891	B
372	A	910	B
405	C	924	C
479	A	954	A
515	B	989	A

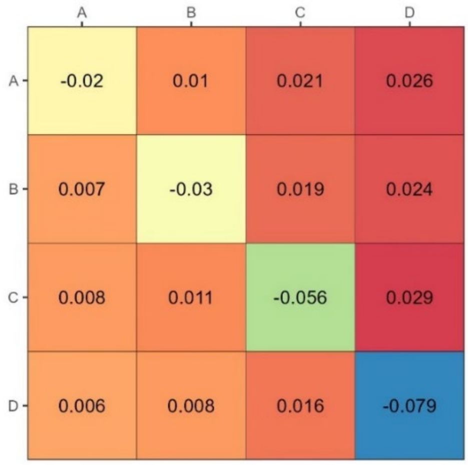

Figure 1. Transition rate matrix for the modelled dataset.

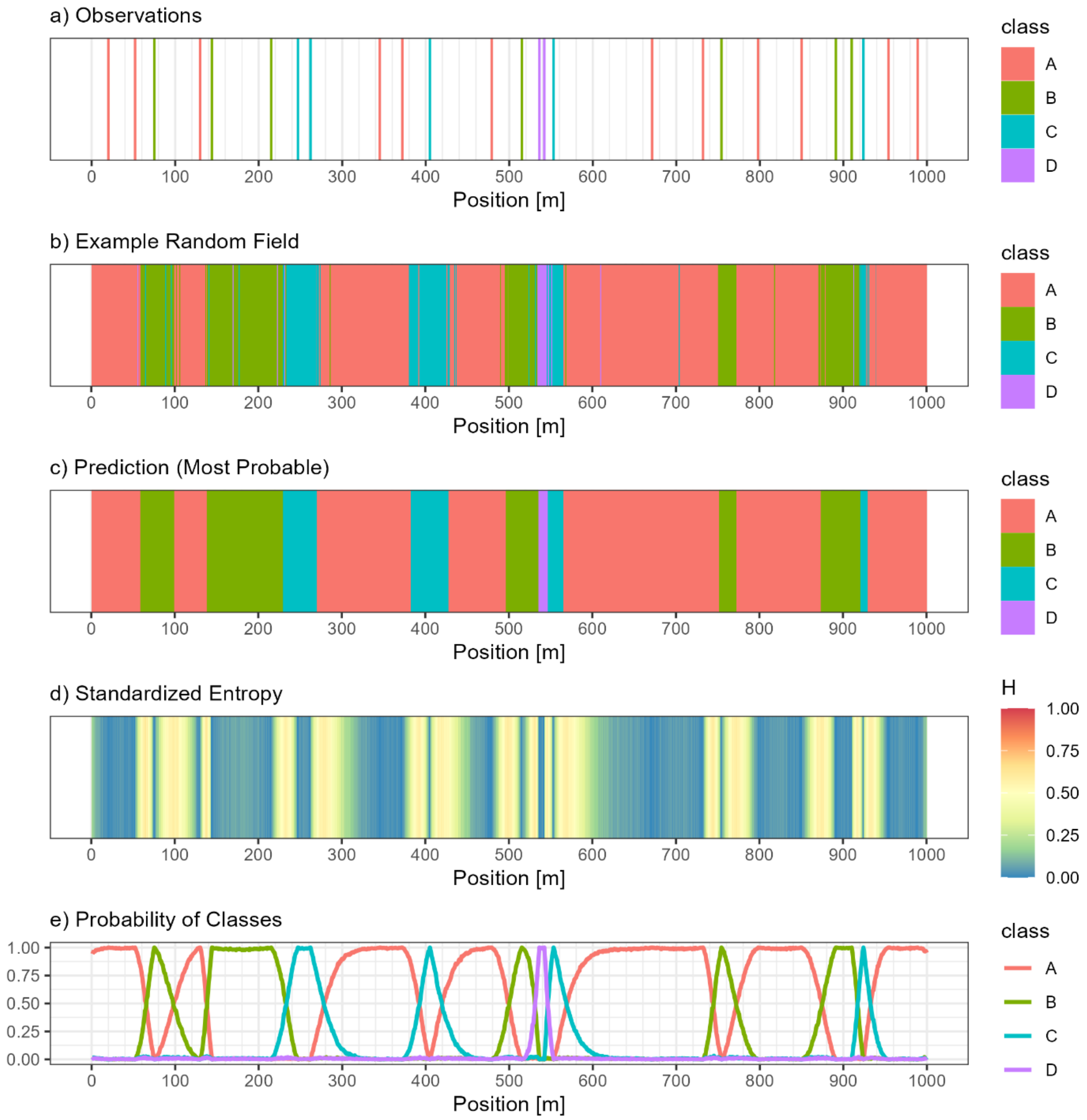

Figure 2. Summary of the Markov random field simulation results: (a) observed categories, (b) an example random field realization, (c) the predicted (most probable) construction class, (d) prediction uncertainty as quantified by the standardized Shannon entropy, and (e) the encounter probability of each construction class versus position.

simulations, statistics, including the probability of each class at a specific location and the most probable class, are estimated. The classification uncertainty is represented by the standardized Shannon entropy (Shannon 1948). The results of the Markov chain random field modelling are summarized by Fig. 2. Specifically, Fig. 2(e) shows the computed encounter probabilities for the construction classes along the tunnel (Eq. (4)), and Fig. 2(b) shows the outcome of one realization of construction classes along the tunnel.

5. Computation of estimated construction time

Having access to 1000 x 1000 realizations of construction classes in a matrix (equivalent to 1000 tunnels of length 1000 m), the total normal construction time T_{N} can be estimated. For this purpose, we apply the triangular distributions for production efforts Q_k that were used by Spross and Lidmar (2023) in their calculation example (Table 2).

The T_{N} is assessed by matching the relevant Q_k with the sampled construction classes for every position along the 1000 simulated tunnels (equivalent to sampling Q_l for every l using Eq. (3)), and then summing the corresponding samples of Q_l along the tunnel lengths (Eq. (1)). The result is shown as a histogram in Fig. 3 and as accumulated time in Fig. 4.

6. Concluding remarks

In this paper we have investigated a method to model the spatial variability in the location of construction classes along a rock tunnel, based on a dataset of observed construction classes at a limited number of locations. The purpose is to improve the KTH model for probabilistic time and cost estimations of tunnels. The investigated method employs Markov Chain random field theory to model and ground conditions along the tunnel and to sample from the corresponding construction classes at each 1-m tunnel segment. This allows straightforward sampling of the production effort Q_l, which is a key parameter within the existing KTH model.

Although the presented method addresses spatial variability explicitly, which is a big step forward in comparison to the deterministic method of assessing proportions used in the current KTH model, it does not address the epistemic uncertainty in the assessment of its model parameters. For rock tunnels, this is a limitation, as the epistemic uncertainty typically is substantial due to limited amount of data from geotechnical investigations in the planning phase. Moreover, we see a need to be able to incorporate subjective information from engineering geologists' expert judgement. To this end, we plan to continue our research, aiming at developing a Bayesian approach that is able to quantify the epistemic uncertainty in the model parameters through combining data from geotechnical investigations with subjective information from expert judgement. This would make the KTH model into a straightforward time and cost estimation tool that allows the user – a client or a contractor – to quantify their risks related to project delay and cost overrun.

Table 2. Assumed triangular distributions, Tri(a, b, c), for production efforts in the four construction classes.

Construction class	a [h/m]	b [h/m]	c [h/m]
A	2.5	3.0	5.0
B	3.0	3.5	6.0
C	4.0	5.0	8.0
D	5.5	8.0	12.5

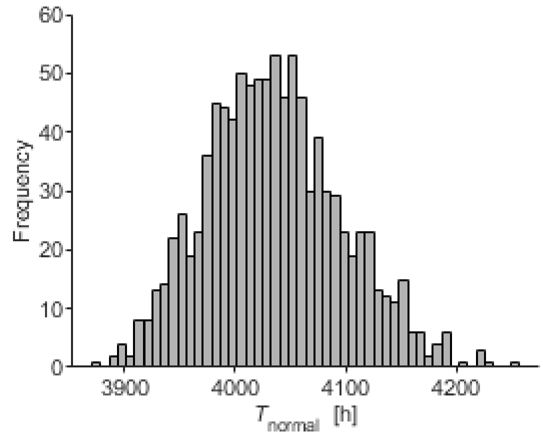

Figure 3. Estimated normal construction time for the illustrative example (1000 simulations).

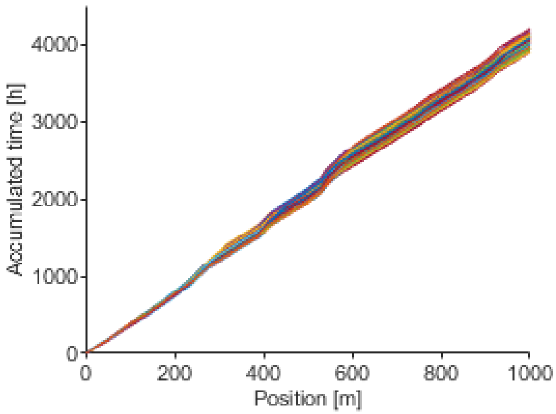

Figure 4. Accumulated time for 100 simulated tunnels. Segments of poorer ground (class C and D) stand out as steeper slopes, e.g. at position 540 m.

Acknowledgements

Johan Spross' research was funded by Formas, a Swedish research council for sustainable development (grant no. 2018-01017).

References

Armstrong, M., Galli, A., Beucher, H., Loch, G., Renard, D., Doligez, B., Eschard, R. and Geffroy, F. *Plurigaussian simulations in geosciences*, Springer Science & Business Media, Heidelberg, 2011.

Carle, S.F. and Fogg, G.E. "Modeling spatial variability with one and multidimensional continuous-lag Markov chains". Math. Geol., 29, pp. 891-918, 1997. https://doi.org/10.1023/A:1022303706942

Einstein, H.H., Indermitte, C., Sinfield, J., Descoeudres, F.P. and Dudt, J.-P. "Decision aids for tunneling". Transp. Res. Rec., 1656(1), pp. 6-13, 1999.

Flyvbjerg, B., Holm, M.S. and Buhl, S. "Underestimating costs in public works projects: error or lie?". J. Am. Plann. Assoc., 68(3), pp. 279-295, 2002. https://doi.org/10.1080/01944360208976273

Goodman, L.A. "The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries". J. Am. Stat. Assoc., 63(324), pp. 1091-1131, 1968. https://doi.org/10.2307/2285873

Guan, Z., Deng, T., Jiang, Y., Zhao, C. and Huang, H. "Probabilistic estimation of ground condition and construction cost for mountain tunnels". Tunn. Undergr. Sp. Tech., 42, pp. 175-183, 2014. https://doi.org/10.1016/j.tust.2014.02.014

Isaksson, T. and Stille, H. "Model for estimation of time and cost for tunnel projects based on risk evaluation". Rock Mech. Rock Eng., 38(5), pp. 373-398, 2005. https://doi.org/10.1007/s00603-005-0048-5

Journel, A.G. "Nonparametric estimation of spatial distributions". Math. Geol., 15, pp. 445-468, 1983. https://doi.org/10.1007/BF01031292

Li, W. "Markov chain random fields for estimation of categorical variables". Math. Geol., 39, pp. 321-335, 2007. https://doi.org/10.1007/s11004-007-9081-0

Mariethoz, G. and Caers, J. *Multiple-point geostatistics: Stochastic modeling with training images*, John Wiley & Sons, Chichester, 2014.

Min, S.Y., Kim, T.K., Lee, J.S. and Einstein, H.H. "Design and construction of a road tunnel in Korea including application of the decision aids for tunneling–a case study". Tunn. Undergr. Sp. Tech., 23(2), pp. 91-102, 2008. https://doi.org/10.1016/j.tust.2007.01.003

Miranda Sarmento, J. and Renneboog, L. "Cost overruns in public sector investment projects". Public Works Manag. Policy, 22(2), pp. 140-164, 2017. https://doi.org/10.1177/1087724X16668357

Mohammadi, M., Spross, J. and Stille, H. "Models to analyze risk in time and cost estimation of tunneling projects". Geotech. Geol. Eng., (in press), 2023a. https://doi.org/10.1007/s10706-023-02627-x

Mohammadi, M., Spross, J. and Stille, H. "Probabilistic time estimation of tunneling projects: The Uri headrace tunnel". Rock Mech. Rock Eng., 56, pp. 703-717, 2023b. https://doi.org/10.1007/s00603-022-03022-3

Sartore, L. "spMC: Modelling spatial random fields with continuous lag Markov chains". R J., 5(2), pp. 16-28, 2013.

Sartore, L., Fabbri, P. and Gaetan, C. "spMC: an R-package for 3D lithological reconstructions based on spatial Markov chains". Comput. Geosci., 94, pp. 40-47, 2016. https://doi.org/10.1016/j.cageo.2016.06.001

Shannon, C.E. "A mathematical theory of communication". Bell Syst. Tech. J., 27(3), pp. 379-423, 1948. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Špačková, O., Šejnoha, J. and Straub, D. "Probabilistic assessment of tunnel construction performance based on data". Tunn. Undergr. Sp. Tech., 37, pp. 62-78, 2013. https://doi.org/10.1016/j.tust.2013.02.006

Spross, J. and Lidmar, J. "Dirichlet distribution for tunnel construction class proportions in probabilistic time and cost estimations". In: *Geo-Risk 2023* (Geotech. Sp. 346), Arlington, VA, 2023, ASCE, Reston, VA, pp. 111-120. https://doi.org/10.1061/9780784484982.012